**AGENDA**

**Wednesday, February 29, 2012**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 am – 5:00 pm</td>
<td><strong>Closed Session: 9:00 am – 3:00 pm</strong>&lt;br&gt;<strong>ACTION:</strong> Review of Secure NAEP Civics Items and Technology and Engineering Literacy (TEL) Task Outlines&lt;br&gt;<em>Lonnie Smith, ETS</em></td>
<td>Secure material sent under separate cover</td>
</tr>
<tr>
<td></td>
<td><strong>Open Session: 3:00 – 5:00 pm</strong>&lt;br&gt;<strong>ACTION:</strong> Review of NAEP Background Questions for U.S. History, Geography, and Civics&lt;br&gt;<em>Committee Members</em></td>
<td>Material sent under separate cover</td>
</tr>
</tbody>
</table>

**Friday, March 2, 2012**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:45 – 9:50 am</td>
<td>Welcome, Introductions, and Agenda Overview&lt;br&gt;<em>Alan Friedman, ADC Chair</em></td>
<td></td>
</tr>
<tr>
<td>9:50 – 11:00 am</td>
<td><strong>ACTION:</strong> Review of NAEP Background Questions for Reading and Mathematics&lt;br&gt;<em>Committee Members</em></td>
<td>Material sent under separate cover</td>
</tr>
<tr>
<td>11:00 – 11:30 am</td>
<td>Briefing on Hewlett Foundation Automated Student Essay Scoring Prize&lt;br&gt;<em>Mark Shermis, University of Akron (via phone)</em></td>
<td>Attachment A</td>
</tr>
<tr>
<td>11:30 am – 12:15 pm</td>
<td><strong>Closed Session</strong>&lt;br&gt;Update on NAEP Mathematics Special Studies&lt;br&gt;<em>Gloria Dion, ETS</em></td>
<td></td>
</tr>
</tbody>
</table>

**Information Item**<br>NAEP Item Review Schedule<br>Attachment B
Mark D. Shermis

Mark D. Shermis, Ph.D., is presently Professor and Dean in the College of Education at The University of Akron. He received his B.A. at the University of Kansas and was on active duty in the U.S. Navy for three years before entering graduate school. After finishing his master's and Ph.D. from the University of Michigan, Dr. Shermis worked for a computer firm and eventually entered academe. Dr. Shermis has played a leading role in bringing computerized adaptive testing to the World Wide Web, and for the last ten years has been involved in research on automated essay scoring. His early efforts resulted in the seminal book on the topic (Jill Burstein, Ph.D., co-editor), Automated Essay Scoring: A Cross-Disciplinary Approach published by Lawrence Erlbaum Associates, Inc.

Dr. Shermis’ latest book, co-authored with Dr. Francis J. DiVesta and published in 2011 by Rowman & Littlefield, is entitled Classroom Assessment in Action, and has focus on performance assessments that teachers use in every day settings. He has numerous publications in such journals as Educational and Psychological Measurement, Psychological Test Bulletin, Educational Measurement: Issues and Practice, and Journal of Psychoeducational Assessment, to name a few. He was recently Chair of the American Psychological Association's Continuing Education Committee and takes an active role in professional governance. He is a licensed psychologist in the states of Florida, California, Indiana, and Ohio, and is a Fellow of the American Psychological Association (Division 5) and the American Educational Research Association. He is a consulting editor for the APA journal Psychological Assessment. Prior to coming to UA, Dr. Shermis served as Professor and Chair of the Department of Educational Psychology at the University of Florida and was Professor and Associate Dean for Research and Grants the Department of Educational and Psychological Studies at Florida International University.

Dr. Shermis serves as the subject matter expert in the Hewlett Foundation Assessment Prize competition for automated scoring.
Hewlett Foundation Essay Scoring Prize

In this session the Assessment Development Committee (ADC) will receive a briefing on the recently announced Hewlett Foundation prize for the automated scoring of student essays.

Each year NAEP administers a huge number of test items that are commonly known as “short” and “extended” constructed response questions. Student responses to these questions may be in the form of a sentence or two, an extended answer, an essay, or a graphic solution to a math or science problem with a written justification, among other types of responses. The NAEP constructed response questions measure an important range of knowledge and skills in multiple subject areas across all grades that NAEP assesses. In years when NAEP administers state and urban district assessments in reading and mathematics, the pool of student responses to be scored numbers in the millions. All of these responses are currently scored by trained raters employed by an NCES contractor.

Automated essay scoring software has been in development since the 1960’s. While a few large-scale testing programs use automated essay scoring (such as the Graduate Management Admission Test or GMAT), the practice is rarely used in the K-12 assessment arena.

The purpose of this session is to brief the ADC on the purpose, design, and status of the Hewlett Foundation essay scoring competition. Both of the Common Core assessment consortia (Smarter Balanced and PARCC) are collaborating on this project, which will likely have significant implications for their assessments. The outcome of the Hewlett competition will have an impact on the future of NAEP as well.

Dr. Mark Shermis of The University of Akron, who serves as the subject matter expert for the Hewlett project, will brief the ADC via phone at the Committee meeting on March 2, 2012. Briefing materials in this tab include a short bio for Dr. Shermis along with information about the purpose, design, and timelines for the competition.
<table>
<thead>
<tr>
<th>Review Package to Board</th>
<th>Board Comments to NCES</th>
<th>Background/ Cognitive</th>
<th>Review Task</th>
<th>Approx Number Items</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 16</td>
<td>October 5</td>
<td>Cognitive</td>
<td>2015 Pilot Reading Passages</td>
<td>25 passages</td>
<td>✓</td>
</tr>
<tr>
<td>November 18</td>
<td>December 8</td>
<td>Cognitive</td>
<td>2014 Technology &amp; Engineering Literacy (TEL) (8)</td>
<td>8 task sketches 5 task demos</td>
<td>✓</td>
</tr>
<tr>
<td>November 14</td>
<td>December 9</td>
<td>Background</td>
<td>2014 Technology &amp; Engineering Literacy (TEL) (8)</td>
<td>60 items (275 with all subitems)</td>
<td>✓</td>
</tr>
<tr>
<td>November 18</td>
<td>January 3</td>
<td>Background</td>
<td>2013 Operational Mathematics (4, 8)</td>
<td>190 items (6 blocks)</td>
<td>✓</td>
</tr>
<tr>
<td>November 18</td>
<td>January 3</td>
<td>Background</td>
<td>2013 Operational Reading (4, 8)</td>
<td>230 items (6 blocks)</td>
<td>✓</td>
</tr>
<tr>
<td>February 10</td>
<td>March 8</td>
<td>Cognitive</td>
<td>2013 Operational Reading (4, 8)</td>
<td>60 items (6 blocks)</td>
<td></td>
</tr>
<tr>
<td>February 10</td>
<td>March 8</td>
<td>Cognitive</td>
<td>2013 Operational Mathematics (4, 8)</td>
<td>90 items (6 blocks)</td>
<td></td>
</tr>
<tr>
<td>February 16</td>
<td>March 8</td>
<td>Cognitive</td>
<td>2014 Pilot Civics (4, 8, 12)</td>
<td>300 items (16 blocks)</td>
<td></td>
</tr>
<tr>
<td>February 29</td>
<td>March 8</td>
<td>Cognitive</td>
<td>2014 Technology &amp; Engineering Literacy (TEL) (8)</td>
<td>10 task demos</td>
<td></td>
</tr>
<tr>
<td>February 16</td>
<td>March 8</td>
<td>Background</td>
<td>2014 Pilots US History, Geography, &amp; Civics (4,8,12); 2015 Pilots Mathematics &amp; Reading (4,8)</td>
<td>1000 items (24 blocks)</td>
<td></td>
</tr>
<tr>
<td>April 12</td>
<td>April 25</td>
<td>Background</td>
<td>2013 Operational Reading &amp; Mathematics (12)</td>
<td>215 items (4 blocks)</td>
<td></td>
</tr>
<tr>
<td>April 25</td>
<td>May 8</td>
<td>Background</td>
<td>2014 Technology &amp; Engineering Literacy (TEL) (8)</td>
<td>&lt;60 items (275 with all subitems)</td>
<td></td>
</tr>
<tr>
<td>Review Package to Board</td>
<td>Board Comments to NCES</td>
<td>Background/ Cognitive</td>
<td>Review Task</td>
<td>Approx Number Items</td>
<td>Status</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>--------</td>
</tr>
<tr>
<td>May 3</td>
<td>May 24</td>
<td>Cognitive</td>
<td>2015 Pilot Mathematics (4, 8)</td>
<td>200 items (12 blocks)</td>
<td></td>
</tr>
<tr>
<td>May 3</td>
<td>May 24</td>
<td>Cognitive</td>
<td>2014 Pilot Geography (4, 8, 12)</td>
<td>225 items (blocks)</td>
<td></td>
</tr>
<tr>
<td>May 3</td>
<td>May 24</td>
<td>Cognitive</td>
<td>2014 Pilot US History (4, 8, 12)</td>
<td>340 items (blocks)</td>
<td></td>
</tr>
<tr>
<td>May 3</td>
<td>May 24</td>
<td>Cognitive</td>
<td>2014 Technology &amp; Engineering Literacy (TEL) (8)</td>
<td>21 tasks (pre-clearance)</td>
<td></td>
</tr>
<tr>
<td>July 5</td>
<td>July 25</td>
<td>Cognitive</td>
<td>2015 Pilot Reading (4, 8)</td>
<td>150 items (10 blocks)</td>
<td></td>
</tr>
<tr>
<td>July 19</td>
<td>August 8</td>
<td>Cognitive</td>
<td>2013 Operational Mathematics (12)</td>
<td>54 items (4 blocks)</td>
<td></td>
</tr>
<tr>
<td>July 19</td>
<td>August 8</td>
<td>Cognitive</td>
<td>2013 Operational Reading (12)</td>
<td>30 items (3 blocks)</td>
<td></td>
</tr>
<tr>
<td>July 19</td>
<td>August 9</td>
<td>Cognitive</td>
<td>2014 Technology and Engineering Literacy (TEL) (8)</td>
<td>21 Tasks, 175 items</td>
<td></td>
</tr>
</tbody>
</table>
Ad Hoc Committee on NAEP Parent Engagement

Final Draft Report:
Reaching Parents with NAEP Resources

March 2, 2012

Committee Members
Tonya Miles, Chair
Louis M. Fabrizio
Shannon Garrison
Doris R. Hicks
Hector Ibarra
Henry Kranendonk
Warren T. Smith
Blair Taylor

Staff
Ray Fields
Ad Hoc Committee on NAEP Parent Engagement

Overview of Recommendations

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

Committee Activity Timeline

November 2010  Recognize Need to Address NAEP Parent Engagement
March 2011    Approve Mission Statement and Establish Ad Hoc Committee on NAEP Parent Engagement
April 2011    First Ad Hoc Committee Teleconference
May 2011      First Committee Meeting
August 2011   Second Committee Meeting
October 2011  Second Teleconference
December 2011 Third Committee Meeting
February 2012 Third Teleconference
March 2012    Final Committee Meeting; Present Recommendations to the Board
Foreword

The National Assessment Governing Board, in overseeing the National Assessment of Educational Progress (NAEP or the Nation’s Report Card), is carrying out an initiative to raise public awareness about the status of student achievement in the United States.

The Governing Board believes that the low levels of student achievement and the persistent, large achievement gaps between student demographic subgroups are cause for alarm—for individuals, for families, for communities, and for the nation’s future.

Although the release of NAEP reports brings periodic public attention to this problem, this attention is not sustained for very long.

Consequently, the Governing Board is implementing an initiative to convey the urgency of improving achievement for all students and of closing achievement gaps between student subgroups by race, ethnicity and income levels, using NAEP data and resources.1

One part of this initiative is aimed at reaching parents. In March 2011, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement, composed of Board members. The Ad Hoc Committee’s assignment was to study ways to reach parents with NAEP data and resources and to present the Committee’s recommendations to the Governing Board by March 2012.

The members of the Ad Hoc Committee have worked diligently over the past year and are pleased to present our report and recommendations on the following pages.

We also would like to express appreciation for the important contributions of the National Center for Education Statistics in supporting the Ad Hoc Committee’s work and in embracing the objective of reaching parents with NAEP data and resources.

Tonya Miles
Chair
Ad Hoc Committee on NAEP Parent Engagement

---

1 The authority for this initiative is found under the Governing Board’s duties in the NAEP authorizing legislation, Public Law 107-279. Specifically, Section 302(e)(1) authorizes the Board to “take appropriate actions needed to improve the form, content, use, and reporting of results” and “plan and execute the initial public release of National Assessment of Educational Progress reports.”
Introduction
The National Assessment Governing Board, recognizing that NAEP report releases were not conveying a sense of urgency, began an initiative in May 2010 to see what the Board could do to “make a difference” in fostering concern and action about the need to improve achievement and reduce achievement gaps, using NAEP data and resources. Toward this goal, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement. The Committee’s task was to develop recommendations on ways to reach parents with NAEP information. The purpose of this report is to document the work of the Ad Hoc Committee and present its recommendations.

Background
U.S. Secretary of Education Arne Duncan addressed the Governing Board on November 19, 2010. He focused on the urgent need to improve student achievement and reduce achievement gaps among student subgroups. He has said publicly that “our nation will pay the price socially and economically” if we fail to act with determination and dispatch and stressed to the Board that “we have to continue to awaken our country to the huge consequences” of inaction.

Secretary Duncan emphasized the important role of parents in improving student achievement. He told the story of President Obama meeting with the President of South Korea, Lee Myung-bak. President Obama asked him about education issues in South Korea. President Lee said his biggest challenge is that parents in South Korea are very assertive in demanding a good education from their schools and great effort from their children. He emphasized that this includes parents of all income levels.

Implicit in this story is the fact that South Korean students, as well as others in the world, outperform U.S. students in mathematics and science on TIMSS.2 Today’s students are tomorrow’s workers and leaders. It follows that failing to improve U.S. student achievement could have disastrous effects on the nation’s future work force and global competitiveness, and that parents have an important role to play in promoting improved student achievement.

Secretary Duncan continued by saying “I wish my biggest problem, my biggest challenge, was parents knocking down my door saying, ‘Get better faster.’” He said that there are good examples in the U.S. of parent initiatives that impact student achievement. But Secretary Duncan wanted to “scale up” parent engagement programs that “are really showing the ability to drive student achievement.”

---

2 TIMSS is the Trends in International Mathematics and Science Study.
The Secretary’s remarks and the Board’s initiative to make a difference served as the backdrop to Board member Tonya Miles asking what can the Board do to make NAEP data available to parents and guardians³ about student achievement, especially about the urgency of addressing achievement gaps by race, ethnicity, and income levels.

The question—“what can the Board do?”—is pertinent and important. Parents have a significant stake in the quality of their local schools and, most immediately, in their own children’s achievement.

Board Chair David Driscoll recognized the opportunity and value of reaching parents with NAEP data. Therefore, at the conclusion of the November 2010 Governing Board meeting, he asked Ms. Miles, and she agreed, to lead a Board initiative to increase parent awareness about and access to NAEP data. The goal was to bring attention to the unacceptably low levels of student achievement in the U.S. and the disgraceful size of the achievement gaps.

The Mission
At the March 2011 Board meeting, the Executive Committee approved the mission statement for and established the Ad Hoc Committee on NAEP Parent Engagement (Appendix A). The Committee would be composed of Board members and chaired by Ms. Miles. The Ad Hoc Committee’s task was to present recommendations to the Governing Board by March 2012. The recommendations would describe steps and strategies the Governing Board and representatives of the NAEP program can take directly, and/or support the efforts of others to increase parent awareness about the urgency to improve the levels of student achievement in the U.S. and the urgency to reduce the size of achievement gaps by race, ethnicity, and income levels, using NAEP data and resources.

The mission statement indicated that the recommendations were to be clear about the limits on the Board’s role under the law. This was to ensure the Committee considered all appropriate options without exceeding the Board’s authority.

The recommendations were to help reach parents in feasible, innovative, and meaningful ways, across all income levels, and whether residing in urban, rural, or suburban areas. Finally, the recommendations were to include strategies to make NAEP parent engagement an ongoing part of the work of the Board and the NAEP program.

³ The term “parents” as used throughout this report is intended to refer to parents and guardians of school children.
Committee Activities
Members of the Ad Hoc Committee have met four times during the May 2011, August 2011, December 2011, and March 2012 Board meetings. The agendas for these meetings are in Appendix B.

The Ad Hoc Committee also has met three times in between Board meetings via conference calls:
- April 15, 2011 - reviewed the Committee’s mission statement and a timeline for completing their work (Appendix C)
- October 12, 2011 - focused on formulating the Committee’s preliminary recommendations for discussion at the December 2011 Board meeting (Appendix D)
- February 8, 2012 - reviewed the Committee’s initial draft report

In addition, Chair Tonya Miles and Ray Fields conducted meetings with leaders of three nationally recognized parent-related organizations. The purpose was to brief them on the Board’s initiative to reach parents with NAEP data, to receive their input and feedback, and to determine their interest in supporting this initiative. The three organizations are the National PTA, the Public Education Network, and the Center on School, Family, and Community Partnerships at Johns Hopkins University.

In connection with the August 2011 meeting in Washington, D.C., the Board conducted an outreach event with parent leaders and national and local parent organizations. The discussion with meeting participants, led by Ms. Miles and Chair Driscoll, resulted in valuable feedback and input on the Board’s parent initiative. A summary of the discussion at this parent outreach meeting is in Appendix E.

Concluding Comment
Parents are the primary advocates for the quality of their children’s education. Having solid information about education achievement improves their ability to advocate. NAEP can be one potentially valuable source of such information. Therefore, it is appropriate to seek ways to reach parents with NAEP data and resources. Some progress already has been made, in a small way, as will be seen in the activities and relationships described below. The recommendations that follow are offered as a set of feasible next steps, all within the Governing Board’s authority. All have the potential to reach parents in meaningful ways.
Recommendations

I. Specify the Target Audience: National, State, and Local Parent Leaders and Parent Organizations

The target audience needs to be defined. Approximately 55 million students are enrolled in public and private K-12 schools in the U.S. It is not feasible to reach the parents of all these children with NAEP data, nor is it within NAEP or the Governing Board’s scope to do so.

Further, the achievement of their own children is the most pressing and immediate interest of parents. Because NAEP does not provide individual student results, this interest of parents is not served by NAEP.

A unique aspect of NAEP is its ability to report patterns of overall and subgroup student performance within and across education systems. These patterns may reflect education system strengths and weaknesses that can affect the achievement of individual students. The NAEP data for the states and 21 urban districts4 provide ample evidence of differences in achievement across comparable groups at points in time and differences in gains in achievement over time. The NAEP data also document persistent and unacceptable achievement gaps between groups. This NAEP information does have potential interest for parents.

Also of potential interest to parents is how their education systems compare internationally. The linking studies the Board has endorsed, beginning in 2011, between NAEP and the international assessments (TIMSS and PIRLS5) will compare student achievement in the U.S. with achievement in other nations.

State and local education policymakers use NAEP data to ask fundamental questions about the levels of student achievement in schools under their authority. For example, Tennessee Commissioner of Education Kevin Huffman said at a November 2011 meeting on NAEP 12th in Nashville, that he analyzes student subgroup results in his state (e.g., students on free and reduced lunch) in comparison to other states. Asking questions about how subgroup performance compares across jurisdictions can help highlight where state or local policies may or may not be working. State and local policymakers as well as parents can ask such questions using NAEP results.

4 The 21 participants in the NAEP Trial Urban District Assessment Program are: Albuquerque, Atlanta, Austin, Baltimore City, Boston, Charlotte, Chicago, Cleveland, Dallas, Detroit, Fresno, Hillsborough County, Houston, Los Angeles, Louisville, Ky. (Jefferson County), Miami (Dade County), Milwaukee, New York City, Philadelphia, San Diego, and Washington, DC.
5 PIRLS is the Progress in International Reading Literacy Study.
The Ad Hoc Committee believes there are groups of active parents and parent organizations who see the connection between system performance and the potential for impact on individual students. These include local and state leaders, often members of recognized parent and community organizations, who regularly work with the leaders of education systems, examine data, and ask fundamental questions to support and foster improved achievement and the closing of achievement gaps. These parent leaders and parent organizations should be the initial target audience for NAEP data and resources.

More specifically, because NAEP provides data for each of the 50 states and 21 urban districts, the initial target audience should be state and local parent leaders and parent organizations associated with these jurisdictions.

2. Establish Relationships with Recognized Parent and Community-based Organizations

To reach the target audience with NAEP data, it is important to work collaboratively with existing parent and community-based organizations. Many of these organizations have state affiliates and/or affiliates associated with local school districts. These organizations have direct access to parent and community leaders through email networks, social media, newsletters, and websites. These mechanisms are potentially effective, viable avenues for the dissemination of NAEP data and resources. In addition, these organizations often conduct national and state conferences, which could afford opportunities for presenting NAEP data and resources.

The Ad Hoc Committee has initiated conversations with the National PTA (NPTA), with positive results (see Appendix F). For example, the NPTA has begun announcing NAEP release events through its email networks and social media. In addition, Tonya Miles has been invited to make a presentation on March 7, 2012 at the NPTA Legislative Conference and on June 21, 2012 at the NPTA Annual Conference. Further, the NPTA assisted in recruiting parents for a meeting on February 16, 2012 to help review NAEP materials for parents.

Likewise, collaborative activity has occurred with the Public Education Network (PEN). Cornelia Orr, Governing Board Executive Director, made a presentation on NAEP and 12th grade academic preparedness at the PEN annual conference in November 2011. PEN also helped recruit experts from among its member organizations for a one-day meeting held on February 14, 2012 to provide input and feedback on NAEP materials for parents. PEN already transmits information about NAEP data and NAEP releases to its members and newsletter subscribers.

The Governing Board should continue to develop the relationships with the NPTA and PEN, and develop similar collaborative relationships with other organizations.
3. **Develop Presentations and Materials Targeted to Parents for Use by Board Members and Others**

Recognizing that the scope and depth of NAEP data and resources can be overwhelming, the Governing Board is working to develop a model PowerPoint presentation and associated materials for parents. Consistent with the information needs of the target audience in Recommendation 1, the presentation and materials can be customized for particular states and urban districts. The materials will include easy to understand charts and graphs and avoid the use of technical terms and jargon. In addition to explaining what NAEP is, the presentation will include NAEP data highlighting the levels of achievement and the gaps between subgroups in a way that conveys urgency. The intent is for these resources to be available for use by Governing Board members in making presentations to the public and by parent and community-based organizations in making presentations specific to their locale.

The National Center for Education Statistics (NCES) is currently developing a general publication for parents. This publication will inform parents about what NAEP is, how it fits into the education landscape, and options to learn or do more. This publication will be debuted at the NPTA conference in June and displayed at the NAEP booth at the conference.

4. **Develop Parent Pages on the Governing Board and NAEP Websites**

Currently, the Governing Board website has no pages aimed at parents as the target audience. The NAEP website, managed by the National Center for Education Statistics (NCES) does have pages for parents whose child has been selected to take NAEP, but not for parents in general.

The Ad Hoc Committee invited NCES to examine what it can do to make NAEP information on the website more accessible to parents generally. As a first important step, NCES made the “parent” navigation button more prominent on the NAEP website landing page. NCES is exploring additional changes to make the NAEP data more accessible to parents. As they develop the parent publication mentioned in Recommendation 3, NCES will update the NAEP web pages to reflect consistent messages. This will expand the NAEP website audience from parents of selected students to all interested parents.

The Ad Hoc Committee asked the Board’s communications and website contractors, Reingold, Inc. and Quotient, to develop page mockups for parent pages on the Governing Board website (Appendix G). These should be further developed and incorporated as components of the Governing Board’s website redesign, which is currently underway. In addition, the model PowerPoint presentation and materials in Recommendation 3 should be available for easy downloading from the Governing Board website.
5. **Conduct a Parent Education Summit in Late Summer/Early Fall 2012**

The Ad Hoc Committee proposes a one-day parent summit on education for the late summer or early fall of 2012. The summit would be conducted in Washington, D.C. and available across the nation via live-streaming internet video, with the potential for live TV and radio coverage via C-SPAN.

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

In addition to Governing Board members, the audience of 150-300 would consist primarily of parent and community leaders, parent organizations, and leaders in education, business, civil rights, the religious community, and legislative policy.

To help convey the non-partisan, universal interest in achieving the summit objective, as well as to focus on its importance for the nation’s future, First Lady Michelle Obama and former First Lady Laura Bush would be invited to share the podium in delivering the keynote address.

A distinguished journalist or media representative, acknowledged for intellect and freedom from bias, would be invited to moderate and provide a concluding summary.

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

Individual and panel presentations would be made to address the national imperative for achieving the summit objective, from a wide range of perspectives which, taken together, would provide a compelling, unassailable argument for the urgent need to take action.

For example (not listed in priority order):

- Religious leaders would provide the moral perspective
- Economists would provide the national economic perspective
- Civil rights leaders would provide the equity perspective
- Military leaders would address the national security imperative
- Business leaders would address the human capital and employment imperative
• Scholars from nationally recognized policy institutions and foundations, representing a diverse range of philosophical perspectives, would provide societal perspectives

• Demographers would address the implications from the perspective of a changing population

• Parent leaders would address the imperative for families and students

• Educators would describe actions that are needed to improve achievement and close achievement gaps
# National Assessment Governing Board
## Executive Committee
### March 1, 2012

## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Presenter(s)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:30 pm</td>
<td>Welcome, Introductions, and Agenda Overview</td>
<td>David Driscoll, Chair</td>
<td></td>
</tr>
<tr>
<td>4:35 pm</td>
<td>Committee Issues and Challenges – March 2012 Meeting</td>
<td>Committee Chairs</td>
<td></td>
</tr>
<tr>
<td>4:45 pm</td>
<td>Updating Board Policy Statement: “Redesigning the National Assessment of Educational Progress”</td>
<td>David Driscoll</td>
<td>Attachment A</td>
</tr>
<tr>
<td></td>
<td>[Plan for Future Discussion on Cross-Cutting Issues]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potential Topics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• NAEP Reporting on 12th Grade Academic Preparedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Comparing NAEP and PISA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Comparing NAEP and Common Core Assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The Future of NAEP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:50 pm</td>
<td>Ad Hoc Committee on NAEP Parent Engagement</td>
<td>Tonya Miles, Ad Hoc Committee Chair</td>
<td>See Ad Hoc Committee Tab</td>
</tr>
<tr>
<td>5:00 – 6:00 pm</td>
<td>NAEP Budget, Assessment Schedule, and Contract Cost Estimates: FY 2013 and Beyond</td>
<td>Cornelia Orr, Executive Director Peggy Carr, Associate Commissioner, NCES</td>
<td>Attachment B</td>
</tr>
<tr>
<td>6:00 pm</td>
<td>Adjourn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Updating the Governing Board Policy Statement
“Redesigning the National Assessment of Educational Progress”

Following the November 1994 Board meeting, then Chair William Randall established a work group on strategic planning for NAEP. The work group was composed of Board members, chaired by Mark Musick, and staffed by Ray Fields.

The need for strategic planning was prompted by several factors. The first trial state assessment—in mathematics at grade 8—had been conducted in 1990, with successive trials in 1992 and 1994 in reading and mathematics at grades 4 and 8, either by single subject and grade or in different subject/grade combinations. There was a degree of unpredictability to the schedule of state assessments, due in large part to the evolving nature of the legislative authorizations for state assessments and a changing budget outlook for NAEP. This led, at times, to late notice to states about changes to the schedule for an imminent assessment. States saw the value in state NAEP and were volunteering, but were frustrated by the absence of a dependable schedule available several years in advance of the assessments to permit orderly in-state planning.

In addition, the Governing Board, almost from its inception, had expressed concern to NCES about the excessive period of time from the end of NAEP testing to the release of assessment results, frequently as long as two years. The Board also was concerned about the fact that the format and content of NAEP reporting was aimed more at researchers than the general public, educators, and policymakers.

While NAEP had been conducted annually during its first decade in the 1970s, the schedule changed to a biannual basis during the 1980s and well into the 1990s. As a consequence, subject coverage and frequency were limited. The Governing Board embarked on the strategic planning process to identify strategies that could lead to efficiencies in assessment and greater subject coverage.

Through an extensive process of structured full-Board deliberation, consultation with NAEP stakeholders and experts, consultation with NCES, and public comment, the work group prepared the policy statement on “Redesigning the National Assessment of Educational Progress.” In August 1996, the Governing Board adopted the policy statement that appears on the following pages.

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

However, the educational environment NAEP is to serve has changed substantially since 1996. Executive Committee members have proposed taking a fresh look at this policy statement and revising or replacing it. The Executive Committee discussion will afford the opportunity to consider this proposal. There also will be a brief discussion of potential topics for future Executive Committee consideration.
Foreword
This policy statement was adopted in 1996, at a time when Congress had codified National Education Goals targeted for accomplishment by the year 2000. It was the expectation that the National Assessment of Educational Progress would be a primary means for monitoring progress in achieving the goal addressing student achievement and this expectation is reflected in the policy below. The National Education Goals legislation is no longer in effect and has been superceded by other national policies, the No Child Left Behind Act of 2001 (NCLB) being the most germane. Therefore, the references to National Education Goals in this policy statement are no longer relevant.

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

A Better Way to Measure Educational Progress in America

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

But is the education provided in your state and in America good enough? How do our 12th graders compare with students in other nations in mathematics and science? Do our 8th grade students have an adequate understanding of the workings of our constitutional democracy? How well do our 4th grade students read, write, and compute? The National Assessment of Educational Progress is the only way for the public to know with accuracy how American students are achieving nationally and state-by-state.
The National Assessment tests at grades 4, 8, and 12. By law, it covers ten subjects, including reading, writing, mathematics, and science. The National Assessment has performance standards that indicate whether student achievement is "good enough." The National Assessment is not a national exam taken by all students. In fact, only several thousand students are tested per grade, comprising carefully drawn samples that represent the nation and the participating states. Since its first test in 1969, the National Assessment has earned a trusted reputation for its quality and credibility. That reputation must be maintained.

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

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

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

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

The current National Assessment design is overburdened, inefficient, and redundant. It is unable to provide the frequent, timely reports on student achievement the American public needs. The challenge is to supply more information, more quickly, with the funding available.
To meet this challenge, the National Assessment design must be changed, building on its strengths while making it more efficient. The design of the National Assessment must be simplified. The purpose of the National Assessment must be sharply focused and its principal audience clearly defined. Because the National Assessment cannot do all that some would have it do, trade-offs must be made among desirable activities. Useful but less important activities may have to be reduced, eliminated, or carried out by others. The National Assessment must "stick to its knitting" in order to be more cost-effective, reach more of the public, provide more information more promptly, and maintain its integrity.

National Assessment Redesign

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

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

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

Purpose and Objectives of the National Assessment of Educational Progress

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

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

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

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

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

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

The Audience for the National Assessment

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

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

Limitations: What the National Assessment Is Not

The National Assessment is intended to describe how well students are performing, but not to explain why. The National Assessment only provides group results; it is not an individual student test. The National Assessment tests academic subjects and does not collect information on individual students’ personal values or attitudes. Each National Assessment test is developed through a national consensus process. This national consensus process takes into account education practices, the results of education research, and changes in the curricula. However, the National Assessment is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques. Nor is the National Assessment an appropriate means, by itself, for improving instruction in individual classrooms, evaluating the effects of specific teaching practices, or determining whether particular approaches to curricula are working.
OBJECTIVE 1: To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations.

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

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

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

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

Provide National Assessment results for states

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

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

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

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

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

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

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

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

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

Vary the amount of detail in testing and in reporting results

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

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

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

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

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

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

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

Simplify the National Assessment design

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

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

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

Simplify the way the National Assessment reports trends in student achievement

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

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

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

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

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

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

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

In 1988, Congress created a non-partisan citizen's group—the National Assessment Governing Board—and authorized it to set explicit performance standards, called achievement levels, for reporting National Assessment results.
The achievement levels describe "how good is good enough" on the various tests that make up the National Assessment. Previously, it might have been reported that the average mathematics score of 4th graders went up (or down) four points on a five-hundred-point scale. There was no way of knowing whether the previous score represented strong or weak performance and whether the amount of change should give cause for concern or celebration. In contrast, the National Assessment now also reports the percentage of students who are performing at or above "basic," "proficient," and "advanced" levels of achievement. Proficient, the central level, represents "competency over challenging subject matter," as demonstrated by how well students perform on the questions on each National Assessment test. Basic denotes partial mastery and advanced signifies superior performance on the National Assessment. Using achievement levels to report results and track changes allows readers to make judgments about whether performance is adequate, whether "progress" is sufficient, and how the National Assessment standards and results compare to those of other tests, such as state and local tests.

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

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

**Use international comparisons**

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

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

Emphasize reporting for grades 4, 8, and 12

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

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

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

• The National Assessment shall continue to test in and report results for grades 4, 8, and 12; however, in selected subjects, one or more of these grades may not be tested.
• Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for
long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.

- Grade 12 results shall be accompanied by clear, highlighted statements about school and student participation, student motivation, and cautions, where appropriate, about interpreting 12th grade achievement results.
- The National Assessment design shall seek to improve school and student participation rates and student motivation at grade 12.
- The National Assessment shall provide practical incentives for school and district participation at grades 4, 8, and 12.

Use innovations in measurement and reporting

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

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

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

- The National Assessment shall assess the merits of advances related to technology and the measurement and reporting of student achievement.
Where warranted, the National Assessment shall implement such advances in order to reduce costs and/or improve test administration, measurement, and reporting.

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

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

**Keep test frameworks and specifications stable**

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

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

The process of developing frameworks and specifications involves consideration of both current classroom teaching practices and important developments in each subject area for inclusion in the National Assessment. In order to ensure that National Assessment data fairly represent student achievement, the test frameworks and specifications are subjected to wide public review before adoption and test questions developed for the National Assessment are reviewed for relevance and quality by representatives from participating states.
An important role of the National Assessment is to report on trends in student achievement over time. For the National Assessment to be able to measure trends, the frameworks (and hence the tests) must remain stable. However, as new knowledge is gained in subject areas and as teaching practices change and evolve, pressures arise to change the test frameworks and tests to keep them current. But, if frameworks, specifications, and tests change too frequently, trends may be lost, costs go up, and reporting time may increase.

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

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

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

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

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

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

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

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

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

Examples of areas in which particular interest has been expressed for using the National Assessment include linking state and local tests with the National Assessment and performing in-depth analysis on National Assessment data. States that link their tests to the National Assessment would have an unbiased external benchmark to help make judgments about their own tests and standards and also would have a means for comparing their tests and standards with those of other states.
The National Assessment shall develop policies, practices, and procedures that assist states, school districts, and others who want to do so at their own cost to link their test results to the National Assessment.

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

The pages that follow immediately are provided as background for the discussion of the NAEP budget and schedule of assessments. They were excerpted from the Department of Education budget justification for FY 2013. The last two pages in this attachment are the current NAEP schedule and the long-term schedule discussion draft, previously shared with the Executive Committee.

A decrease of $5 million is requested for NAEP in FY 2013 and a decrease of $1 million is requested for the Governing Board.

If approved by Congress, this would set new baselines for the NAEP and Governing Board budgets, with serious effects not just in FY 2013, but in the out-years as well. For example, over the five-year period of the new NAEP contracts planned for award in September 2012, the impact would be a $25 million reduction in funds available for assessing student achievement.

The first page of the excerpt (numbered X-5) provides an overview of the President’s request for the Institute of Education Sciences, showing where there are requested increases and decreases. The remaining pages provide details related to the request for NAEP and the Governing Board.

Your attention is directed to page X-49 and, especially, the highlighted text on page X-50.

The complete Department of Education FY 2013 budget justification can be found at http://www2.ed.gov/about/overview/budget/budget13/justifications/index.html

More details about the implications for the budget and NAEP schedule of assessments will be provided at the Governing Board meeting.
## Summary of Changes
(dollars in thousands)

<table>
<thead>
<tr>
<th>Year</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>$593,664</td>
</tr>
<tr>
<td>2013</td>
<td>$621,150</td>
</tr>
<tr>
<td>Net change</td>
<td>+27,486</td>
</tr>
</tbody>
</table>

### Increases

<table>
<thead>
<tr>
<th>Program</th>
<th>2012 base</th>
<th>Change from base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase for Research, Development, and Dissemination to conduct additional research, development, dissemination, and evaluation activities.</td>
<td>$189,787</td>
<td>+$12,486</td>
</tr>
<tr>
<td>Increase for Statistics to allow NCES to collect State-level Program for International Student Assessment data.</td>
<td>108,748</td>
<td>+6,000</td>
</tr>
<tr>
<td>Increase for Statewide Data Systems to provide funding to support postsecondary data initiatives designed to improve information on students as they progress from high school to postsecondary education and the workforce.</td>
<td>38,077</td>
<td>+15,000</td>
</tr>
</tbody>
</table>

Subtotal, increases: +33,486

### Decreases

<table>
<thead>
<tr>
<th>Program</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease for National Assessment to reflect need to focus the program on key assessments</td>
<td>129,616</td>
</tr>
<tr>
<td>Decrease for National Assessment Governing Board to reflect need to focus on key activities</td>
<td>8,690</td>
</tr>
</tbody>
</table>

Subtotal, decreases: -6,000

Net change: +27,486
Assessment
(National Assessment of Educational Progress Authorization Act)
(dollars in thousands)
FY 2013 Authorization: 0

Budget Authority:

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Assessment of Educational Progress</td>
<td>$129,616</td>
<td>$124,616</td>
<td>-$5,000</td>
</tr>
<tr>
<td>National Assessment Governing Board</td>
<td>8,690</td>
<td>7,690</td>
<td>-1,000</td>
</tr>
<tr>
<td>Total</td>
<td>138,306</td>
<td>132,306</td>
<td>-6,000</td>
</tr>
</tbody>
</table>

1 The GEPA extension expired September 30, 2009; the Administration proposes to continue funding in FY 2013 under appropriations language.

PROGRAM DESCRIPTION

The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what American students know and can do. Also known as The Nation’s Report Card, NAEP collects and analyzes data on, measures, and reports on the status of and trends in student learning over time, subject-by-subject. By making objective information on student performance available to policymakers, educators, parents, and others, NAEP has become an integral part of the Nation’s measurement of educational progress.

Assessment frequency is specified in the authorizing statute. The Commissioner for Education Statistics must conduct:

- National reading and mathematics assessments in public and private schools at grades 4 and 8 at least once every 2 years;
- National grade 12 reading and mathematics assessments in public and private schools on a regular schedule; and
- Biennial State assessments of student achievement in reading and mathematics in grades 4 and 8.

If time and resources allow, the Commissioner may conduct additional national and State assessments in grades 4, 8, and 12 in public and private schools at regularly scheduled intervals in additional subject matters, including writing, science, history, geography, civics, economics, foreign languages, and arts; may conduct grade 12 State reading and mathematics assessments; and may conduct long-term trend assessments of academic achievement at ages 9, 13, and 17 in reading and mathematics. Whenever feasible, information must be collected
Assessment

and reported by race, ethnicity, socioeconomic status, gender, disability, and limited-English proficiency. The NAEP schedule is publicly available at http://www/nagb.org/.

The National Assessment Governing Board (NAGB) is responsible for formulating policy for NAEP. NAGB is composed of 25 voting members including Governors, State legislators, chief State school officers, a superintendent, State and local board of education members, testing and measurement experts, a representative of business or industry, curriculum specialists, principals, classroom teachers, and parents. The Director of the Institute of Education Sciences serves as an ex officio, nonvoting member of the Board. Using a national consensus approach, NAGB develops appropriate assessment objectives and achievement levels for each grade in each subject area to be assessed. The Assessment budget supports the following major program components:

- **National NAEP.** The main NAEP assessments report results for the Nation and are designed to follow the curriculum frameworks developed by NAGB. They periodically measure student achievement in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects;

- **Grade 4 and 8 State NAEP.** State assessments address the needs of State-level policymakers for reliable data concerning student achievement in their States in reading, mathematics, science, and writing. In 2002, the Department began paying for State participation in biennial reading and mathematics assessments in grades 4 and 8. Periodic assessments also are administered in science and writing;

- **Grade 4 and 8 Trial Urban District Assessment (TUDA).** Begun in 2002, the TUDA provides information on student achievement in a small number of urban school districts. Participation is voluntary;

- **Long-term NAEP.** In its long-term trend program, NAEP administers identical instruments from one assessment year to the next, measuring student achievement in reading and mathematics. These assessments do not evolve based on changes in curricular or educational practices; and

- **Evaluation and validation studies.** Congress mandates that the Secretary provide for continuing review of the national and State assessments and student performance levels by one or more nationally recognized evaluation organizations. NAEP funds also support studies to examine critical validity issues involving NAEP design, interpretation, and operations.

In order to inform the American public about the performance of the Nation's students, NAEP produces a series of public audience and technical reports. All NAEP reports are available through the Internet (http://nces.ed.gov/nationsreportcard/). In addition, an online data tool (http://nces.ed.gov/nationsreportcard/naepdata/) allows users to create their own data tables with national and State data.

The statute requires biennial State assessments in reading and mathematics in grades 4 and 8 and requires reporting NAEP results, where feasible, by disability and limited-English proficiency as well as by race, ethnicity, socioeconomic status, and gender. The Federal Government is
specifically prohibited from using NAEP to influence standards, assessments, curriculum, or instructional practices at the State and local levels, or from using NAEP to evaluate individual students or teachers or provide rewards or sanctions for individual students, teachers, schools, or school districts. In addition, the statute specifies that nothing in the law shall be construed to prescribe the use of NAEP for student promotion or graduation purposes, and that NAEP should not affect home schools. Maintenance of a system of records containing personally identifiable information on students is prohibited, and assessments must not evaluate or assess personal or family beliefs or attitudes.

The statute ensures the Department’s ability to maintain test integrity by allowing the Statistics Commissioner to decline to release cognitive test items that will be used in future assessments for 10 years (and longer if important to protect long-term trend data) while continuing to provide for public access to assessment materials in secure settings. The statute requires that the public be notified about such access; requires that access be provided within 45 days in a mutually convenient setting; and establishes procedures for receiving, reviewing, and reporting complaints. The law provides criminal penalties for unauthorized release of assessment instruments.

The statute also mandates that participation is voluntary for students and schools, as well as for local educational agencies. Each participating State must give permission for the release of the results of its State assessment. However, under Title I of ESEA, each State participating in the Title I program must develop a State plan that demonstrates, among other things, that the State has developed high quality assessments that will be used to determine student progress (ESEA, Title I, Part A, Section 1111). In addition, each State, in its plan, had to agree to participate in the biennial grades 4 and 8 reading and mathematics NAEP assessments beginning in the 2002-2003 school year, if the Secretary paid for the costs of participation. Any State with an approved plan under section 1111 is deemed to have given its permission for the release of its grades 4 and 8 reading and mathematics NAEP data.

Funding levels for both NAEP and NAGB for the past 5 fiscal years were:

<table>
<thead>
<tr>
<th>Year</th>
<th>Amount (in thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>$104,053</td>
</tr>
<tr>
<td>2009</td>
<td>138,844</td>
</tr>
<tr>
<td>2010</td>
<td>138,844</td>
</tr>
<tr>
<td>2011</td>
<td>138,566</td>
</tr>
<tr>
<td>2012</td>
<td>138,306</td>
</tr>
</tbody>
</table>

**FY 2013 BUDGET REQUEST**

The Administration requests $132.3 million for Assessment in 2013, a decrease of $6.0 million from 2012. Of this amount, $124.6 million would provide support for the National Assessment of Educational Progress (NAEP) program and $7.7 million would support the National Assessment Governing Board (NAGB). NAGB is responsible for formulating policy for NAEP and develops appropriate assessment objectives and achievement levels for each grade in each subject area to be assessed. The NAEP State-level assessments are held every other year, meaning that
Assessment

costs are considerably higher in some years and lower in others. The Administration requests that these funds remain available for 2 years, as they have been in recent years. Extending the availability of funds for an additional year allows the Department the flexibility it needs to fund the assessments. The Administration believes that the funds requested are sufficient to enable NAEP to fulfill its mission and continue to provide the critically important information needed on student achievement over time.

NAEP funds for a particular fiscal year provide support for the analysis and reporting of assessments conducted in prior fiscal years, the administration of current year assessments, and preparation for future assessments. The current plans are to use the 2013 funds for:

• Conducting the 2013 national and State reading and math assessments at grades 4, 8, and 12. State participation in 12th grade NAEP is voluntary, with 11 States participating in 2009. In addition, the 2013 assessments will once again include data for certain large urban districts. In 2009, 18 urban districts participated; 21 participated in 2011.

• Analysis and reporting of assessments conducted in prior years, including the 2012 economics and long-term trend assessments.

• Preparation for 2014 assessments. The current NAGS schedule includes assessments in U.S. history, civics, and geography, as well as a technology and engineering literacy assessment.

The requested funding for NAGS would allow it to carry out its policy-setting responsibilities for NAEP, including selecting subject areas to be assessed; developing student achievement levels for each grade and subject tested; taking appropriate actions to improve the form, content, use, and reporting of NAEP; developing test objectives and specifications for assessments in each subject; handling the initial public release of NAEP reports; ensuring that all NAEP materials are free from racial, cultural, gender, and regional bias and are secular, neutral, and non-ideological; developing and implementing procedures for the review of NAEP methodology, content, frameworks, reporting, and dissemination; and reviewing complaints about NAEP submitted by parents and other members of the public and determining whether revisions to NAEP are necessary and appropriate. NAGS also conducts special studies to inform NAEP. The 2013 request would allow NAGS to finish work on 12th grade preparedness initiated in 2011 and to complete Web-based interactive versions of the reading, mathematics, science, and writing frameworks.
<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>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>ECONOMICS (12)</td>
<td>Writing (8)</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 Mathematics*</td>
<td>READING (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>SCIENCE**</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td>SCIENCE (4, 8)</td>
</tr>
<tr>
<td>2010</td>
<td>U.S. History</td>
<td>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td>Science (8)</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>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td>2013</td>
<td>Reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>U.S. History</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td>Science (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>TECHNOLOGY AND ENGINEERING LITERACY (8)**</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Reading</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8)</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td>Math (4, 8, 12)</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.
4. The Governing Board intends to conduct assessments at the 12th grade in World History and Foreign Language during the assessment period 2018-2022.
<table>
<thead>
<tr>
<th>Year</th>
<th>National</th>
<th>State</th>
</tr>
</thead>
<tbody>
<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><strong>WRITING (8, 12)</strong></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Economics (12)</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td>2013</td>
<td>Reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>U.S. History</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td>Science (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>TECHNOLOGY AND ENGINEERING LITERACY (8)**</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Reading</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8)</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Add: Economics (12)</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td>Science (4, 8, 12)</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>
<tr>
<td>2018</td>
<td>U.S. History</td>
<td>CIVICS**</td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technology and Engineering Literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8, 12)**</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Reading</td>
<td>Reading (4, 8, 12)**</td>
</tr>
<tr>
<td></td>
<td>Mathematics**</td>
<td>Math (4, 8, 12)**</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8, 12)**</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>Long-term trend</td>
<td>NOTE: administer by computer?</td>
</tr>
<tr>
<td></td>
<td>Economics (12)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOREIGN LANGUAGE (12)**</td>
<td></td>
</tr>
<tr>
<td>2021</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>
<tr>
<td>2022</td>
<td>U.S. HISTORY**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEOGRAPHY**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WORLD HISTORY (12)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technology And Engineering Literacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4, 8, 12)**</td>
<td></td>
</tr>
</tbody>
</table>

**Assessments involving test administration by computer.

NOTES:

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

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

(3) In 2009, 12th grade assessments in reading and mathematics at the state level will be conducted as a pilot in 11 volunteering states.

(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>Session</th>
<th>Notes</th>
</tr>
</thead>
</table>
| 9:45 - 9:50 am | Welcome, Introductions, and Agenda Overview  
*Lou Fabrizio, COSDAM Chair* | Attachment A                  |
| 9:50 – 11:15 am | Update on 12th Grade Preparedness Research Program  
- Statistical Relationship Analyses: Performance of Florida Students on Preparedness Indicators and in First Year of College Relative to Performance on NAEP  
  *(Informational Item for Discussion)*  
- Research on Job Training Curriculum Comparisons to NAEP  
  *Susan Loomis, Governing Board Staff*  
| 9:50 – 10:05 am |                                                                                             | Attachment A-1               |
| 10:05 - 10:20 am |                                                                                             | Attachment A-2               |
| 10:20 - 10:50 am |                                                                                             |                              |
| 10:50 – 11:25 am |                                                                                             |                              |
| 11:25 - 11:30 am | Recommendations of Future COSDAM Agenda Topics |                              |
| **CLOSED SESSION** | Writing Achievement Levels-Setting Update  
- Revised Writing Achievement Levels Descriptions  
- Field Trial with Revised Achievement Levels Descriptions: January 27, 2012  
- Writing Achievement Levels Setting Study: February 7-11, 2011  
  *Luz Bay, Measured Progress Project Director*  
  *Susan Loomis* | Attachment B                              |
Program of Preparedness Research
Updates and Final Steps for Reporting

OVERVIEW OF PROGRAM OF PREPAREDNESS RESEARCH

Based on the Program of Preparedness Research adopted by the Board in March 2009, four categories of research studies are recommended to produce results for reporting 12th grade preparedness for grade 12 NAEP 2009 reading and mathematics. The following four categories are included.

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

Additionally, the Texas Commissioner of Higher Education offered the opportunity to conduct a benchmarking study with Texas higher education institutions, and a pilot study to examine the feasibility has been implemented. More information is provided below.

As part of the ongoing updates to COSDAM, this document includes an overview of each study type, followed by an update on new developments since the December 2011 Board meeting. A timeline is included to show only the remaining steps in the process to be completed for reporting results in 2012.

Updates for March 2012
For each type of study, there is a description of the study including background information that has already been shared with COSDAM, as well as an update including information about the current status of the research. Briefly, the updates for March 2012 are as follows:

1. Content alignment studies: completed; no updates
2. Statistical relationship studies: final briefing on analysis of data for state representative sample of 2009 NAEP examinees in Florida, including statistical relationships of NAEP performance with ACT, SAT, and ACCUPLACER and data for Florida students’ first year in college
3. Judgmental standard setting studies: staff recommendation on reporting; information about additional research about curriculum requirements for job training programs
4. Survey of higher education: 85% response rate attained; preliminary data reported to COSDAM
TIMELINES

Studies listed in the chart below are for both reading and mathematics. This timeline is provided to help focus attention on the final steps to be completed before reporting results from the preparedness studies. Many interim steps are excluded from the chart.

For the content alignment, judgmental standard setting, and higher education survey, a full federal procurement process was required which involved many steps and compliance with specified timelines.

In addition, a pilot study was required for each type of study, except for the analysis of statistical relationships. The analyses of statistical relationships were reviewed by the Technical Advisors for 12th Grade Preparedness Research at several stages, and additional analyses were subsequently produced and reviewed. Board staff worked in collaboration with The College Board (SAT) and Florida Department of Education (FLDOE) staff to develop data sharing agreements for the studies of NAEP statistical relationships with SAT and FLDOE data. Board staff worked in collaboration with NCES staff, College Board staff, FLDOE staff, ETS staff, and Westat staff to develop procedures for maintaining the confidentiality of student data, matching records, and producing statistical results.

Design documents were developed by Board staff working with consultants for the content alignment and judgmental standard setting studies. Development of the higher education survey was a lengthy process of collaboration between Board staff and contractors, first at ACT and finally at Westat, and between Board staff and the Office of Management and Budget.

For the content alignment and judgmental standard setting studies, there was a lengthy process of identifying programs and panelists appropriate for the studies. Panelist recruitment is always a labor and time intensive process, and the effort was even greater for the judgmental standard setting studies involving panelists in areas entirely new to the Governing Board’s work.

The following chart shows the types of studies and indicates the current status and the date when studies will be complete and final reports submitted.
TIMELINES FOR REPORTING RESULTS OF STUDIES OF 12th GRADE PREPARED RESEARCH FOR 2009 NAEP READING AND MATHEMATICS

Content Alignment
- ACT-NAEP Content Comparison
- SAT-NAEP Content Alignment
- ACCUPLACER-NAEP Content Alignment
- WorkKeys-NAEP Content Alignment
  Final 2010

Statistical Relationships
- SAT-NAEP National Samples: COMPLETE; FINAL REPORT December 2011
- NAEP-FLDOE Assessment Data: COMPLETE; FINAL REPORT March 2012
- NAEP-FLDOE Post-Secondary Data: COMPLETE; March 2012; FINAL REPORT March 2012

Judgmental Standard Setting
- Evaluation of Course Requirements for Occupational Job Training Programs: UNDERWAY; FINAL REPORTS September 2012

Higher Education Course Placement Survey
- Survey Data Collection: August - December 2011
- Analysis Completed: March 2012

Report on 12th Grade Academic Preparedness Research
- Presentation to Governing Board: March 2012
- Governing Board Review: May 2012
- Public Report: June 2012
- Technical Report: June 2012
Content Alignment Studies

**Background:** 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. Content alignment studies for grade 12 NAEP in reading and mathematics will be implemented for the assessments for which data and materials are available. We have not yet succeeded in negotiating the use of The Armed Services Vocational Aptitude Battery (ASVAB) in our research, but Governing Board staff continue to pursue this potential.

A design document was developed by Dr. Norman Webb for the NAEP preparedness research alignment studies, and this design was implemented for the studies of the 2009 NAEP with the SAT and ACUPLACER in reading and mathematics. This design, with minor modifications, has also been used for the alignment of the 2009 NAEP with WorkKeys tests in these subject areas.

Content alignment studies for the first phase of the Board’s Program of Preparedness Research have been completed for NAEP in reading and in mathematics with WorkKeys, the SAT, and ACCUPLACER. In addition, a content alignment study was designed and conducted by ACT for the ACT and NAEP in reading and mathematics before the content alignment design document was developed. A brief summary of those studies was shared with COSDAM at the November 2010 meeting.

**Update:** Final reports have been received and a summary version has been shared with COSDAM and posted at [www.nagb.org](http://www.nagb.org). The complete reports for each study are available upon request. This work is final. There is no additional information for the March 2012 briefing.

Studies to Establish Statistical Relationships

**Background:** Highest priority is generally placed on these studies. Currently, two main sets of studies have been conducted under this heading. One is a study to relate SAT scores in reading and in mathematics to the national sample of NAEP scores for grade 12. The goal is to provide a statistical linking of SAT and NAEP scores for all students in the 2009 grade 12 NAEP who had taken the SAT by June 2009.

ETS staff reported that the match rate of approximately 33% of NAEP scores to SAT scores compares favorably to the national SAT participation rate of approximately 36% of public school students. The 11 states that participated in the pilot state-NAEP for grade 12 included some of the more populous states such as Massachusetts and New Jersey that are largely “SAT states,” but the 11 included more “ACT states.”

The correlation between NAEP and SAT reading was found to be lower than that for mathematics. The correlation between the NAEP reading scores and Florida state assessment...
scores was also lower than for mathematics. Research into those relationships, as suggested by the Technical Advisors for 12th Grade Preparedness Research, included: (1) inclusion of SAT scores in the NAEP conditioning model, (2) identification and removal of outliers, (3) evaluation of demographics of outliers, i.e., sensitivity analysis, (4) evaluation of alternative SAT scores (e.g. highest, most recent, composite driven), and (5) disaggregation of NAEP reading students’ scores based on block content of their assessment booklet.

The final sample used for linking the NAEP reading and SAT critical reading included approximately 16,200 students. The correlation between the two reading scales was 0.74. For NAEP and SAT mathematics, the linking sample included approximately 15,300 students, and the correlation between the math scales was 0.91.

Several methods were examined for evaluating the statistical relationship between NAEP and SAT scores for both mathematics and reading. A concordance was possible for mathematics, but the relatively low correlation did not support this method for establishing the SAT reference point for NAEP reading. Each method results in a different way of interpreting the results, and each can be helpful in understanding the overall relationship of NAEP and SAT for reporting preparedness. The preparedness benchmark for college readiness adopted by the College Board is 1550 for the composite score and 500 for writing, mathematics, or critical reading subtests. These benchmarks represent the score on the SAT that students would need in order to have a .65 probability of earning a B- freshman year grade point average. For each of the linking methods examined, the results indicate that the SAT readiness benchmark of 500 is very close to the NAEP Proficient cut score for both reading and mathematics, and the scale alignment varies by the probability of the estimate.

The second set of studies analyzes a series of statistical relationship for Florida’s NAEP examinees. NAEP’s 2009 state-representative sample of Florida 12th graders was used to match NAEP scores for reading and mathematics to student scores on several tests collected by the Florida Department of Education (FLDOE). The data sharing agreement with FLDOE provides access to scores for the SAT, ACCUPLACER, and WorkKeys. Additionally, ACT, Inc. has given permission to the Florida Department of Education to share ACT scores with the Governing Board for purposes of conducting the grade 12 preparedness research. We also plan to obtain employment data and salary data for Florida examinees, but access to those data was not included under the current data sharing agreement. A plan to allow for electronic transfer of data was developed to keep secure the identity of students, consistent with the NAEP legislation, FLDOE requirements, and requirements of each assessment program.

Records for roughly half of the Florida grade 12 NAEP examinees in 2009 could be matched to an ACT score and half to an SAT score. This match rate is consistent with other data for Florida students. The match of WorkKeys scores to the total 2009 state NAEP sample of 12th graders was only about 6%. FLDOE reported that around 89,300 Florida 12th graders were enrolled in vocational-technical programs in school year 2008-09. The match of WorkKeys examinees to NAEP examinees was not sufficient to warrant additional analyses for the 2009 cycle. The state of Florida has only recently implemented the testing of high school students in vocational programs with the WorkKeys exam, and we anticipate that the number of examinees will increase in subsequent years.
**Update:** Included in Attachment A-1 is a description of the analysis of the statistical relationship between NAEP grade 12 data and the state-representative sample of 12th grade Florida students who took the NAEP in 2009 in either reading or mathematics. The analysis for Florida students includes both statistical linking studies with ACT, SAT, and ACCUPLACER test data and data on the first year in college for students who entered college. This research is on-going, and data will be provided annually to update the post-secondary performance of Florida students who were in the 12th grade NAEP sample in 2009. The complete report will be finalized and shared with COSDAM prior to the March 2, 2012 meeting.

**Judgmental Standard Setting Studies**

**Background:** A series of judgmental standard setting studies was planned to produce preparedness reference points on the NAEP scale for entry into job training programs and for placement in college credit-bearing courses. Within this category of studies, the Technical Panel for 12th Grade Preparedness Research placed highest priority on the judgmental studies related to preparedness for job training programs in 5-7 exemplar jobs. This priority is largely related to the paucity of national data available for statistical studies in these areas. Unlike most other studies of preparedness for post-secondary activities in college or job training programs, the Governing Board has not assumed that prepared for college and the workplace are the same. Rather, our studies are aimed at determining whether the level of performance on NAEP is approximately the same or significantly different for entry in job training programs for the occupations included in our research studies and placement in credit-bearing college courses that fulfill general education requirements for a bachelor’s degree.

In order to maximize the standardization of judgmental standard setting (JSS) studies within and across post-secondary areas, a design document was developed to specify the number of panelists, the eligibility criteria for panelists, the procedures for drafting and finalizing borderline performance descriptions, the methodology to be implemented, feedback to be provided, key aspects to be evaluated, and reports to be produced. The methodology and basic procedures specified for the design of these studies were those implemented for the achievement levels-setting process for the 2006 grade 12 economics NAEP and for the 2009 science NAEP for grades 4, 8, and 12.

The five exemplar jobs approved by COSDAM for inclusion in these studies are as follows:

1. automotive master technicians
2. computer support specialists
3. heating, ventilation, and air conditioning technicians
4. licensed practical nurses
5. pharmacy technicians
An overview of the pilot study for automotive master technicians and college course placement, conducted April 26-29, was presented to COSDAM on May 13, 2011. Implementation of the operational studies for job training programs in 5 occupations and for college course placement was completed on July 1, 2011. Operational sessions were conducted for two post-secondary areas each and for both reading mathematics. A pair of replicate panels with 10 panelists each was convened for each subject and post-secondary area for a total of 24 operational panels.

A comprehensive report on the results for the judgmental standard setting studies was provided to COSDAM in November 2011 via a WebEx virtual meeting. At that time, COSDAM discussed the general approach to the preparedness studies and the need to evaluate results of the judgmental standard setting studies in the context of the overall set of studies. COSDAM recommended that reports emphasize the fact that the studies included job training programs for only five occupations and noted that the results for five other occupations could point to different interpretations regarding the minimal level of academic preparedness.

Additional research is underway to collect more information from the job training programs included in these JSS studies. This work is being conducted under the direction of WestEd as the primary contractor and the Educational Policy Improvement Center (EPIC) as the subcontractor. The study requires collection of course materials for both entry level courses and courses at the completion of job training programs. The materials will be reviewed to determine the knowledge, skills, and abilities (KSAs) that students need to have to enter the job training program, i.e. pre-requisites, and those that are taught in the program. The KSAs represented in course requirements will be compared to the borderline performance descriptions developed in the JSS studies for each job training program in each subject to evaluate the extent to which the descriptions used to set the cut scores on NAEP for each job training program align with the pre-requisite knowledge, skills, and abilities, those taught in the beginning courses, and those taught in the final courses. The information from course materials will also be compared to the NAEP item pool for each subject to evaluate the alignment of knowledge, skills and abilities assessed by NAEP relative to those required as pre-requisites, in courses taught at program entry, and in courses taught just at program completion. Finally, the knowledge, skills, and abilities represented by the course materials will be compared to items in the NAEP item pool near the cut scores set by the panelists to represent minimal preparedness for entry in each training program. This information is being collected to evaluate the reasonableness of the cut score recommendations across the job training programs.

**Update:** A report on the project is included as Attachment A-2. A timeline for the key activities of the project is included in that report. Information regarding preliminary findings will be provided to COSDAM at the meeting on March 2, 2012.
Higher Education Survey

**Background:** A survey of two-year and four-year post-secondary institutions was implemented to gather information regarding (1) the placement tests used and (2) the cut scores in reading and mathematics below which remedial reading and mathematics course placement results, and at or above which placement in credit-bearing entry level courses results. The sample was nationally representative of accredited postsecondary education institutions. Results will be analyzed according to several attributes of the institutions, including enrollment size and level of selectivity for admissions.

Several small-scale studies were conducted to refine the survey instrument. A pilot study of 120 postsecondary institutions yielded an overall response rate of 86 percent. Based on the results of the pilot study, minor changes to the questionnaire were made.

The survey was submitted to the Office of Management and Budget (OMB), and clearance was achieved in July, 2011. Distribution of the survey to institutions began in August, 2011.

**Update:** An 86% response rate was achieved with responses from 1522 institutions. Data tables have been prepared and presented to Governing Board staff for analysis. Ray Fields will join the COSDAM meeting to discuss preliminary findings.

Benchmarking Studies

**Background:** The Governing Board has discussed the possibility of conducting benchmarking studies as part of the 12th grade preparedness research. This type of study was not included in the recommendations of the Technical Panel for 12th Grade NAEP Preparedness Research for the first cycle of research, largely because of the challenges involved in collecting the data. They suggested that the design would potentially provide useful preparedness information and help to answer questions that were likely to arise from the research studies in the first cycle.

Benchmarking studies in the preparedness research context are studies in which NAEP is administered to groups of interest, e.g., college freshmen enrolled in credit-bearing college level courses that fulfill general education requirements for a four-year degree without the need for remediation. Determining the average NAEP performance of this group would then provide a “benchmark” score that can be considered as one of the reference points on the NAEP scale. A benchmarking study in combination with reference points from other studies in the Program of Preparedness Research can assist the Board in determining the areas of the NAEP scale that indicate preparedness. A benchmarking study of Texas college freshmen was planned, and it had the support of the Texas Commissioner of Higher Education and the cooperation of nine Texas higher education institutions. A small scale pilot study to evaluate the feasibility of the study design was implemented.

The Governing Board and the National Center for Education Statistics (NCES) collaborated on the implementation of this study, which was carried out by Westat. Westat, the NAEP sampling and administration contractor to NCES, conducted interviews with officials in each of the 9 higher education institutions in Texas that volunteered to participate in the pilot study at the invitation of the Texas Commissioner of Higher Education. The institutions included both two-
and four-year institutions with considerable diversity among the enrollments across the institutions. Those interviews were conducted to gather specific information related to logistics, student sampling, and student recruitment prerequisite for the pilot study. The response from each campus was positive and enthusiastic. The data collection phase for the pilot ended on October 15, 2010. Of the eligible sample of 1,234 students, 255 actually attended a NAEP session, for an overall response rate of 20.7 percent.

As announced at the November 2010 meeting of COSDAM, NCES, Westat, and Governing Board staff met to discuss alternatives. Board staff decided that we will not proceed to the operational phase of this study due to low participation rates and the lack of feasible alternatives to increase participation.

No additional benchmarking studies are planned for the 2009 NAEP preparedness research.

High School Transcript Study Data

Background: Board staff plan to use data from the 2009 High School Transcript Study in reporting the grade 12 NAEP preparedness research as a source of contextual information for the results of the Board’s program of preparedness research. The transcript data are available for mathematics and science courses and not for courses related to student performance in reading. Course information is standardized as part of the High School Transcript Study to enable national comparisons. Data were collected from over 700 high schools in the 2009 study and include a variety of information in addition to course titles, grades, and credits earned. The data also include, for example:

- course and program offerings at each high school in the sample;
- graduation requirements at each high school in the sample;
- type of course (e.g., remedial, honors, Advanced Placement, International Baccalaureate);
- location of the course (e.g., on campus, distance learning, vocational education center);
- approximate percentage of graduates who enrolled in 2-year, 4-year, and vocational postsecondary institutions for each high school in the sample; and
- graduation status at the end of grade 12 (e.g., type of diploma earned, still enrolled, dropped out).

Data on course-taking patterns can supplement the Board’s preparedness research in the following ways.

1. Provide contextual indicators on the NAEP scale
   Each study in the full compilation of the Board’s Program of Preparedness Research is expected to result in a preparedness “reference point” on the NAEP scale. A reference point could be a score point or a range of scores on the NAEP scale. Findings from the full compilation of preparedness studies will be examined in relation to particular descriptive statistics, such as:
   - Average NAEP score for students who took a “rigorous” curriculum;
   - Average NAEP score for students who took an Advanced Placement course; and
• Average NAEP score for students who took an International Baccalaureate course. These data can be useful in interpreting the various reference points associated with preparedness and can be used to evaluate the logical relationships and mutually confirmatory evidence of validity.

2. Identify course-taking patterns associated with reference points
Selected reference points, or score ranges, can be used to identify course-taking patterns for students whose scores fall within these same score ranges:
• Use score ranges associated with preparedness that are identified in the Judgmental Standard Setting Studies and the Statistical Linking Studies in the Board’s Program of Preparedness Research, and identify patterns in course-taking for students whose scores fall within or above these score ranges.
• Use the longitudinal data provided by Florida to identify the average NAEP score for students who placed into college-level courses, and identify patterns in high school course-taking for students who scored at or above this level.

Update: Data from the 2009 high school transcript study will be incorporated into the reports on 12th grade preparedness research.

12th Grade Preparedness Validity Framework
Development of validity framework continues for the 12th Grade NAEP Program of Preparedness Research. Versions of the document have been reviewed by experts in validity and in program evaluation, as well as by COSDAM. Comments from the NAEP Validity Studies Panel have been received, and changes to the document in response to that input are in progress.

Update: Staff will complete work on the validity framework and share the document with Technical Advisors for Preparedness Research, other advisory groups, and with individuals and groups identified as having an interest in the research. The comments and recommendations will be shared with COSDAM and the Governing Board to inform reporting of 2009 preparedness data and research for the 2013 and subsequent grade 12 assessments.

Procurements for Reporting 12th Grade NAEP Preparedness Research
A contract was awarded to Widmeyer Communications to work with the Governing Board staff for production of the “public” report on the overall results of the 2009 NAEP preparedness research to be produced as both a print document and web-based document and addressed to a general audience. Production of completely electronic web-based technical report will be primarily addressed to the research community. Work is underway on developing the overall conceptual design of the reports. Widmeyer staff are attending meetings of the High School Commission on 12th Grade NAEP Preparedness Research, the Technical Advisors for 12th Grade NAEP Preparedness Research, and meetings of Board committees—COSDAM and Reporting and Dissemination to gain greater understanding of the research and related Board policies.
The Reporting and Dissemination Committee will be the primary focus of updates on preparation of the final reports documents, and COSDAM will continue to be the primary focus of updates on the studies in the reports.

The Governing Board will be briefed on the program of research and findings to date at the March 2012 meeting and on the final reports at the May 2012 meeting. The final reports are expected to be ready for distribution by June 2012.

OVERVIEW OF ASSESSMENTS

For additional background information, the following list presents a brief description of the assessments that the Technical Panel on 12th Grade Preparedness Research recommended for analysis in NAEP preparedness research. Many of these assessments are the primary focus of the proposed content alignment studies and statistical relationship studies. In each case, only the mathematics and reading portions of the assessments will be analyzed, although analyses with the composite scores may be conducted.

- **ACCUPLACER** – ACCUPLACER is a computer adaptive test used for college course placement decisions in two-year and four-year institutions. It is produced by the College Board and includes assessments of sentence skills, reading comprehension, arithmetic, elementary algebra, college level math, and written essays.

- **ACT** – The ACT assessment is a college admissions test used by colleges and universities to determine the level of knowledge and skills in applicant pools, including reading, English, and mathematics tests. ACT has *College Readiness Standards* that connect reading or mathematics knowledge and skills and probabilities of a college course grade of “C” or higher (75%) or “B” or higher (50%) with particular score ranges on the ACT assessment.

- **ACT WorkKeys** – WorkKeys is an assessment designed for use by employers to evaluate the knowledge and skills of a prospective employee relative to a job profile. WorkKeys scores are used more generally to certify workplace readiness. WorkKeys assesses knowledge and skills in communication (business writing, listening, reading for information, writing) as well as problem solving (applied technology, applied mathematics, locating information, observation). There is also an interpersonal skills section of WorkKeys.

- **SAT** – The SAT reasoning test is a college admissions test produced by the College Board. It is used by colleges and universities to evaluate the knowledge and skills of applicant pools in critical reading, mathematics, and writing. The College Board has provided SAT score data to be used in research studies to establish a statistical relationship between the SAT and NAEP.
NAEP 12th Grade Preparedness Research:
Overview of Analyses Relating Florida Students’ Performance on NAEP to Preparedness Indicators and Postsecondary Performance

Rebecca Moran, David Freund, and Andreas Oranje, ETS

As part of the National Assessment Governing Board’s efforts to enable NAEP to report on the preparedness of U.S. twelfth graders for postsecondary education or entry into job training programs, studies were conducted to statistically relate performance on NAEP with results from other assessments that serve as indicators of preparedness for college entry, course placement, and entry into the workforce (National Assessment Governing Board, 2009). Both nationally-representative data, such as those used to establish a statistical link between NAEP and SAT, and data representative of students in individual states were of interest for such statistical relationship studies. The 2009 12th grade NAEP Reading and Math assessments included a first-time pilot state assessment for 11 states; Florida was one of the participants. The Florida Department of Education (DOE) maintains a longitudinal data base (K-20 Education Data Warehouse) that includes college entrance and placement test scores and first-year college performance data for those students who attended public colleges in the state of Florida during the 2009-2010 academic year.

This document describes the data and procedures used to evaluate Florida students’ performance in high school and first year of college relative to scores on the NAEP 12th grade reading and mathematics assessments and other test scores. This will be followed by a description of the analyses that serve as a follow-up to the statistical relationships established between NAEP and SAT at the national level. Of particular note are the analyses of postsecondary data to provide validity evidence for the potential preparedness reference points on the NAEP scales identified in the national statistical relationship studies.

Data

This study used data from Florida public school students who participated in the 2009 NAEP 12th grade reading or mathematics assessments, approximately 3,200 in math and 3,400 in reading. Analyses were conducted with the use of NAEP sampling weights to appropriately represent 12th grade public school students in Florida in 2009.

Matching NAEP and Florida DOE Data

The Governing Board entered into an agreement with the Florida DOE to obtain longitudinal data for public school students selected to participate in the 2009 NAEP 12th grade assessment. The process of matching data between the Florida database and NAEP participants was carried out in coordination with NAEP contractors, Westat and ETS, and the Florida DOE.

A critical requirement of the matching of student records was to protect students’ identity and maintain confidentiality. This was assured through the assignment of a unique pseudo ID for students sampled to
participate in NAEP. At the time of sample selection of students for operational NAEP, Florida DOE staff appended the pseudo ID to files within the Florida DOE and transmitted the pseudo ID to Westat with other administration data. On all subsequent data files containing Florida data (e.g., ACT scores), only the pseudo ID was included on the files. The pseudo ID was used by Westat to match files from Florida back to the NAEP data files. Westat in turn provided files to ETS with the additional Florida data appended to NAEP student records. Throughout the process, ETS had no access to any Personally Identifiable Information (PII), such as names, birthdates, or social security numbers. This process was essentially identical to the matching process conducted with the NAEP-SAT national linking study.

Data Elements evaluated for use in the Florida Preparedness Research

Of the variables available in the rich Florida longitudinal dataset for the 2009 12th grade cohort, those examined for use in this research are described briefly below. Some of these data elements lacked sufficient power (i.e., small sample sizes in the linked set) and, therefore, value for extensive use in the current research.

1. **Florida Comprehensive Assessment Test** (FCAT) is Florida’s K-12 state assessment. Scores on the reading and mathematics tests from 3rd through 10th grades were available in the longitudinal dataset. Match rates were very high, with 10th grade scores in reading and math matched to approximately 94% of the NAEP test takers. However, concerns about the relevance of relating students’ 12th grade NAEP performance to FCAT scores earned two years earlier, while in 10th grade, precluded further analyses being pursued with the FCAT data.

2. **WorkKeys®** is a job skills assessment system that helps employers select, hire, train, develop, and retain a high-performance workforce. WorkKeys® includes three relevant tests: Applied Mathematics, Locating information, and Reading for Information. Matched sampled sizes were about 300 students (about 10% of the NAEP sample in each subject) for each WorkKeys® test and therefore were inadequate for further analysis.

3. **Advanced Placement (AP)** college-level exams enable students to earn college credit and advanced placement in college courses. Approximately 36% of students in the matched NAEP-Florida dataset took one or more AP tests. However, only 16% of the NAEP reading sample took a relevant reading AP exam (English or English Literature) and only 8% of the NAEP sample took a relevant math AP exam (calculus). The small sample sizes limited the efficacy of these data for further analysis.

4. **High School Program:** One of the background questions asked of students on the 12th grade NAEP assessment was “Which of the following best describes your high school program?” Response options included (1) General, (2) Academic or college preparatory, and (3) Vocational or technical school. For the linked NAEP/Florida sample, approximately 47% of students indicate their program was “general”, 43% indicated “academic or college preparatory”, and 9% indicated “vocational or technical school”. These data were examined in much greater detail; the technical report to be shared at the March Governing Board COSDAM meeting will provide more results.
5. **SAT and ACT College Entry Exams and ACCUPLACER College Placement Exam.** Approximately 43% of Florida’s NAEP sample took the SAT test; 47% took the ACT test, and 18% took the ACCUPLACER test.

6. **College Enrollment Status, First-Year Course-taking and Grade-point Average:** Data were obtained for students attending public colleges and universities in Florida for the 2009-2010 academic year. Approximately 54% of the students in Florida’s 2009 NAEP 12th grade sample attended a public postsecondary institution in Florida, with 36% attending community colleges and 17% attending four-year colleges and universities.

### Analyses Conducted

The purpose of this research activity was to explore the relationships between Florida students’ performance on the 12th grade NAEP assessments and other indicators of postsecondary preparedness to provide validity evidence for the potential preparedness reference points on the NAEP scales that were identified by the national NAEP-SAT statistical relationship study. Those potential reference points are as follows:

<table>
<thead>
<tr>
<th>Statistical Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percentage at or above 500 on SAT</strong></td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>67%</td>
</tr>
<tr>
<td>80%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAT Subscore = 500</strong></td>
</tr>
<tr>
<td>165</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NAEP “Proficient”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>174</strong></td>
</tr>
</tbody>
</table>

Thus, NAEP scale score distributions were examined in relation to:

- SAT performance, defined in terms of whether students met the SAT college readiness benchmark in each subject area (determined by College Board to be a score of 500 on the SAT critical reading and mathematics measures)
- ACT performance; the benchmarks for college readiness established by ACT are 22 for math and 21 for reading
- ACCUPLACER performance, where the cut scores used in Florida for placement into credit-bearing courses are 72 for math and 83 for reading
• Students’ self-reported program of study in high school, whether vocational/technical, general, or academic/college preparatory. This information was collected as part of the NAEP student background questionnaire.
• College enrollment: community college, four-year college/university, or unknown
• First-year college course-taking: remedial or credit-bearing
• First-year college grade point average: above or below a B-

Summary of Results

Overall, the patterns of results observed for this cohort of Florida students did not contradict the potential preparedness reference points on the NAEP reading and mathematics scales identified through the national NAEP-SAT linking study. For instance,

• Average NAEP scores for students who met the SAT preparedness benchmarks were near the NAEP Proficient cut scores and fell in or above the range of possible NAEP preparedness reference points. The average scores for these students were roughly one standard deviation higher than average NAEP scores for their peers who did not meet the SAT preparedness benchmark.
• Performance on NAEP was similarly differentiated for students who met vs. did not meet the ACT preparedness benchmarks.
• Average NAEP scores for students attending four-year colleges fell within or exceeded the potential NAEP preparedness reference points, whereas average scores for their peers attending community colleges were below the reference points.
• A greater percentage of students taking no remedial courses during the first year of college scored at or above the NAEP preparedness reference points compared to students who required one or more remedial courses.

However, the limitations of the Florida data, namely the availability of data only for students enrolled in Florida public postsecondary institutions, must be taken into consideration when interpreting these results.

Figure 1 provides sample sizes and percentages for the 2009 NAEP 12th grade Florida sample disaggregated by high school program, test-taking, college attendance, and remedial course-taking.
Figure 1: Sample sizes and Percentages for 2009 NAEP/Florida Grade 12 Math

- **H.S. Program**: N = 3200 (5% of nation)
  - Voc/Tech: N = 300; 9%
  - **Academic/College Prep**: N = 1500; 43%
  - General: N = 1600; 47%
- **Took SAT**: N = 1400; 43%
- **Took ACT**: N = 1500; 47%
- **Took Accuplacer**: N = 600; 18%
- **Attended Florida Public College**: N = 1800; 54%
  - Remedial Courses: Yes N = 600; 19%, No N = 1200; 35%
  - Community College: N = 1200; 36%
  - 4-year College: N = 600; 17%
Job Training Program Curriculum Study for NAEP Preparedness Research

Submitted by WestEd

The National Assessment Governing Board (Governing Board) adopted a Program of Preparedness Research in March 2009 that included judgmental standard-setting (JSS) studies for the 12th grade National Assessment of Educational Progress (NAEP). These studies produced preparedness reference points on the NAEP scale for entry into job-training programs and for placement in college credit-bearing courses, representing the academic knowledge and skills required for postsecondary course and training program placement. A total of 180 job training programs were represented in the judgmental standard setting studies focusing on five occupations:

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automotive master technician</td>
<td>41</td>
</tr>
<tr>
<td>Computer support specialist</td>
<td>31</td>
</tr>
<tr>
<td>Heating, ventilation, and air conditioning technician</td>
<td>31</td>
</tr>
<tr>
<td>Licensed practical nurse</td>
<td>40</td>
</tr>
<tr>
<td>Pharmacy technician</td>
<td>37</td>
</tr>
</tbody>
</table>

The Governing Board requested additional research to examine the validity of findings obtained from the JSS studies and to better understand the knowledge, skills, and abilities in reading and mathematics required for these occupational training programs. This additional research is intended to provide a clearer understanding of the knowledge, skills, and abilities (KSAs) required for entry- and exit-level coursework in designated job training programs within these occupations. This study will help to determine if the KSAs required of students in the training programs are appropriately represented by the borderline preparedness descriptions (BPDs) and by the NAEP items near the reference points developed in the JSS studies to represent the minimal level of academic knowledge and skills in the subject matter necessary for a student to be prepared to enter the job training course.

**METHODOLOGY**

This study addresses the following research questions:

1. What mathematics and reading KSAs are prerequisite to the introductory-level courses, and what mathematics and reading KSAs are taught in the introductory courses for the job-training programs for each occupation?
2. What mathematics and reading KSAs are students expected to have attained at the conclusion of the job-training programs for each occupation?
3. How do the prerequisites (KSA expectations for entry) for job training programs in each occupation relate to descriptions of minimal academic preparedness on NAEP (as described by the BPDs from the JSS studies)?

4. How do prerequisites (KSA expectations for entry) for job training programs in each occupation relate to the content assessed by NAEP (as determined by NAEP items representing minimal academic preparedness)?

This study comprises three primary phases:
1. Identification and collection of course artifacts
2. Review of course artifacts by Review Teams
3. Review of resulting KSAs by NAEP Content Expert Teams

**Identification and Collection of Course Artifacts**

Programs from the five occupations used in the JSS studies will comprise the population of programs for this study; from this population, a minimum of 20 programs per occupation will be recruited from the 180 programs represented on the JSS panels.

Occupational job-training instructors who served on the JSS panels are being recruited to participate in this study. These job training instructors are being asked to identify courses that best address the objectives of this study and to submit artifacts for those courses. These instructors also have the option of nominating colleagues who teach one or more courses selected for the study to participate in this activity. Course artifacts will be collected for all programs in each occupational area that agree to participate, with course submission remaining open until either materials are obtained from a minimum of 20 programs or the population of programs has been exhausted.

Each participating program instructor is being asked to (1) identify foundational textbooks for her/his program; (2) verify program information collected by EPIC (e.g., accreditation status, course sequencing, school and department admission requirements, degree accreditation, and credit requirements); and (3) submit course artifacts for two introductory courses and two concluding courses. Course artifacts may be submitted via a web-based upload tool, email, facsimile, and physical mail.

**Introductory courses**

Introductory courses differ across programs within an occupation, and across occupations, in terms of standardization and sequencing. As such, “entry-level” courses could embody one or more of numerous definitions, including (1) those that occur lowest in the course sequence for a program, regardless of course title; (2) those that are core “Introduction to...” or “Foundations of...” courses that occur across the majority of programs, and (3) those that are identified by instructors as being most representative of the mathematics and reading expectations for entry-
level students in the program. Because the focus of the study is on identifying the mathematics and reading skills expected upon entry into introductory-level courses in the job-training programs for each occupation, courses are being selected for inclusion using the third definition.

Concluding courses
Concluding, or exit-level, courses also differ in level of standardization, and multiple options for identifying such courses also exist. For consistency, the same approach is being used to identify the exit-level courses for inclusion in the study: instructors are asked to identify those courses that best represent the mathematics and reading knowledge and skills that students are expected to know upon program completion.

For each training program, a set of course materials will be collected for introductory courses and a set for concluding courses. The following types of artifacts will be submitted and assembled into a course packet (with at least one of each type of artifact required):

1. Course syllabus
2. Textbook title(s) (with author and ISBN)
3. Textbook table of contents (instructor will copy and upload or EPIC will download from publisher website)
4. Course exam (one or more), preferably the mid-term or earlier for introductory courses and the final exam for concluding courses
5. Text-based assignment (one or more), with corresponding passage, that best illustrates mathematics and reading KSAs needed by students—one or more for introductory courses and one or more for concluding courses
6. Stand-alone assignment (one or more) such as a lab, worksheet, problem sheet, essay, or group project that best represents mathematics and reading KSAs needed for students—one or more for introductory courses and one or more for concluding courses

Instructors representing institutions that offer more than one program within an occupational area are asked to complete a submission for one program and to complete submissions for additional degree programs if selected courses are different than those already submitted.
Review of Course Artifacts
Once course artifacts are gathered, a team of content experts will be trained to consistently and reliably apply a coding scheme to the course artifacts to identify prerequisite and taught content for each of the occupational training programs. Two Review Teams, one for mathematics and one for reading, have been recruited. Each consists of two mathematics or two reading experts and one occupational area (e.g., automotive master technician) expert. The content experts in mathematics and reading were recruited from a pool of trained analysts who have substantial experience in this type of work. The occupational-area experts recruited for each of the Review Teams (one for mathematics and one for reading) were drawn from pools of mathematics and reading JSS panelists who were nominated by the JSS studies’ content and process facilitators as being well qualified for this type of work.

Review Teams will independently code the course packets for their content area. In order to maximize the efficiency of the Review Teams, an initial set of foundational KSAs is used to analyze course materials. These foundational KSAs include the NAEP frameworks and additional KSAs derived from the National Career Clusters™ Essential Knowledge and Skill Statements, synthesized to reduce redundancy and to present only those KSAs relevant to mathematics and reading.

Once the Review Teams’ review of course materials is complete, EPIC staff will aggregate the individual ratings for each course within each program to summarize the mathematics and reading KSAs that are prerequisite to and taught in introductory-level courses and that students are expected to have attained at program completion. Responses will be aggregated to create overall content maps describing the relationship between frameworks and prerequisite KSAs for each occupation. In addition to tabular data displays, the data will be displayed using color shading, as well as summary statistics, to show the extent of overlap in content between standards and programs. Content maps, grouped by key characteristics, will also be created for programs, to show the impact of key program characteristics that impact findings. EPIC staff will review the content maps to identify similarities and differences across program types within occupations and will note differences in findings due to program characteristics. Final results will be provided both overall and by key program characteristics. EPIC staff will also compute descriptive statistics to summarize the Review Teams’ demand ratings overall (by occupation) and by program type, should program characteristics have an impact on the demand of occupational courses.

Review of Knowledge, Skills, and Abilities Required for Training Courses
Two NAEP Expert Teams, one team for mathematics and one for reading, each consisting of two experts, will review the prerequisite and taught KSAs (as identified by the Review Teams) in the context of NAEP. They will describe the relationships between the prerequisite content and both the BPDs and the content on the 2009 NAEP, evaluating the results of the Review Team analyses to describe
KSAs assessed by NAEP that are not included in the job-training programs and KSAs included in the job-training programs that are not part of the NAEP frameworks or assessments.

**Comparison to BPDs**
Using the Review Teams’ determination of KSA requirements and course artifacts, the NAEP Expert Teams will synthesize and describe the relationship between the content that is prerequisite to and taught in occupational programs and the content described in the BPDs for that program. Conclusions will be provided overall for each occupation, identifying differences related to program characteristics.

**Comparison to NAEP items**
Each NAEP Expert Team will also compare KSAs identified for each program’s introductory courses (drawing upon the content maps and BPD comparisons) to the NAEP item pools. Starting with a set of items near the cut scores identified in the JSS studies, they will judge the correspondence between the course prerequisite KSAs and the KSAs needed to correctly respond to the items with a .67 probability. They will be asked to identify the items in the range of the cut score plus one standard deviation that are prerequisite to or required in the courses. They will also be asked to examine items below the cut score and above the range in the first analysis to determine if the KSAs represented in the curricular requirements are largely above or below this range.

**PILOT STUDY**
In order to address unanticipated challenges that may arise when implementing the proposed design, materials, and/or logistics, a pilot—or feasibility—study is being implemented. The automotive master technician occupation has been selected for the pilot study. Lessons learned through the pilot study will be used to refine the study design as needed for the subsequent four occupations.

However, to expedite the collection of and reporting on introductory course data—which is of particular and immediate importance to the Governing Board—coding and analysis of introductory course data, which begins with the automotive master technician occupation, will continue for the remaining occupational areas before the analysis of concluding courses commences. Similarly, the NAEP Expert Teams will conduct their analyses for each occupation’s introductory courses prior to conducting their analyses of the concluding courses; in addition to providing introductory and concluding course analyses, the Expert Teams will also provide an overall review of each program and occupational area as well as a summary report of KSAs identified as pre-requisites to automotive master technician programs.

Decision points identified during the pilot study will be discussed with the Governing Board, and decisions will be documented and implemented for data collection and analyses within the remaining occupations.
KEY ACTIVITIES IN PROPOSED PROJECT SCHEDULE

<table>
<thead>
<tr>
<th>PLANNING ACTIVITIES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Submission Tool released to participants</td>
<td>1/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PILOT STUDY ACTIVITIES (INTRODUCTORY COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course artifact collection</td>
<td>1/12/12–2/14/12</td>
</tr>
<tr>
<td>Review Teams course packet reviews</td>
<td>2/3/12–2/24/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>3/1/12–3/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REMAINING OCCUPATIONS ACTIVITIES (INTRODUCTORY COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course artifact collection†</td>
<td>2/13/12–3/9/12</td>
</tr>
<tr>
<td>Review Teams course packet reviews</td>
<td>3/2/12–4/20/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>3/12/12–5/7/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PILOT STUDY ACTIVITIES (CONCLUDING COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Teams course packet reviews</td>
<td>4/24/12–5/4/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>5/4/12–5/16/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REMAINING OCCUPATIONS ACTIVITIES (CONCLUDING COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Teams course packet reviews</td>
<td>5/8/12–6/29/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>5/18/12–7/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REPORTING</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>2/9/12</td>
</tr>
<tr>
<td>Draft pilot report submitted to Governing Board</td>
<td>3/31/12</td>
</tr>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>4/19/12</td>
</tr>
<tr>
<td>Final pilot report submitted to Governing Board</td>
<td>4/30/12</td>
</tr>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>7/5/12</td>
</tr>
<tr>
<td>Draft final report submitted to Governing Board</td>
<td>7/31/12</td>
</tr>
<tr>
<td>Final report submitted to Governing Board</td>
<td>9/7/12</td>
</tr>
</tbody>
</table>

PROGRESS UPDATE
An initial program analysis identified for each institution the number of qualifying degree programs as well as the courses required for degree completion. There is a total of 41 automotive master technician training programs in the study, and 38 institutions with 61 programs were invited to participate. Of these 61 programs, 12 programs within 10 schools have completed a total of 78 course submissions, including 40 introductory courses (21 reading and 19 mathematics courses) and 38 concluding courses (22 reading and 16 mathematics). Another 23 course submissions are in progress for 14 programs. EPIC continues to call and email instructors in the auto programs to assist with course document submission.

1 Course artifacts are being gathered for introductory and concluding courses concurrently.
The automotive master technician Review Teams have completed reviews of six courses each in reading and mathematics. The remaining four occupations will begin submitting course documents the week of February 13. EPIC has confirmed the participation of NAEP Expert Teams to conduct the summary review of automotive master technician introductory course findings, and their review instrument is in development. EPIC is preparing a progress report with initial pilot findings to inform and improve data collection and analysis for the remaining four occupational areas.
The first fully computerized National Assessment of Educational Progress (NAEP) was administered in 2011 for writing at grades 8 and 12. Writing tasks were presented in a variety of formats including written documents and instructions, video clips, and audio clips.

For the writing ALS process, the Body of Work (BoW) method was the process implemented for panelists to judge student performances relative to the NAEP achievement levels descriptions for Basic, Proficient, and Advanced. The Body of Work method is a holistic judgmental process whereby student work is judged according to criteria (NAEP achievement levels descriptions) and classified accordingly. Performances judged to be lower than the description of the Basic achievement level were to be classified as below Basic. Achievement levels-setting panels include teachers, other educators, and representatives of the general public. A total of 100 student booklets were evaluated by each panelist relative to the achievement levels descriptions. Panelists were first given 50 booklets to classify. They used the same 50 booklets to classify booklets two times, with feedback and discussions after each classification. Cut scores were computed for each panelist, and the grade level cut score was computed as the median of the panelists’ cut scores for each grade. A different set of 50 booklets were distributed for the third round of classifications, and the third classifications were used to compute the final cut scores.

**Achievement Levels Descriptions**

Achievement levels descriptions (ALDs) were developed for use in the panel meetings, including the operational achievement levels-setting meetings. A lengthy process involving many writing experts was implemented to develop and evaluate the descriptions that were recommended for use in the 2011 achievement levels-setting process. The Committee on Standards, Design and Methodology (COSDAM) approved the achievement levels descriptions at the August 2011 meeting, with the understanding that final approval would be given in May 2012 to make these the “official” descriptions as part of the overall process of setting the achievement levels.

As a result of the pilot study implemented in November 2011, staff recommended that the achievement levels descriptions be reviewed again. Panelists commented on the ambiguity of some aspects of the descriptions, and staff observed that panelists had a tendency to translate less precise wording to mean lower performance requirements (e.g., “some” was taken to mean “any.”) Further, the results of a study comparing performance on the 2007 and 2011 assessments relative to the 2011 descriptions raised further concerns regarding the achievement levels descriptions. Although several factors could be contributing to the differences resulting from this comparison, COSDAM agreed that the achievement levels descriptions should be re-evaluated.
Further, COSDAM recommended that a small-scale study be conducted to try out the ALDs prior to the operational study, if modifications to the ALDs were made.

Staff worked with content experts to review and modify the achievement levels descriptions. The goal was to make the wording more precise and to further address the calibration of levels within and across grades. As always, the NAEP policy definitions were the primary guide for calibration of the descriptions of what students should know and be able to do. The modified descriptions were reviewed by eight additional content experts, most of whom had been part of the NAEP Writing Framework panel. The recommendations of these experts were evaluated and incorporated to further modify the ALDs.

As a result of comments from panelists in the study to try out the ALDs (Field Trial 2, described below), the ALDs were modified yet again. Panelists considered the description of Basic level performance to be more rigorous than the policy definition for Basic performance. Content experts again worked with staff to modify the descriptions. Finally, agreement was reached that the descriptions were clear and appropriately calibrated across levels and grades. The descriptions used in the operational achievement levels-setting process are included in Attachment B-1. These descriptions are recommended for consideration by COSDAM as part of the overall set of information to be used in reporting results for the 2011 Writing NAEP.

Field Trial #2
One field trial was planned to study the logistics of using two computers in the ALS process. Field trial #1 was implemented in September 2011 with panelists for grade 12 only.

A second field trial was implemented to try out the revised achievement levels descriptions and one other modification to the process. A paper selection process has been used in most NAEP achievement levels-setting procedures, and it was added to the Body of Work process to be implemented for the 2011 Writing NAEP. The addition of this procedure was recommended by the Governing Board staff, and the Technical Advisory Committee for Standard Setting approved the plan presented by Measured Progress (the Writing ALS contractor) for incorporating the training step into the procedure. The paper selection procedure provides an opportunity for panelists to review samples of student work, judge the level of achievement most closely matched to the performance exhibited in the paper, and discuss their judgments with others in the group. After all the papers have been discussed, panelists are then given the scores for the papers. Scores are presented to help panelists understand that there is no systematic correspondence between the rubric scores and the achievement levels represented by the student responses.

The field trial was conducted on January 27, 2012 with 39 panelists: 19 for grade 8 and 20 for grade 12, all recruited within a 50-mile radius of the meeting location in Dover, New Hampshire. This was a sample of convenience in that most panelists serve or had served as scorers for Measured Progress. Panelists included the three types required for the NAEP ALS process: grade-level teachers, non-teacher educators including higher education faculty in the subject matter, and representatives of the general public who are trained in or work in the subject matter
The distribution of panelists in each of the three categories was approximately the same as that required for achievement levels setting: 55% teachers, 15% non-teacher educators, and 30% general public representatives. The study was conducted for the 2011 Writing NAEP for grades 8 and 12. Panelists made only one round of judgments in which they classified 50 student booklets into achievement levels categories.

The results of this study using the revised achievement levels descriptions and incorporating the paper selection process as additional training were generally similar to those for the pilot study. For grade 8, the classifications in this single round by field trial panelists resulted in 4% fewer students performing at or above the Basic level than was the case for the pilot study, about 16% fewer performing at or above the Proficient level, and 2% more at the Advanced level. The most noticeable difference was a sizable increase (12%) in the percentage of 12th grade students that would be classified at the Advanced level. The percentage performing at or above the grade 12 Basic level was the same, and the percentage performing at or above the Proficient level increased by 9% over the level for the pilot study.

**Operational Achievement Levels-Setting Study**

The operational achievement levels-setting study was implemented in St. Louis February 7-10, 2012, with a special study added at the end implemented February 10-11, 2012. The goal was to recruit 30 panelists for each grade level to include 55% writing teachers at the grade level; 15% writing educators, such as college writing instructors, curriculum directors, and other educators who are not currently teaching writing in the K-12 system; and 30% general public representatives who are trained in writing and or currently employed in a position requiring a significant writing component. The content facilitators considered the panelists to be exceptionally strong in terms of their writing experience and knowledge of writing. The panelists included a mayor and several authors as well as outstanding teachers of writing and leaders of writing programs. The panelists for the operational achievement levels-setting study are described in the Table 1 below.

Two additional modifications were made for the operational ALS process. For previous achievement levels-setting studies for the 2011 Writing NAEP, student booklets were presented to panelists in order from lowest student performance to highest. The Technical Advisory Committee for Standard Setting concurred with the Governing Board staff recommendation to change the order of presentation from highest to lowest performance. Panelists were told that they could classify booklets in any order, and they were encouraged to look at one or two booklets at a few locations across the range before they began classifying. The second modification was to have panelists record their level of confidence in the classification of each booklet for the second and third classification. The second classification was made with the same set of booklets used for round 1, but the third classification involved a completely new set of booklets. Staff recommended that the confidence rating be collected to provide information of the relative confidence in the two sets of ratings, as well as to provide data for further research regarding this methodology because the Body of Work method had not been used previously in NAEP achievement levels-setting. Again, this recommendation was discussed with the Technical Advisors, and they approved of the plan.
Special Study: A special study was conducted at the conclusion of the achievement levels-setting study, February 10-11, 2012. The purpose of the study was to produce estimates of the performance of students on the 2007 Writing NAEP comparable to that on the 2011 Writing NAEP.

All panelists were invited to participate in the study, but unavailability to participate in the special study did not preclude participation on the ALS panel. A total of 35 panelists (approximately 2/3 of the ALS count) agreed to participate: 17 for grade 8 and 18 for grade 12. The distribution of panelists participating in the special study is presented in Table 2 below.
Table 2
NAEP Writing 2011 Achievement Levels Setting Panelists for Special Study with 2007 Writing NAEP

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Attributes</th>
<th>Grade 8</th>
<th>Grade 12</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Panelist Type</td>
<td>Teachers</td>
<td>8</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Non-teacher Educators</td>
<td>4</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>General Public</td>
<td>5</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>13</td>
<td>76</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>4</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>Caucasian</td>
<td>15</td>
<td>88</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Non-Caucasian</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>NAEP Region</td>
<td>Midwest</td>
<td>5</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Northeast</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>South</td>
<td>4</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>7</td>
<td>41</td>
<td>6</td>
</tr>
</tbody>
</table>

The special study was designed to provide data for comparing performance on the 2007 Writing NAEP to that on the 2011 NAEP when judged relative to the 2011 achievement levels descriptions. The special study was implemented following the pilot study; but because the achievement levels descriptions were modified for use in the operational ALS, it was necessary to implement the study design again using the revised descriptions.

Closed Session March 2, 2012
Outcomes of the achievement levels-setting study and the special study will be reviewed by the Technical Advisory Committee on Standard Setting at a meeting scheduled for February 23-24, 2012. Their recommendations, along with complete study results for the added field trial, the operational ALS, and the special study will be shared with COSDAM for discussion at the meeting on March 2, 2012.

The current schedule calls for a recommendation on 2011 writing achievement levels by COSDAM for action by the Governing Board at the May 2012 meeting in San Antonio.
WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 4

BASIC

Fourth-grade students writing at the Basic level should be able to address the tasks appropriately and at least partially accomplish their communicative purposes. Texts should be appropriately structured. Many of the ideas in the texts should be developed, and their texts should include supporting details and examples that are relevant to the topic, purpose, and audience. Most sentences should be well structured, and texts may be composed mostly of simple sentences. Many of the words and phrases should be appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation skills should be sufficiently accurate to convey general meaning, although there may be some errors that detract from meaning.

PROFICIENT

Fourth-grade students writing at the Proficient level should be able to address the tasks appropriately and accomplish their communicative purposes. Texts should be appropriately structured and coherent. Most of the ideas in their texts should be developed effectively, and their texts should include supporting details and examples that support the main ideas. Texts should have well structured sentences and a variety of sentence types—simple, compound, and complex. Words and phrases should be thoughtfully selected and appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation should be sufficiently accurate to communicate clearly with the reader. There may be some errors in the texts, but these errors should not impede meaning.

ADVANCED

Fourth-grade students writing at the Advanced level should be able to address the tasks appropriately and accomplish their communicative purposes in effective ways. Texts should be well structured and coherent. The ideas in the texts should be developed fully and effectively. Their texts should include supporting details and examples that are closely related to the topic, purpose, and audience and that enhance communicative effectiveness. Sentences should be well structured, and texts should include a variety of sentence types (simple, compound, and complex) to enhance their communicative effectiveness. Words and phrases should be chosen skillfully, and they should both enrich meaning in the texts and enhance communicative effectiveness. Spelling, grammar, usage, capitalization, and punctuation should be mostly accurate and well developed, and they should be used appropriately. Grammatical, mechanical, and usage choices should contribute to communicative effectiveness. There may be a few errors, but they should not impede meaning.
WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 8

BASIC

Eighth-grade students writing at the Basic level should be able to address the tasks appropriately and mostly accomplish their communicative purposes. Their texts should be coherent and effectively structured. Many of the ideas in their texts should be developed effectively. Supporting details and examples should be relevant to the main ideas they support. Voice should align with the topic, purpose, and audience. Texts should include appropriately varied uses of simple, compound, and complex sentences. Words and phrases should be relevant to the topics, purposes, and audiences. Knowledge of spelling, grammar, usage, capitalization, and punctuation should be made evident; however, there may be some errors in the texts that impede meaning.

PROFICIENT

Eighth-grade students writing at the Proficient level should be able to develop responses that clearly accomplish their communicative purposes. Their texts should be coherent and well structured, and they should include appropriate connections and transitions. Most of the ideas in the texts should be developed logically, coherently, and effectively. Supporting details and examples should be relevant to the main ideas they support, and contribute to overall communicative effectiveness. Voice should be relevant to the tasks and support communicative effectiveness. Texts should include a variety of simple, compound, and complex sentence types combined effectively. Words and phrases should be chosen thoughtfully and used in ways that contribute to communicative effectiveness. Solid knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. There may be some errors, but these errors should not impede meaning.

ADVANCED

Eighth-grade students writing at the Advanced level should be able to construct skillful responses that accomplish their communicative purposes effectively. Their texts should be coherent and well structured throughout, and they should include effective connections and transitions. Ideas in the texts should be developed logically, coherently, and effectively. Supporting details and examples should skillfully and effectively support and extend the main ideas in the texts. Voice should be distinct and enhance communicative effectiveness. Texts should include a well-chosen variety of sentence types, and the sentence structure variations should enhance communicative effectiveness. Words and phrases should be chosen strategically, with precision, and in ways that enhance communicative effectiveness. An extensive knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. Appropriate use of these features should enhance communicative effectiveness. There may be a few errors, but these errors should not impede meaning.
WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 12

BASIC

Twelfth-grade students writing at the Basic level should be able to respond effectively to the tasks and accomplish their communicative purposes. Their texts should be coherent and well structured. Most of the ideas in their texts should be developed effectively. Relevant details and examples should be used to support and extend the main ideas in the texts. Voice should support the communicative purposes of the texts. Texts should include appropriately varied simple, compound, and complex sentence types. Words and phrases should be suitable for the topics, purposes, and audiences. Substantial knowledge of spelling, grammar, usage, capitalization, and punctuation should be clearly evident. There may be some errors in the texts, but these errors should not generally impede meaning.

PROFICIENT

Twelfth-grade students writing at the Proficient level should address the tasks effectively and fully accomplish their communicative purposes. Their texts should be coherent and well structured with respect to these purposes, and they should include well-crafted and effective connections and transitions. Their ideas should be developed in a logical, clear, and effective manner. Relevant details and examples should support and extend the main ideas of the texts and contribute to their overall communicative effectiveness. Voice should be relevant to the tasks and contribute to overall communicative effectiveness. Texts should include a variety of simple, compound, and complex sentence types that contribute to overall communicative effectiveness. Words and phrases should be chosen purposefully and used skillfully to enhance the effectiveness of the texts. A solid knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. There may be some errors in the texts, but they should not impede meaning.

ADVANCED

Twelfth-grade students writing at the Advanced level should be able to address the tasks strategically, fully accomplish their communicative purposes, and demonstrate a skillful and creative approach to constructing and delivering their messages. Their texts should be coherent and well structured; they should include skillfully constructed and effective connections and transitions; and they should be rhetorically powerful. All of the ideas in their texts should be developed clearly, logically, effectively, and in focused and sophisticated ways. Supporting details and examples should be well crafted; they should skillfully support and extend the main ideas; and they should strengthen both communicative effectiveness and rhetorical power of the texts. A distinct voice that enhances the communicative effectiveness and rhetorical power of the texts should be evident. Texts should include a variety of sentence structures and types that are skillfully crafted and enhance communicative effectiveness and rhetorical power. Words and phrases should be chosen purposefully, with precision, and in ways that enhance communicative effectiveness and rhetorical power. A highly developed knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts and function in ways that enhance communicative effectiveness and rhetorical power. There may be a few errors in the texts, but they should not impede meaning.
**National Assessment Governing Board**  
**Reporting and Dissemination Committee**  

**March 2, 2012**  
**9:45 a.m. - 12:15 p.m.**

**AGENDA**

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Presenter(s)</th>
<th>Attachment</th>
</tr>
</thead>
</table>
| 9:45 – 10:20 am | NAEP Background Questions:  
Expert Panel Report and Recommendations  
*Marshall S. Smith, Panel Chair* |                                                   | Attachment A |
| 10:20 – 10:45 am | Implementation of Making a Difference Initiatives:  
Speakers’ Tool Kit, Presentation for Parents, and Focused Reports  
*Amy Buckley, Reingold Communications* |                                                   | Attachment B |
| 10:45 – 11:05 am | Private School Participation and Reporting  
*Arnold Goldstein, NCES* |                                                   | Attachment C |
| 11:05 – 11:15 am | Review of Recent NAEP Release: TUDA 2011  
*Stephaan Harris, NAGB  
Amy Buckley, Reingold Communications* |                                                   | Attachment D |
| 11:15 – 11:25 am | Projected Schedule for Future NAEP Reports  
*Arnold Goldstein, NCES* |                                                   | Attachment E |
| 11:25 – 11:35 pm | **ACTION:** Release Plan for NAEP 2011 Science Report Card  
*Stephaan Harris, NAGB* |                                                   | Attachment F |
| 11:35 am – 12:15 pm | Policy on NAEP Testing and Reporting on Students with Disabilities and English-Language Learners  
a. Implementation Update  
*Arnold Goldstein, NCES*  
b. Letter from Florida Education Commissioner  
*Larry Feinberg, NAGB* |                                                   | Attachment G |
NOTE TO Reporting and Dissemination Committee
on Expert Panel on NAEP Background Questions

For the past 25 years the National Assessment of Educational Progress (NAEP) has asked hundreds of background or noncognitive questions of the students, teachers, and schools in its samples. These are meant to enrich the reporting of NAEP’s academic results but for more than a decade little use has been made of them in NAEP reports. Responses to the background questions are available through the NAEP Data Explorer on the Internet, but they have gathered little attention and had little impact despite the considerable efforts that have gone into collecting and tabulating this information.

PURPOSE OF THE EXPERT PANEL

Following a discussion by the Reporting and Dissemination Committee in August 2011, the expert panel was convened by Board staff to recommend how to make better use of existing NAEP background questions and to propose an analytic agenda or framework for additional topics and questions that would be useful in developing education policy and of value to the public.

The panel’s deliberations and recommendations support two of the Board’s statutory responsibilities: to select and approve NAEP background questions, and to improve the form, use, and reporting of the National Assessment. (P.L. 107-279) http://www.nagb.org/who-we-are/naep-law.htm.

COMPOSITION

The panel has six members, composed of persons familiar with NAEP and its background questions. Panel members have expertise in survey research, educational testing, and education policy analysis.

DELIBERATIONS AND REPORT

The panel held an initial all-day meeting in Washington, DC on November 16, 2011. The agenda included extensive briefings by staff of the National Center for Education Statistics (NCES) and ETS, the contractor that prepares and analyzes NAEP background questionnaires. It also included discussion of how background questions are collected and used both in NAEP and in the international assessments, PISA and TIMSS. After that, the panel met by teleconference, and members exchanged drafts and comment.

The panel chair, Marshall S. Smith, will present the final report at this meeting of the Reporting and Dissemination Committee. He will discuss it with the full Board in the afternoon.

The report appears under a separate tab in the agenda book.
EXPERT PANEL ON NAEP BACKGROUND QUESTIONS

MEMBERSHIP

Marshall S. Smith (Chairman)
Visiting Scholar, Carnegie Foundation for the Advancement of Teaching
Former U.S. Under Secretary of Education
Palo Alto, CA

Naomi Chudowsky
Co-Director
Caldera Research, LLC
Bend, OR

Alan Ginsburg
Education Consultant
Former Director of Policy and Program Evaluation
U.S. Department of Education
Washington, DC

Robert Hauser
Executive Director
Division of Behavioral and Social Sciences and Education
National Research Council
Washington, DC

Jennifer Jennings
Assistant Professor of Sociology
New York University
New York, NY

Sharon Lewis
Research Director
Council of the Great City Schools
Washington, DC
Background Information

Framework for the

National Assessment

of Educational Progress

EXCERPTS

National Assessment Governing Board
Adopted August 1, 2003
National Assessment Governing Board

Darvin M. Winick
Chair

John H. Stevens
Chair, Ad Hoc Committee on Background Questions

Charles E. Smith
Executive Director

Lawrence Feinberg
Project Officer

Background Framework Project

Mary Lyn Bourque, Writer/Editor
Mid-Atlantic Psychometric Services

David Grissmer
RAND

Consultants

Paul E. Barton
Ina V.S. Mullis
Michael Podgursky
Mark Reckase
Herbert J. Walberg
Preface

by the National Assessment Governing Board

The National Assessment of Educational Progress (NAEP) has been established by law to monitor the academic achievement of American students. In addition to its academic assessments, NAEP has collected information from hundreds of non-cognitive or background questions about students, their educational experiences in class and at home, their teachers, and their schools. Some of these questions provide data for NAEP’s reporting categories, but far more have been used to give context to NAEP results or to track factors associated with academic achievement. Some have been used by scholars in social science research.

Concerns have been raised about the selection of background variables, the quality of the information obtained, and the validity of inferences drawn from it. There is also concern about the burden that collecting background information places on respondents and on the NAEP program. After the National Assessment Governing Board was granted final authority over the background questions in early 2002, it adopted a policy to focus NAEP background data on the primary purpose of the National Assessment—to provide sound, timely information on the academic achievement of American students. The Board also initiated a process to prepare a general framework to guide the collection and reporting of background data.

It is important to understand the National Assessment is not designed to prove cause-and-effect relationships; it cannot prescribe what should be done. But its descriptions of the educational circumstances of students at various achievement levels—considered in light of research from other sources—may provide important information for public discussion and policy action.

This framework will define the purpose and scope of NAEP’s system of collecting background information, including background questionnaires and other sources of non-cognitive data. It will establish criteria for reporting background information as part of the National Assessment. The approach it suggests provides for asking various groups of questions to various samples of students at various times.

The framework reflects the following key principles:
The selection of background topics and questions shall be designed to fulfill all legal requirements for the National Assessment and to carry out decisions regarding what NAEP will report and how to report it.

Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be “directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information.”

The collection of background data shall be designed to obtain information that is objective, valid, reliable, and of consistently high quality.

The system of background data collection shall be efficient and designed to minimize the burden on respondents and on the NAEP program. As much data as possible should be obtained from school records and other reliable data sources.

These principles shall apply both to the collection of general background information and to subject-specific background questions. The frameworks for the latter must be focused and prioritized, indicating a core set of variables for regular reporting and a more comprehensive set to be collected and reported less frequently.

The priority order for background information is as follows: (1) reporting categories, as required by law; (2) contextual factors with a well-established relationship to achievement; and (3) subject-specific information.

There is one other consideration—the new role of the National Assessment in the No Child Left Behind Act of 2001. Under this law, all states receiving federal Title I aid are required to participate every two years in NAEP’s state-level samples of reading and mathematics in grades 4 and 8. The results will provide an independent yardstick to compare trends on NAEP with performance on each state’s own set of required exams.

Because No Child Left Behind places particular emphasis on closing the persistent performance gaps between various student groups, NAEP must be able to report on changes in achievement for all groups specified by law. Through its background questions, the National Assessment might also provide useful information about the students left behind and those who are ahead of them, including the sorts of
schools that high-achieving and low-achieving students attend, the courses they take, the patterns of how they are taught, and the qualifications of their teachers. Over time, such descriptive information will allow NAEP to track changes in contextual and instructional factors related to student achievement and in the distribution of important educational resources.

In sum, the purpose of this Background Information Framework is to focus the collection and reporting of background data by the National Assessment and to establish clear priorities and limits. We hope to make it possible that with far fewer non-cognitive questions than it has had in the recent past, NAEP will serve the purposes of law and provide the American public and decision makers with useful information. We are committed to improving the quality of data collected and the reporting of results.
Executive Summary

The National Assessment of Educational Progress (NAEP) is a federally authorized survey of student achievement at grades 4, 8, and 12 in various subject areas, such as mathematics, reading, writing, science, U.S. history, the arts, and foreign languages. The No Child Left Behind Act of 2001 (P.L. 107-110) requires the assessment to collect data on specified student groups, including race/ethnicity, gender, socio-economic status, disability, and limited English proficiency. It requires fair and accurate presentation of achievement data and permits the collection of background or descriptive information that is related to academic achievement and aids in fair reporting of results. The intent of the law is to provide representative-sample data on student achievement for the nation, the states, and subpopulations of students and to monitor progress over time.

The National Assessment Governing Board (NAGB) sets policy for NAEP and determines the content framework for each assessment. As a result of the No Child Left Behind Act, the Board is responsible for selecting and approving all of NAEP’s non-cognitive or background questions, as well as the cognitive items over which it has had final authority since 1988. This Background Information Framework will guide the development and selection of non-cognitive topics and questions, starting with the NAEP 2006 assessment. It will fulfill the purposes of law and implement Board policy.

When NAEP began in 1969-70, its background information was limited to gender, race/ethnicity, and literacy materials at home. During the 1980s the array of non-cognitive questions expanded greatly, both to provide more contextual information and in an effort—never fully realized—to use the assessment for educational research.

This background data framework will refocus the collection of non-cognitive variables on NAEP’s primary mission: providing a fair and accurate measure of student achievement and on achievement trends over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results. NAEP may contribute to research into improving education policy and practice, but its role in this respect is limited and the framework is not a comprehensive list of possible factors to explore.
Since by law NAEP may only collect information that is “directly related to the appraisal of academic achievement,” it must concentrate on non-cognitive variables that are known from other research to have such a relationship. The law also specifically prohibits NAEP from asking about personal or family beliefs and attitudes. These points are emphasized in the Governing Board Policy Statement on the Collection and Reporting of Background Data by the National Assessment (adopted on May 18, 2002). That policy is incorporated into this framework. It is attached in the appendix.

PRIORITIES

The following priorities for collecting and reporting non-cognitive information should be followed in planning background questionnaires, the frequency with which questions are asked, and the samples from which data are collected.

1. **Student reporting categories** that are required by law must be collected as a regular component of all NAEP assessments. These include race, ethnicity, gender, socioeconomic status, disability, and limited English proficiency. A core of SES information should be collected in every assessment, such as type of community and poverty status. An expanded set of SES variables may be included periodically or administered to limited samples.

2. **Other factors that provide a context for results** should be sampled periodically, or on a rotating basis, over several NAEP cycles, although a limited set may be asked in every assessment. Contextual factors may include courses taken, student mobility, school safety and discipline, teacher-related factors such as demographics and experience, other factors related to students and schools, and educationally-relevant variables outside school. Although many non-cognitive variables may be of interest, they must be limited to meet the needs of NAEP reporting. In all cases, they must be clearly related to academic achievement or to the fair presentation of achievement results.

3. **Subject-specific background information** should be gathered at the same time that achievement in a subject is assessed. This may include relevant course content and requirements, teacher preparation, and other factors related to student achievement. Questions will not be designed to determine effective practices, but to show patterns and trends of factors of interest, based on previous research. Like the contextual information, most of these variables should be sampled periodically, or on a rotating basis, over several administrations of the subject exam, although a limited core set may be repeated every time the assessment is given.
SELECTION CRITERIA

Key criteria for selecting non-cognitive topics and questions are as follows:

- **Does the current or proposed non-cognitive variable relate to the primary purpose of NAEP and how?** The primary purpose of NAEP is to report on the academic achievement of students to the American public. It is not to report on the causes of that achievement. Other surveys with longitudinal data are far better suited to examining causality. NAEP’s choice of which non-cognitive variables to measure should be guided by how and to what extent the variables selected will support NAEP’s primary mission.

- **Do the current or proposed non-cognitive variables meet professional standards for reliability and validity?** The NAEP legislation requires that the assessment “use widely accepted professional testing standards (P.L. 107-110, Sec. 411 (b) (5)).” This requirement applies equally to non-cognitive and academic variables.

- **How stable is the non-cognitive variable from period to period?** If a variable shows little change from year to year, it should be reviewed to determine whether it should be deleted or used on a periodic basis rather than in every assessment.

- **If new questions are added, have others been deleted in order to limit the burden and expense of NAEP’s background questionnaires?** There will always be pressure to collect more information. Mechanisms must be developed to make sure the burden of background questionnaires does not expand over time.

- **Does a question address specific behavior rather than conclusions?** Even for such questions, however, caution is advisable because self-reports are often unreliable.

- **Will the topic or question meet the test of broad public acceptability and not be viewed as intrusive or prying?** NAEP’s non-cognitive questions are not kept secure, and all of them are to be posted on the Internet. Possible objections should be considered in deciding whether or not a question will be asked.

- **Does the topic or question deal with a factor in which trends over time are important?**
• **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.

### DATA COLLECTION

Whenever possible, NAEP should use information from school records and other reliable data collections in order to improve the validity of the information collected and limit the background questionnaires in NAEP itself. In exploring the utility of different data sources, the following criteria should be considered: (1) reliability, (2) universality, (3) currency, (4) respondent burden, (5) logistics, (6) efficiency and cost-effectiveness, and (7) the impact on timeliness of NAEP reporting.

Of the student reporting categories in Priority 1, information on gender, race/ethnicity, disability status, and limited English proficiency shall be collected in a uniform manner in all NAEP samples. NAEP is also required to collect information on socio-economic status. This will continue to be done in all samples, although there may be some variation in the number of factors on which data are obtained with a uniform core and more extensive data gathering in some cases.

Because socio-economic status cannot be measured simply or directly, NAEP has used “proxy” variables, such as eligibility for free or reduced-price lunch (a measure of poverty), parent education, and number of reading materials in the home. The framework provides that NAEP explore development of a composite index for SES derived from the proxy variables currently collected. To the extent that the index can be sharpened by additional data from readily available sources, such as zip codes and census, this option should also be considered. Occasionally and in limited samples, more extensive SES questions may be asked. Although NAEP may never be able to produce a full composite of SES, based on family income, education, and occupation, efforts should be made to find an approximation that is more informative than the current set of proxy variables.

For the past two decades, NAEP has collected information on a lengthy list of student, teacher, school, and beyond-school factors that may provide a context for achievement results and are of interest to policymakers, researchers, and the public. Yet, NAEP’s design as a cross-sectional survey places serious limitations on the inferences that can properly be drawn from this information. We propose a careful review of the contextual factors in NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. Different questions should be cycled in and out of the assessment
periodically, and the use of data from non-NAEP sources should increase. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.

The collection of subject-specific background information should be focused, limited, and prioritized as part of the subject-matter frameworks adopted by the Board. For each subject there should be a small core set of background items administered to the full sample each time a subject is assessed. An additional, more comprehensive set of questions should be administered periodically or to smaller subsamples.

NCES will prepare for Board review and approval a plan indicating the frequency, sample size, and schedule of rotation for all background variables and questions on which information is to be collected by NAEP. This should include both questionnaires and alternate data sources to obtain core reporting data, subject-specific information, and data on achievement-related contextual variables from a variety of NAEP samples—national only, national and state, and a subset of the national sample. The plan should indicate the frequency and schedule of rotation for each of the questions proposed. It should also indicate any questions needed for quality control purposes. The recommendations should be prepared with input from researchers and state policy analysts, as appropriate, and updated on a regular basis.

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. As much data as possible should be obtained from school records and other reliable data sources. The average individual response time to answer background questionnaires for each assessment, as calculated in accordance with Office of Management and Budget (OMB) procedures, shall be limited as follows: ten minutes for each student, 20 minutes for each teacher, and 30 minutes for each school.

**REPORTING**

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on patterns and trends of non-cognitive variables known to have a relationship to academic achievement and may contain disaggregated data on school conditions and practices for various groups of students. Data on courses taken before NAEP assessments (either from transcripts or questionnaires) is of great public interest and can be related to academic results.
In addition, supplemental reports may be prepared that focus on particular aspects of the background data collected. In all cases, NAEP reports published by the National Center for Education Statistics must not state conclusions as to cause and effect relationships and avoid simplistic presentations that imply best practice.

All background questions and data collected by NAEP should be posted on the Internet so the public may be able to consider them in discussing results. Complete data files should be made available to researchers for further analysis.

RESEARCH

As a cross-sectional survey without longitudinal data, the National Assessment is able to document school conditions and practices. It can report on achievement results. But it cannot properly be used to establish direct cause-and-effect relationships. Still, over the past three decades, NAEP has been part of two important research endeavors—exploring changes in the black-white test score gap since 1970 and seeking to establish the impact of state-level reforms during the 1990s. By monitoring achievement well, NAEP has provided sound data for researchers to use. NAEP results have been critical in identifying research hypotheses. Its large data sets have been combined with other information to tease out meaning and policy implications, though NAEP’s own reports have properly steered clear of these activities.

The Governing Board believes that by doing its main task of monitoring educational achievement well NAEP can make a valuable contribution to education research. The NCES program of secondary analysis grants for researchers to analyze NAEP data should continue. Educational researchers should be involved, under the auspices of NCES, in developing NAEP background questionnaires, validity studies, and other data collection efforts to carry out the provisions of this framework.

The primary purpose of NAEP is to provide fair and accurate information on student achievement. Its primary audience is the American public. The Governing Board believes that in serving its purpose and audience well, NAEP can contribute to educational research. It welcomes the interest and efforts of researchers.
Fourth-graders who read for fun almost every day score higher

Students were asked how often they read for fun on their own time. Students selected one of four responses indicating “never or hardly ever,” “once or twice a month,” “once or twice a week,” or “almost every day.” In 2011, fourth-graders who reported reading for fun almost every day scored higher on average than those who did so less frequently, and students who reported never or hardly ever reading for fun scored lowest (figure 12).

Figure 12. Average scores in fourth-grade NAEP reading, by students’ responses to a question about how often they read for fun on their own time: 2011

Forty-six percent of fourth-graders reported reading for fun almost every day in 2011 (table 6). Since students were asked the same question in some of the earlier assessment years, the percentages can be compared over time. The percentage of students who reported reading for fun almost every day was higher in 2011 than in all the previous assessment years, and the percentage who reported never or hardly ever reading for fun was lower in 2011 than in all other years.

Table 6. Percentage of students assessed in fourth-grade NAEP reading, by how often they read for fun on their own time: Various years, 2002-11

<table>
<thead>
<tr>
<th>Frequency of reading for fun</th>
<th>2002</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never or hardly ever</td>
<td>15*</td>
<td>15*</td>
<td>16*</td>
<td>18*</td>
<td>15*</td>
<td>14</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>14*</td>
<td>15*</td>
<td>15*</td>
<td>16*</td>
<td>15*</td>
<td>14</td>
</tr>
<tr>
<td>Once or twice a week</td>
<td>26</td>
<td>25</td>
<td>26*</td>
<td>27*</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Almost every day</td>
<td>45*</td>
<td>45*</td>
<td>43*</td>
<td>40*</td>
<td>44*</td>
<td>46</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.
NOTE: Detail may not sum to totals because of rounding.

The extent to which students reported reading for fun differed by gender, race/ethnicity, and eligibility for NSLP (table 7). In 2011, the percentage of students who reported reading for fun almost every day was

- higher for female students than for male students,
- higher for Asian students than for other racial/ethnic groups, and
- higher for students who were not eligible for NSLP than for students who were eligible.

Table 7. Percentage of students assessed in fourth-grade NAEP reading, by how often they read for fun on their own time and selected student characteristics: 2011

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Never or hardly ever</th>
<th>Once or twice a month</th>
<th>Once or twice a week</th>
<th>Almost every day</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>18</td>
<td>17</td>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>12</td>
<td>25</td>
<td>53</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>15</td>
<td>15</td>
<td>24</td>
<td>46</td>
</tr>
<tr>
<td>Black</td>
<td>17</td>
<td>14</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13</td>
<td>15</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Asian</td>
<td>9</td>
<td>11</td>
<td>26</td>
<td>54</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>18</td>
<td>16</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>Native Hawaiian/Other Pacific Islander</td>
<td>12</td>
<td>17</td>
<td>28</td>
<td>43</td>
</tr>
<tr>
<td>Two or more races</td>
<td>14</td>
<td>14</td>
<td>26</td>
<td>46</td>
</tr>
<tr>
<td><strong>Eligibility for free/reduced-price school lunch</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>15</td>
<td>15</td>
<td>26</td>
<td>44</td>
</tr>
<tr>
<td>Not eligible</td>
<td>13</td>
<td>14</td>
<td>25</td>
<td>47</td>
</tr>
</tbody>
</table>

NOTE: Black includes African American, and Hispanic includes Latino. Race categories exclude Hispanic origin. Detail may not sum to totals because of rounding.

More students have teachers not permitting calculators during mathematics lessons in 2011 than in previous years

Teachers reported on the extent to which they permitted students to use calculators during mathematics lessons. Teachers selected one of three responses indicating “unrestricted use,” “restricted use,” or “calculators are not permitted.”

Sixty-two percent of fourth-graders had teachers who reported permitting the restricted use of calculators in 2011 (Table 6). Because teachers were asked the same question as part of the 2005, 2007, and 2009 assessments, the percentages can be compared over time. A higher percentage of students had teachers who did not permit the use of calculators in 2011 than in earlier assessment years, while the percentage permitting restricted use was lower in 2011 than in earlier years.

Table 6. Percentage of students assessed in fourth-grade NAEP mathematics, by the extent of calculator use in mathematics lessons: Various years, 2005–11

<table>
<thead>
<tr>
<th>Extent of calculator use</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted use</td>
<td>5*</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Restricted use</td>
<td>75*</td>
<td>69*</td>
<td>67*</td>
<td>62</td>
</tr>
<tr>
<td>Calculators are not permitted</td>
<td>20*</td>
<td>27*</td>
<td>29*</td>
<td>34</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.

The extent to which students had teachers who permitted calculator use for mathematics lessons was different for those who were or were not eligible for NSLP. The percentage of students whose teachers permitted restricted use of calculators was higher for students who were not eligible for NSLP than for students who were eligible, and the percentage of students whose teachers did not permit them to use calculators was higher for eligible students (Figure 12).

Figure 12. Percentage of students assessed in fourth-grade NAEP mathematics, by eligibility for free/reduced-price school lunch and extent of calculator use in mathematics lessons: 2011

In 2011, students whose teachers permitted restricted use of calculators during mathematics lessons scored higher on average than students whose teachers allowed unrestricted use or did not permit the use of calculators (figure 13).

**Figure 13.** Average scores in fourth-grade NAEP mathematics, by teachers’ responses to a question about the extent to which their students use calculators during mathematics lessons: 2011

To what extent are students permitted to use calculators during mathematics lessons?

- Unrestricted use: 236
- Restricted use: 243
- Calculators are not permitted: 237

Students who have more frequent class discussions score higher

As part of the 2011 NAEP reading assessment, eighth-graders were asked how often they had class discussions about something their English class had read. Students chose from four options: “never or hardly ever,” “a few times a year,” “once or twice a month,” or “at least once a week.”

In 2011, eighth-graders who reported having class discussions more frequently scored higher on average than those who reported doing so less frequently (figure 28). For example, the average score for students who reported having discussions at least once a week was higher than the score for students who did so once or twice a month. Those who reported never or hardly ever having discussions scored lowest.

Figure 28. Average scores in eighth-grade NAEP reading, by students’ responses to a question about how often they had an English class discussion about something the whole class read during the school year: 2011

Forty-eight percent of students reported having class discussions at least once a week in 2011, which was higher than the percentage in 2002 (table 13). The percentage of students who reported having class discussions a few times a year was also higher in 2011 than in 2002, while the percentages of students who reported never or hardly ever having discussions, or doing so once or twice a month, were lower in 2011 than in 2002.

Table 13. Percentage of students assessed in eighth-grade NAEP reading, by how often they had an English class discussion about something the whole class has read during the school year: Various years, 2002-11

<table>
<thead>
<tr>
<th>Frequency of class discussion</th>
<th>2002</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never or hardly ever</td>
<td>12*</td>
<td>12*</td>
<td>11*</td>
<td>11*</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>A few times a year</td>
<td>13*</td>
<td>13*</td>
<td>13*</td>
<td>14*</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>30*</td>
<td>30*</td>
<td>29*</td>
<td>30*</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>At least once a week</td>
<td>45*</td>
<td>45*</td>
<td>46*</td>
<td>45*</td>
<td>47</td>
<td>48</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.

NOTE: Detail may not sum to totals because of rounding.

The extent to which students had class discussions about something they had read differed by school type, location, and enrollment (table 14). In 2011, the percentages of students who reported having class discussions at least once a week were

- lower for students attending public schools than for those attending private schools,
- higher for students attending schools in city and suburban locations than for those attending schools in town or rural locations, and
- higher for students attending schools with enrollments of 1 to 399 students than with larger school enrollments.

Table 14. Percentage of students assessed in eighth-grade NAEP reading, by how often they had an English class discussion about something the whole class read during the school year and selected student characteristics: 2011

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency of class discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never or hardly ever</td>
</tr>
<tr>
<td><strong>Type of school</strong></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>11</td>
</tr>
<tr>
<td>Private</td>
<td>8</td>
</tr>
<tr>
<td><strong>School location</strong></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>10</td>
</tr>
<tr>
<td>Suburb</td>
<td>10</td>
</tr>
<tr>
<td>Town</td>
<td>12</td>
</tr>
<tr>
<td>Rural</td>
<td>11</td>
</tr>
<tr>
<td><strong>School enrollment</strong></td>
<td></td>
</tr>
<tr>
<td>1-399</td>
<td>10</td>
</tr>
<tr>
<td>400-599</td>
<td>11</td>
</tr>
<tr>
<td>600-799</td>
<td>10</td>
</tr>
<tr>
<td>800-999</td>
<td>10</td>
</tr>
<tr>
<td>1000 or more</td>
<td>11</td>
</tr>
</tbody>
</table>

NOTE: Detail may not sum to totals because of rounding.

One-third of students taking algebra I in eighth grade

Eighth-graders participating in the 2011 NAEP mathematics assessment were asked what math class they were taking that year. Students selected one course from the following list:

- Geometry
- Algebra II
- Algebra I (one-year course)
- First year of a two-year Algebra I course
- Second year of a two-year Algebra I course
- Introduction to algebra or pre-algebra
- Basic or general eighth-grade math
- Integrated or sequential math
- Other math class

Thirty-four percent of eighth-graders reported taking algebra I (one-year course) in 2011, which was higher than the percentages of students who reported taking each of the other types of mathematics classes listed (Table 11). The next highest percentage of students reported taking basic or general mathematics followed by those taking an introductory algebra class.

The percentage of students who reported taking algebra I in 2011 was not significantly different from 2009 but was higher than the percentage who reported taking it in 2005. The percentage of students who reported taking an introductory algebra class was lower in 2011 than in 2009 and 2005. There has been no significant change in the percentage of students taking a basic or general mathematics class.

Table 11. Percentage of students assessed in eighth-grade NAEP mathematics, by the type of mathematics class taken during the school year: Various years, 2005-11

<table>
<thead>
<tr>
<th>Type of class taken</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometry</td>
<td>4*</td>
<td>4*</td>
<td>4*</td>
<td>5</td>
</tr>
<tr>
<td>Algebra II</td>
<td>3*</td>
<td>3*</td>
<td>3*</td>
<td>4</td>
</tr>
<tr>
<td>Algebra I (one-year course)</td>
<td>30*</td>
<td>31*</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>First year of a two-year Algebra I course</td>
<td>3*</td>
<td>3*</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Second year of a two-year Algebra I course</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Introduction to algebra or pre-algebra</td>
<td>27*</td>
<td>27*</td>
<td>25*</td>
<td>23</td>
</tr>
<tr>
<td>Basic or general eighth-grade math</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Integrated or sequential math</td>
<td>1*</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other math class</td>
<td>5*</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.

NOTE: Detail may not sum to totals because of rounding.

Students who reported taking algebra I scored higher on average than students taking an introductory algebra class or a basic or general mathematics class (figure 29). The average score for students who reported taking a basic mathematics class was lower than the score for students taking an introduction to algebra.

**Figure 29.** Average scores in eighth-grade NAEP mathematics, by the type of mathematics class students took during the school year: 2011

The proportions of students taking certain mathematics courses in 2011 varied by race/ethnicity (table 12). For example, with one exception, the percentage of Asian students taking algebra I was higher than the percentages of other racial/ethnic groups (the percentage of Asian students was not significantly different from the percentage of Native Hawaiian/Other Pacific Islander students taking algebra I). The percentage of American Indian/Alaska Native students taking an introductory algebra class was higher than the percentages of other racial/ethnic groups. The percentages of students taking a basic math course were higher for Black, Hispanic, and American Indian/Alaska Native students than for White, Asian, and multiracial students.

**Table 12.** Percentage of students assessed in eighth-grade NAEP mathematics, by race/ethnicity and the type of mathematics class taken during the school year: 2011

<table>
<thead>
<tr>
<th>Type of class taken</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>American Indian/Alaska Native</th>
<th>Native Hawaiian/Other Pacific Islander</th>
<th>Two or more races</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra I (one-year course)</td>
<td>36</td>
<td>28</td>
<td>33</td>
<td>45</td>
<td>24</td>
<td>37</td>
<td>34</td>
</tr>
<tr>
<td>Introduction to algebra or pre-algebra</td>
<td>25</td>
<td>23</td>
<td>20</td>
<td>13</td>
<td>32</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Basic or general eighth-grade math</td>
<td>23</td>
<td>30</td>
<td>29</td>
<td>13</td>
<td>29</td>
<td>26</td>
<td>23</td>
</tr>
</tbody>
</table>

NOTE: Results are not shown for the other types of mathematics classes taken by students. Black includes African American, and Hispanic includes Latino. Race categories exclude Hispanic origin.

Status of Parent Outreach and Engagement

One of the National Assessment Governing Board’s top priorities in 2012 is to better engage parent leaders and organizations and develop communication materials that educate them about the National Assessment of Educational Progress (NAEP) and make its data and resources more accessible tools for parents. The Board’s challenge is to discover and develop ways to make NAEP resources increasingly meaningful and useful to parents, particularly because NAEP does not provide individual student results.

In 2011, the Governing Board formed the Ad Hoc Committee on NAEP Parent Engagement to increase parent awareness about the urgency of improving student achievement and narrowing achievement gaps—particularly by race/ethnicity and income level. In August 2011, prior to its quarterly meeting, the Governing Board convened an outreach event with parent group representatives and other education advocates in the Washington, DC metropolitan area to discuss: 1) the relevance of NAEP for parents; 2) how NAEP results can be communicated to parents to raise awareness of student achievement and achievement gaps; and 3) how parents can use NAEP data to advocate for improved student performance in their own communities and schools.

Since the August outreach event, the Governing Board has been developing draft materials for use with and by parent leaders and parent organizations. Materials developed to date (as of February 2012) include the following:

- PowerPoint presentation that can serve as a master template and be customized for local presentations
- State and urban district assessment profiles that focus on student achievement and incorporate related data such as high school and college completion rates and demographic statistics
- Targeted web page for parents on the Governing Board website (nagb.org) where parents can access materials, tools, resources, and other relevant information.

Initial drafts of these materials were presented to the Ad Hoc Parent Committee at the December Board meeting, and revisions have been ongoing based on their feedback.

The Governing Board is entering a new phase of soliciting and obtaining additional feedback from parent leaders and those involved with developing education data messages for the parent audience. Beginning in February 2012 and continuing into the spring, the
Governing Board is hosting five parent input meetings across the country and convening parent leaders and data experts, to get feedback on the draft materials and discuss ways in which the Governing Board can effectively engage parents with NAEP.

The first input meeting took place on February 14 in Washington, DC, with members of the Parent Education Network (PEN). The eight participants—all local education fund (LEF) constituents representing school districts across the country—provided feedback on the draft materials, shared how the LEFs are using data to communicate with parents, and discussed other approaches and channels for engaging parents. Following this, a second input meeting will be held with the representatives of the National PTA in Washington, DC. On March 7, Board Member Tonya Miles will make a presentation to the National PTA Legislative Conference, using the materials that have been further refined based on the PEN and National PTA meetings. Three additional input meetings will be held—two in Los Angeles and one in Miami—during March and April, aiming to include parents for whom English is a second language.

Based on the collective feedback from all of these parent input meetings, the Governing Board will finalize the draft PowerPoint and related materials, and develop a comprehensive plan for integrating this core suite of materials into Board member speaking opportunities and other communications channels.
At the Board’s December 2011 meeting, the members discussed and prioritized various strategies proposed under Chairman Driscoll’s “Making a Difference” initiative. An idea that received overwhelmingly strong interest among all committees was a Speakers’ Toolkit, which would include the development and distribution of materials about Governing Board initiatives that Board members and staff can use when talking with stakeholder audiences. The toolkit complements a key goal of Chairman Driscoll that is also found in the Board’s Strategic Communications Plan: going beyond the NAEP Report Card releases to extend the life of the data and engage the public and constituent groups on how NAEP can be a resource for them. This key area is critical in helping the Board communicate the importance of NAEP, improve the dissemination of results, encourage broader use of the results, and emphasize the urgency of local action. [The Communications Plan is included at the end of this attachment.]

The Communications Plan recommends the creation of a Speakers’ Toolkit that provides Governing Board members and staff with the materials and information needed to effectively reach these goals. Governing Board members and staff have used template materials and provided presentations to various stakeholder groups, including the Public Education Network, the National PTA and the National Council for the Social Studies.

The toolkit will equip Board members and staff to prepare for and deliver presentations to various stakeholder audiences. The toolkit will include new materials as well as existing presentations, talking points and speeches that Board members and staff have prepared for specific events or meetings such as the release of NAEP Report Cards, Board outreach and NAEP High School Commission symposia events, and parent input meetings. All toolkit materials will be archived and stored on the Governing Board members’ website. Toolkit content will be updated and tailored for each audience and outreach opportunity.

Potential Speakers’ Toolkit materials include:
- PowerPoint slides
- PowerPoint scripts
- Talking points
- One-page overviews or brochures
- Fact sheets and relevant data from NAEP releases
- Handout suggestions
- Photo images
- Recommendations for utilizing the materials with different event formats

Speakers’ Toolkit next steps:
- Conduct interviews to identify the range of uses for the toolkit, desirable tools, logistical organization of resources, and priorities for making them available.
- Identify and vet with a review panel a set of topics and resources to be included in the toolkit and development timelines.
- Develop a prototype toolkit and pilot test it with the review panel and other interested parties.
Incorporate feedback on the prototype and complete toolkit development.

Discuss next steps, such as the process for ongoing maintenance of toolkit content, implementation of the toolkit on the Board members’ website, and launching the toolkit with Governing Board members and other champions.

Status of Focused Reports

During the December 2011 discussion of “Making a Difference” initiatives, another strategy that many Board members deemed a high priority was the development of focused reports – reports that would concentrate on specific NAEP data and trends of potential interest that were not explicitly showcased in NAEP Report Cards.

Focused reports were also a key strategy that the Reporting and Dissemination Committee agreed was a priority initiative. Many Committee members viewed focused reports as a highly desirable effort to increase the usefulness of NAEP data and reporting, and extend NAEP's mission and role in public discussions of important education issues. These reports could analyze or repackage existing data or may be based on new studies with background question modules on particular topics and possibly special research samples.

The Committee requested NCES to provide it with a list of published and planned reports for further discussion at the March 2012 meeting. Focused reports on a variety of NAEP aspects have been released in the recent past, and NCES has plans for several more to be published in the near future. Additionally, Committee members suggested several areas that would benefit from a focused report that could be of wide interest and create a significant impact, including charter schools and digital learning. Board staff have offered several additional suggestions of focused reports for discussion.
# Published Focused Reports Since 1990

<table>
<thead>
<tr>
<th>Publication</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on NAEP: NAEP and the Visual Arts: Framework, Field Test, and Assessment</td>
<td>August 13, 1998</td>
</tr>
<tr>
<td>Focus on NAEP: NAEP and Theatre: Framework, Field Test, and Assessment</td>
<td>August 13, 1998</td>
</tr>
<tr>
<td>Focus on NAEP: NAEP and Music: Framework, Field Test, and Assessment</td>
<td>August 13, 1998</td>
</tr>
<tr>
<td>Focus on NAEP: NAEP and Dance: Framework and Field Tests</td>
<td>August 13, 1998</td>
</tr>
<tr>
<td>Focus on NAEP: New Software Makes NAEP Data User-Friendly</td>
<td>May 12, 1997</td>
</tr>
<tr>
<td>Focus on NAEP: Inclusion of Students from Special Populations</td>
<td>July 31, 1996</td>
</tr>
<tr>
<td>Focus on NAEP: 1994 NAEP Assessment in Geography</td>
<td>December 29, 1995</td>
</tr>
<tr>
<td>Focus on NAEP: 1994 NAEP Assessment in U.S. History</td>
<td>October 30, 1995</td>
</tr>
<tr>
<td>Focus on NAEP: 1994 NAEP Teacher Background Questionnaire</td>
<td>October 19, 1994</td>
</tr>
<tr>
<td>Focus on NAEP: 1994 NAEP Assessment in Reading</td>
<td>March 4, 1994</td>
</tr>
</tbody>
</table>
Planned Future Focused Reports from NCES

1. Focus on NAEP: The NAEP Social Studies Assessments

2. Focus on NAEP: Simpsons Paradox

3. Focus on NAEP: 12th Grade Participation and Engagement

Possible Topics for Focused Reports

1. Computer-delivered education: Digital learning in its many varieties including virtual schools and hybrid or blended courses (that combine online and face-to-face instruction)

2. Charter schools: A Ten-Year Report—2013 compared to 2003 (when a charter school report was last published)

3. School safety and discipline (including suspensions and expulsions)

4. Opportunity-to-Learn: Teachers, Curriculum, and Instruction. How they are distributed by race/ethnicity, socio-economic status, and achievement levels

5. Learning after-school and at-home

6. Education policies and instructional practices of high-performing or high-growth states and districts

7. Gender gaps

8. Black male students

9. Learning in the South: A report across the curriculum on the SREB states

10. Other regional reports: New England states and the Midwest (Big Ten) with NAEP data across the curriculum

11. Private Schools: Achievement and school practices across the curriculum with trends over two decades

12. Eighth-grade algebra: How it has grown and achievement patterns and trends
13. 21st Century Skills: Compile NAEP released questions with student performance data that illustrate widely-desired competencies in communication, collaboration, critical thinking, and creativity (4 C’s)

14. Rural education

**NOTE:** Some of these reports may use data from other NCES and federal government surveys in addition to NAEP background questions and achievement results.
STRATEGIC COMMUNICATIONS PLAN

INTRODUCTION

The theme of this plan is getting beyond the scores and NAEP releases to expand outreach of the Governing Board and NAEP. The two main objectives are:

1. Enhance and elevate the NAEP brand as the gold star of academic assessment and thought leadership in advancing excellence in achievement reporting.

2. Strengthen the relevance and use of NAEP – The Nation’s Report Card – results and NAEP research and resources by existing and new audiences.

To achieve these communications objectives, there will be a new approach that includes the following areas of engagement:

1. Practice consistent, year-round outreach and engagement with stakeholders and audiences.
2. Enhance collaboration with NCES and other entities involved with NAEP.
3. Use multiple communications channels, including social media.
4. Mobilize stakeholders and partners.

The Governing Board defines its audience as the general public. Effective communication requires breaking down the audience into segments based on their level of interest in the Board’s work and education in general, how they might use the Board’s information, and their capacity and tools to influence and effect change. This segmentation will allow resources to be targeted and used most efficiently by delivering messaging and information that are most pertinent to each audience.

The target audiences are as follows:

- **General Public** – the broad grouping of individuals who would be inclined to be receptive to effective messaging and information about NAEP.

- **Education Policymakers** – federal, state, and local officials with responsibility for enacting legislation and policies affecting elementary and secondary education.

- **Higher Education** – educators and administrators of postsecondary institutions, including two- and four-year colleges and trade schools.
- **Business Leaders** – Public and private sector employers, including the military, which are interested in the knowledge and skills of entry-level workers.

- **Education and Workforce Stakeholder Groups** – membership, advocacy and policy groups addressing education and workforce issues.

- **K-12 Teachers** – professionals in public, private, or charter schools who teach K-12.

- **Parents** – families of K-12 students in public, charter, and private schools.

Each of these audiences will require specific messaging and a well-defined “call to action,” such as:

- Learn more about NAEP and the Governing Board.

- Understand how NAEP and the work of the Governing Board may be relevant to issues that are important to you.

- Use NAEP and the Governing Board as a resource in your pursuits.

**COMMUNICATION STRATEGIES**

The Governing Board’s communications plan is built on six distinct but integrated strategies focused on the most effective ways to educate and engage these target audiences. The strategies are designed to leverage the reach and impact of messaging delivered through other organizations, the media, and the Internet. At the same time, they provide the flexibility needed to pursue timely opportunities. Specifically, these strategies will use:

I. **Report Card Releases** – Reinventing the release events to reach broad audiences with greater impact and use the releases as a catalyst for other Board communications efforts.

II. **Stakeholder and Partnership Outreach** – Identifying organizations with valuable contacts and communications vehicles for spreading the Governing Board’s messaging.

III. **Traditional Media** – Using targeted media relations with traditional print and media outlets that provide skilled and trusted educational reporting.

IV. **Social Media** – Identifying and participating in emerging electronic media that reach the Board’s targeted audiences and offer interactive, real-time discussion formats.

V. **Website Development** – Enhancing the site to ensure that target audiences can readily find it, use it as a resource for both learning about and informing education initiatives, and pursue web tactics to increase traffic and impact.

VI. **NAEP Communications Alignment** – NAGB and NCES working together to review branding, materials and outreach.
I. REPORT CARD RELEASE STRATEGY

Release of The Nation’s Report Card will be conducted as part of a comprehensive, integrated communication campaign with a series of planned activities designed to generate traditional news coverage, to disseminate information about the assessment to stakeholder groups, and to further position The Nation’s Report Card as the most trusted national yardstick of student achievement. This can be accomplished through the following:

- **Webinar-style NAEP Releases.** With declining attendance and higher costs of renting venues, the traditional press conference is not giving the Board the best return on its investment. We recommend online webinar releases whenever possible and appropriate for future releases. So panelists can participate via Web-Ex and graphics and presentations can be seen by the viewing public. However, for releases like TUDA and Grade 12 Reading and Mathematics that involve specific cities or states, we leave open the option of having the more traditional style of release in a city or state that would involve local leaders as guests and panelists and add a unique angle to the release.

- **Strategic Release Dates.** Choose Report Card release dates (within the dictates of Board policy and NCES timeline) that optimally use media cycles, coinciding events, and other opportunities to leverage attention so that the release is driven by a date not vice versa.

- **More Accessibility to Media and Other Stakeholders.** The Board can take important and innovative steps to expand Report Card outreach to media and others, by facilitating better access through methods such as:
  - Pursue meetings and deskside briefings with key education journalists to illuminate them on various data, trends, and related efforts.
  - Issue a post-event news release that updates the reactions to NAEP results, gathering some of the best quotes from superintendents, parents, and other stakeholders and using them in another round of outreach to relevant groups.
  - Conduct phone chats with journalists and stakeholders before and after the release to help shape and influence media stories on NAEP.
  - Pitch the participation of event panelists and the Board chair and executive director in online events, including web chats, online forums, or discussion room Q&As with major news organizations such as the Washington Post.

- **Utilizing Web Site and Social Media.** The Board should harness its web site and social media opportunities to extend the life of each Report Card. Several ideas include:
  - Obtain video and audio sound bites of Governing Board staff, members, and other panelists form each event to disseminate to media and post online.
  - In advance of each release, create a “splash” page on the www.nagb.org to host all materials related to the event, including bios of panelists, facts from past and
related releases, information about relevant Board task forces and commissions to build momentum for the event.

- Develop an integrated social media strategy that links to the splash page that will help create a following on social networking sites leading up to the launch.

II. **STAKEHOLDER AND PARTNERSHIP OUTREACH**

As a highly respected, independent source of unique objective data, the Governing Board is an attractive partner for numerous organizations. Relationships are mutually beneficial: the Board gains the support of other respected organizations and another outlet for its message, while the partner’s stature and message are also enhanced. Partnership activities can range from simply establishing website links to publishing reports and newsletters; co-sponsoring workshops, events, and forums; creating awards programs; actively participating in partners’ initiatives and conferences; and disseminating NAEP resources to organizational constituents.

**Potential Partner Types**

- The Media
- Colleges and Universities
- Think Tanks
- Education Advocates
- Parent Groups
- Foundations
- Private Companies
- Minority Advocacy Groups
- Governmental Organizations
- Individuals
- Other Testing Entities

**Recommended Partnership Activities**

Implementing a partnership strategy involves several steps to review, vet and establish the optimal partnership. The following list suggests a handful of specific ideas for activities for the Governing Board to undertake with potential partners. It ranges from big events to daily interactions and demonstrates the cumulative power of partnership development.

This course of action will entail such initial tasks as developing a list of recommended partners and related database; conducting research on priority stakeholders in each audience category; creating a partnership scorecard that identifies the specific opportunity, approach, and outcome for each group; developing partnership outreach materials and other content; and conducting ongoing stakeholder monitoring to identify partnership opportunities.
• **Events**
  - Present NAEP and related issues at education conferences.
  - Join with a teacher group like Teach for America, Phi Delta Kappa, or the National Staff Development Council to hold workshops for teachers on how to use NAEP.
  - Increase partnership with NCES and NAEP State Coordinators and local education groups to host state conferences and/or workshops in states or TUDA districts.
  - Partner with national and local PTAs to hold workshops for parents.

• **Content**
  - Co-sponsor a series of monthly webinars, with a different NAEP-related topic.
  - Create electronic newsletters on Board and NAEP subject-specific topics, using NAEP data and other information.
  - Publish booklets or one-pagers on Board initiatives, task forces, or important topics.
  - Partner with a media outlet or a local university to do background reports on TUDA cities to put the TUDA data in richer context.

• **Other Outreach**
  - Co-sponsor sections on the websites of NAEP partners, such as the Council of Chief state School Officers, and establish linking agreements with each.
  - Create an association of school districts that commit to using NAEP as a resource, partnering with them on assessment matters and making resources available school staff and parents on how NAEP works.
  - Work with the Hechinger Institute (a non-profit organization based at Columbia University that focuses on training education reporters and producing in-depth national and investigative journalism on education) to showcase NAEP as a resource for reporters.
  - Join with a teacher’s group to give an annual award to a district, school, or principal that demonstrates best use of NAEP to improve instruction.

### III. TRADITIONAL MEDIA STRATEGY

The traditional print and broadcast media are important vehicles for public education. However, NAEP coverage in the media has been largely limited to Report Card releases. The extent and value of traditional media coverage can be increased through a number of tactics and tools. These might include media events, a Board directory and experts “tip sheet,” op-eds, a story bank, and improved website usability for the press. The Report Card releases will be used as a catalyst for generating ongoing use of NAEP data in coverage of broader educational policy issues. Ideas include:

- **More Events.** Create additional media events to release new frameworks, for example, or respond to emerging issues, and not just rely on Report Cards to generate news.
Media Training. Conduct media training for Board members so they are comfortable and prepared for interviews.

Experts Directory. Develop an expert’s directory of Board members, alumni and staff available for interviews and speaking opportunities, as appropriate.

Op-eds. Write and pitch op-eds to various newspapers, magazines, and online sites on NAEP-related topics and Board endeavors.

Develop Contacts. Cultivate media contacts and resources by regularly keeping in touch, seizing opportunities to send occasional emails and making phone calls.

Advance Outreach. Conduct media pre-calls to create initial effective media placements on Board releases, events, and ongoing work.

Interactive Website. Create dynamic online press kits and updating the “what’s new” section with press releases and video releases to entice more media interest.

Multiple Platforms. Reporters for mainstream media now routinely produce web stories, videos, audio Q and As, and blog entries for each assignment. Outreach efforts should acknowledge these areas and tailor story ideas to a number of formats, helping reporters repurpose the material for different platforms.

Story Bank. The Board should create a bank of broader story ideas that came out of release events, reports, and publications, and pitch those to journalists.

Database Expansion. Expand media lists to include influential bloggers, online journalists, and others outside of traditional mainstream media.

IV. Social Media Strategy

The Governing Board can engage in social media effectively while honoring its mission and maintaining its position of independence. Tactics include the following.

Create Facebook and Twitter Accounts. The Board should develop profile pages for Facebook and Twitter to allow it to quickly and easily communicate with others using a variety of social media tools, including blogs, videos, images, tags, lists of friends, forums, and messaging. Alerts and postings on Board happenings and resources – events, data, background variables, etc. – can easily be disseminated and daily or weekly account updates keep the Board in the spotlight between releases. Also, Board members and staff with Facebook and Twitter accounts already can help promote Board activities.

Blogs by Board Members. Board members can rotate in writing a blog for www.nagb.org, with postings prompted by test score trends, framework issues, news topics, and the like. Board members can share insights, pose questions, and provoke
thoughtful discussion without overstepping their bounds. Ideally, the content would then be picked up by other bloggers who will send it to others, generating a viral effect.

- **Disseminate E-mail Newsletters.** The Board can develop a robust newsletter that includes content of interest to various audience groups, including teachers, associations, alumni, parents, and students who may not be aware of the Board and NAEP. It will help to forge connections and a sense of community among these audiences.

V. **WEBSITE STRATEGY**

To position the Governing Board as a leading voice and authority on the complex issues of academic assessment and advancing educational innovation and excellence, its website should be positioned to play a more prominent role in achieving its objectives. This requires a redesign that supports and promotes the various communications channels and content of the entire communications plan outlined above, including:

- **Website Design.** The overall design should support the key content areas the website is targeting and be organized for easy navigation by subject or audience.

- **Search Engine Optimization (SEO).** Reingold, the Board’s communications contractor, will work with the Board and its web contractor, Quotient, to ensure the website receives full credit from search engines for content as it is published. This will involve ensuring design, word usage, tags, and the like will be positioned to help www.nagb.org show up on searches, so that people looking up phrases like “national assessment” and “high school achievement” would find us as well.

- **Keyword Research.** This process will help the Board identify high-traffic subject areas and the associated keywords or search terms most frequently used to research them. It will help shape the organization and development of content in the “language” of the Board’s target audiences, using keywords and phrases they use when navigating search engines to find information and relevant content. Because nearly 90 percent of all clicks from search engine results pages originate on the first results page, it is critical to understand which words and phrases the Board can realistically compete for to achieve a first-page position and then ensure those keywords and phrases appear in the target page’s URL, tile, meta description, image alt text, video narration, and/or body text.

- **Content Development.** Once the above preliminary work is done, the site’s content that is interesting and relevant to the Board’s target audiences must be continuously developed, integrating your targeted keywords, posted in the appropriate areas of the site, and refreshed regularly.

- **Link-Building & Outreach.** The Board should develop an effective link-building campaign that includes initial research to identify a broad list of other relevant and authoritative websites, blogs, forums and other outlets based upon the www.nagb.org content and keyword strategy and approved by Board members and staff. Reingold can then approach these sites with requests that should identify a specific page on their
website and connect that content/topic back to a specific page on the www.nagb.org website with complimentary content, information or resources. The strategy would increase Board exposure and improve SEO efforts.

VI. NAEP COMMUNICATIONS ALIGNMENT

In the campaign’s first six months, Reingold will help the Governing Board work with NCES and other internal stakeholders to develop the foundation for expanded outreach. This foundation will focus on specific tasks under the strategies for stakeholder outreach and partnerships, traditional media, social media, Report Card releases, and the website.

Overarching Tasks

- **Review Governing Board branding.** Reingold will help the Governing Board and NCES to review the NAEP brand platform, determining how well its messaging and graphic elements distinguish and elevate NAEP and communicate the roles of the Governing Board and NCES.

- **Establish working group with NCES.** The Governing Board will create a NAEP working group with NCES to examine the activities and outreach undertaken by each group to determine if optimization is possible through greater coordination and collaboration. The group also can review the effectiveness of all NAEP materials and the Report Card release process, provide feedback, and recommend improvements.

- **Collaborate.** The Board staff and Reingold will define release plan roles, discuss deadlines, and streamline approval processes for release materials with NCES in a timeframe that enables optimal messaging, materials and content development.

- **Synergy.** The Board and NCES will work to align outreach strategies in communications and the website. For example, if NCES and its contractors sponsor a NAEP booth at a convention, then the Board can look into offering a member or staffer to give a presentation related to NAEP. Also, the Board and NCES can link to each other’s sites more regularly on NAEP-related items so that each group is contributing to increased exposure for the other.
Private School Participation Rates

and Plans for 2012 and 2013

Private school participation rates in the 2011 reading and mathematics assessments were sufficiently high to allow reporting of private schools overall, Catholic schools, Conservative Christian schools, and Lutheran schools. As with previous assessments, participation of other private schools, such as independent schools, did not meet standards for reporting as a separate group. However, students from those schools that did participate are included in the overall private school results as well as national results for all students.

In 2012, NAEP is conducting two assessments. These are the long-term trend assessments of students aged 9, 13, and 17; and an assessment of economics at grade 12. Both are national-level assessments, and private schools will be part of the samples. There will not be oversampling of private schools for these assessments that would allow separate reporting of Lutheran and Conservative Christian schools. Results for private school students overall, and Catholic school students, will be reported if response rates are sufficient.

In 2013, mathematics and reading will be assessed at grades 4, 8, and 12. Private schools will be part of the national samples at all grades. As with the 2012 assessments, oversampling is not planned for Lutheran and Conservative Christian schools in these assessments.
Private school students score higher than those in public schools

In 2011, the average mathematics score for fourth-graders attending public schools was 7 points lower than the overall score for students attending private schools, and 5 points lower than for students attending Catholic schools specifically (figure 10). There may be many reasons why private school students perform differently, on average, from public school students. Differences in demographic composition, availability of resources, admissions policies, parental involvement, and other factors not measured in NAEP may influence student achievement scores.

The average score for public school students was 1 point higher in 2011 than in 2009, while there was no significant change in the score for private school students overall or for Catholic school students over the same period. Scores for all three groups were higher in 2011 than in 1990; however, the 7-point score gap between private and public school students in 2011 was not significantly different from the gap in 1990.

Figure 10. Trend in fourth-grade NAEP mathematics average scores, by type of school

<table>
<thead>
<tr>
<th>Year</th>
<th>Public</th>
<th>Private</th>
<th>Catholic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1991</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1992</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1993</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1994</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1995</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1996</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1997</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1998</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>1999</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2000</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2001</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2002</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2003</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2004</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2005</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2006</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2007</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2008</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2009</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2010</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>2011</td>
<td>192</td>
<td>89</td>
<td>7</td>
</tr>
</tbody>
</table>


92% of fourth-graders attended public schools in 2011, and 8% attended private schools, including 4% in Catholic schools (table 4). In comparison to 1990, the percentage of students attending public schools in 2011 was larger, and the percentage attending private schools was smaller.

Table 4. Percentage distribution of students assessed in fourth-grade NAEP mathematics, by type of school: Various years, 1990-2011

<table>
<thead>
<tr>
<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>Public</td>
<td>89*</td>
<td>88*</td>
<td>89*</td>
<td>90*</td>
<td>90*</td>
<td>90*</td>
<td>90*</td>
<td>90*</td>
<td>90*</td>
</tr>
<tr>
<td>Private</td>
<td>11*</td>
<td>12*</td>
<td>11*</td>
<td>10*</td>
<td>10*</td>
<td>10*</td>
<td>9*</td>
<td>9*</td>
<td>9*</td>
</tr>
<tr>
<td>Catholic</td>
<td>7*</td>
<td>8*</td>
<td>8*</td>
<td>5*</td>
<td>5*</td>
<td>5*</td>
<td>4*</td>
<td>4*</td>
<td>4*</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.
1 Accommodations not permitted.
NOTE: Private schools include Catholic, other religious, and nonsectarian private schools. Detail may not sum to totals because of rounding.
No significant change in score gap between public and private school students

In 2011, the average reading score for eighth-graders attending public schools was 19 points lower than the overall score for students attending private schools, and 20 points lower than for students attending Catholic schools specifically (figure 26). The score gap between private and public school students in 2011 was not significantly different from the gap in either 2009 or 1992.

The average score for public school students was 1 point higher in 2011 than in 2009 and 6 points higher than in 1992, while there was no significant change in the score for private school students overall in comparison to either previous assessment year. The average score for Catholic school students did not change significantly from 2009 to 2011, but was 8 points higher in 2011 than in 1992.

3 The score-point difference is based on the difference between the unrounded scores as opposed to the rounded scores shown in the figure.

Figure 26. Trend in eighth-grade NAEP reading average scores, by type of school

Ninety-one percent of eighth-graders attended public schools in 2011, and 9 percent attended private schools, including 4 percent in Catholic schools (table 11). In comparison to 1992, the percentage of students attending public schools in 2011 was larger, and the percentages attending private schools and Catholic schools were smaller.

Table 11. Percentage distribution of students assessed in eighth-grade NAEP reading, by type of school: Various years, 1992-2011

<table>
<thead>
<tr>
<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>Public</td>
<td>89*</td>
<td>89*</td>
<td>89</td>
<td>91</td>
<td>91*</td>
<td>91</td>
<td>91</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>Private</td>
<td>11*</td>
<td>11*</td>
<td>11</td>
<td>9</td>
<td>9*</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Catholic</td>
<td>6*</td>
<td>7*</td>
<td>7</td>
<td>5*</td>
<td>5*</td>
<td>5*</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from 2011.
1 Accommodations not permitted.
NOTE: Private schools include Catholic, other religious, and nonsectarian private schools. Detail may not sum to totals because of rounding.

NAEP assesses representative samples of private school students at grades 4, 8, and 12 (or at ages 9, 13, and 17 for the long-term trend assessment) with each assessment. Private school results are reported whenever at least 70% of the schools in the sample participate. The symbol ‡ shows up in the tables or charts whenever the data could not be reported.

For a quick but detailed look at results for private and public schools, use the Private School Quick Data tool below. Select the subject and grade, then select whether you want to see the data broken out by two categories (public and private schools) or by five categories (including Catholic, other private schools, and schools of the Bureau of Indian Education and the Department of Defense). Then press "Show table." This will take you to a results table in the NAEP Data Explorer (NDE), in a new tab or window.

Once you see the results in the NDE, you can explore them further by customizing your tables or viewing them in a variety of chart formats, such as these:

To learn more about how to use the NDE, watch a short video or use the Quick Reference Guide (595K PDF). Learn about additional NDE features from the tutorial or access Help from every page of the tool.

The links on the Private and Other Nonpublic Schools page show you the results from the Nationsreportcard.gov website, where you will find detailed information about the report for that assessment.

Last updated 04 February 2011 (NB)
Executive Summary

This report is the first to focus on private school students’ performance on NAEP assessments. It provides results in reading, mathematics, science, and writing in 2000, 2002, 2003, and 2005. Specifically, it focuses on the three private school types that combined enroll the greatest proportion of private school students (Catholic, Lutheran, and Conservative Christian) as well as private schools overall. It also compares the performance of students in these schools to that of public school students to provide additional perspective.

Comparing student performance among the three types of private schools highlights several differences at grades 4 and 8 and a few at grade 12. Among the three types of private schools, few significant differences in performance were found at grade 12. The exceptions were that in 2000, the average score in science for grade 12 students in Catholic schools was 6 points higher than for students in Lutheran schools, and in the 2000 mathematics assessment, a higher percentage of twelfth-graders in Catholic schools performed at or above Proficient than twelfth-graders in Conservative Christian schools. Where differences existed at grades 4 and 8, students in Lutheran schools generally outperformed those in Conservative Christian schools. In some grade/subject combinations, Lutheran school students outperformed Catholic school students, and Catholic school students outperformed Conservative Christian school students.

Students at grades 4, 8, and 12 in all categories of private schools had higher average scores in reading, mathematics, science, and writing than their counterparts in public schools. In addition, higher percentages of students in private schools performed at or above Proficient compared to those in public schools.

Average scores in mathematics at grades 4 and 8 increased between 2000 and 2003 for both public and private schools overall. Students in Catholic schools also had higher average mathematics scores in 2003 than in 2000 in both grades.

The three types of private schools have few differences in their student demographics, except that Catholic schools generally enroll a greater proportion of Hispanic students than Lutheran schools. In general, private schools enroll a higher proportion of White students than public schools, while public schools have a higher proportion of Black and Hispanic students. Private schools also enroll a smaller proportion of students with disabilities, English language learners, and students eligible for free or reduced-price school lunch.

Black and Hispanic fourth-graders in all private schools combined had higher average mathematics scores in 2003 than in 2000. However, no significant differences in scores were found across the same time period for Black and Hispanic private school students in grade 4 reading or grade 8 mathematics.

A word of caution is needed: The data in this report provide a summary of the performance of students in public and private schools. The number of assessed students in some types of private schools is small, so it is not always feasible to make statistically meaningful comparisons between the performance of public school students and students in particular types of private schools. Factors not reported here, such as admission policies and parental involvement, can also influence student achievement.
Introduction

The National Assessment of Educational Progress (NAEP) measures the knowledge of students in grades 4, 8, and 12 in a variety of subject areas. About 10 percent of the entire U.S. school population, almost 5.3 million students, attended private schools during the 2001–02 school year (Broughman and Pugh 2004). This report provides NAEP results for private schools in 2000, 2002, 2003, and some in 2005. The subjects addressed in this report are reading and writing at grades 4 and 8, and mathematics and science at grades 4, 8, and 12.

What Information Is Available From NAEP About Private Schools?

In the 1970s and 1980s, the NAEP long-term trend assessment reported data separately for public school students and for all private school students, without distinguishing among types of private schools. Beginning in 1990, main NAEP also has reported performance data separately for private school students in Catholic and in non-Catholic schools. NAEP further increased the number of reporting categories for private schools for the 2000 assessments to include Catholic, Lutheran, and Conservative Christian schools. Combined, these schools enroll the majority of private school students, and they participated in NAEP at most grades in 2000, 2002, and 2003. This is the first NAEP report to compare the performance of students in these three types of private schools.

In this report, results of the 2000, 2002, and 2003 NAEP assessments for Catholic, Lutheran, and Conservative Christian schools are compared with each other. Although all categories of private schools are included in the total for “overall” private, only these three categories are shown separately. An update on 2005 results for Catholic and Lutheran schools is also provided.

Comparisons to public schools are also provided as points of reference. Results are given for students overall, as well as for student groups defined by race/ethnicity and—at grades 8 and 12 only—by the highest level of education reached by the students’ parents. Comparisons over time for the 2000–2003 mathematics and reading assessments are included whenever possible. Changes in the gaps between private and public school students’ performance are also discussed. The comparisons discussed in this report between students in the different types of private schools, and between students in private schools and public schools, are statistically significant unless otherwise stated.

What Findings Are Discussed in This Report?

For the past 30 years, NAEP has reported that students in private schools outperform students in public schools. This report confirms that point, but also looks more closely at NAEP results for three types of private schools: Catholic, Lutheran, and Conservative Christian. Combined, these schools enroll the majority of private school students, and they participated in NAEP at most grades in 2000, 2002, and 2003. This is the first NAEP report to compare the performance of students in these three types of private schools.

In this report, results of the 2000, 2002, and 2003 NAEP assessments for Catholic, Lutheran, and Conservative Christian schools are compared with each other. Although all categories of private schools are included in the total for “overall” private, only these three categories are shown separately. An update on 2005 results for Catholic and Lutheran schools is also provided.

Table 1. Grades assessed with reportable data, by subject and type of private school: Various years, 2000–2005

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Enrollment in private schools: Fall 2001</th>
<th>Reading</th>
<th>Mathematics</th>
<th>Science</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lutheran</td>
<td>219,397 (0.4%)</td>
<td>‡</td>
<td>4/8/‡</td>
<td>4/8</td>
<td>4/8/‡</td>
</tr>
<tr>
<td>Conservative Christian</td>
<td>823,469 (1.6%)</td>
<td>‡</td>
<td>4/8/‡</td>
<td>4/8</td>
<td>4/8/‡</td>
</tr>
<tr>
<td>Other Religious</td>
<td>882,009 (1.7%)</td>
<td>‡</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Nonsectarian</td>
<td>901,114 (1.7%)</td>
<td>‡</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Other Private</td>
<td>– (–)</td>
<td>–</td>
<td>‡/‡</td>
<td>‡/‡</td>
<td>‡/‡</td>
</tr>
</tbody>
</table>

– Not available because data were not collected.
‡ Reporting standards not met. Data are not reported because participation rates failed to meet minimum NCES standards for reporting.

How Are Results Reported?

Results are reported in two ways: as average scale scores and as percentages of students attaining NAEP achievement levels. Average scale scores in NAEP measure what students know and can do, and are reported on 0–500 scales in mathematics and reading, with all three grades on the same scale; science and writing are reported on 0–300 scales with each of the three grades on a separate scale.

Three achievement levels—Basic, Proficient, and Advanced—have been developed by the National Assessment Governing Board (NAGB) to provide a context for interpreting student performance on NAEP assessments. These achievement levels state what students should know and be able to do in each subject area and at each grade assessed. Further information on achievement levels and sample questions associated with these achievement levels can be found in previous NAEP reports (see, for example, Braswell et al. 2005; Donahue, Daane, and Jin 2005) or online at http://nces.ed.gov/nationsreportcard/itemmaps/ or http://www.nagb.org/pubs/pubs.html.

As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that achievement levels are to be used on a trial basis and should be interpreted and used with caution. However, NCES and NAGB have affirmed that these performance standards are useful for understanding trends in achievement. NAEP achievement levels have been widely used by national and state officials for over a decade.

Cautions in Interpretation

It is important to note that a relationship between a variable and measures of educational achievement, like the ones presented in this report, does not imply that a difference in the variable causes differences in educational achievement. Higher performance scores in private schools do not imply that the private schools are better than public schools, as they often serve different populations of students. In addition, the results are cross-sectional, rather than longitudinal, so they only provide a snapshot for any given point in time. Comparing students of a particular demographic group may provide more information; however, only one characteristic is compared at a time. For example, this report compares the performance of Black students in different types of schools, but it does not compare the performance of Black students who are eligible for free school lunch across school types, because of limitations of the sample. There are many reasons why the performance of one group of students differs from another, including factors that are not measured in NAEP.

Some key results are presented in the body of the report. Additional data for reading and mathematics are found in the Technical and Data Appendix. As indicated in the appendix tables, some of the data presented in the appendix should be interpreted with caution due to the uncertainty of the variability of the estimates. Also, estimates based on smaller student groups are likely to have relatively large standard errors. These large standard errors mean that some differences that seem large may not be statistically significant. Because private school results are based on smaller samples, they are less likely to show significant differences than the results from public schools. Standard errors, as well as additional data on science and writing, can be found using the NAEP data tool at http://nces.ed.gov/nationsreportcard/nde/. Further explanation is provided in the Technical and Data Appendix.

Private School Types

- **Catholic** schools included parochial, diocesan, and private order schools.
- **Lutheran** schools included all those that indicated an affiliation with any branch of the Lutheran Church.
- **Conservative Christian** schools included all those that indicated membership in Accelerated Christian Education, American Association of Christian Schools, Association of Christian Schools International, Association of Christian Teachers and Schools, or the Oral Roberts University Educational Fellowship.
- **Other Religious** included all other schools that indicated an affiliation with any other religious organization. This category was tabulated separately only in 2000.
- **Nonsectarian** schools included all private schools without an affiliation to any religious organization or institution. This category was tabulated separately only in 2000.
- **Other Private** schools included the combined data for “Nonsectarian” and “Other Religious” schools, when data in those categories were too few to report separately. This category was created in 2002.

What Are NAEP Achievement Levels?

Achievement levels are performance standards set by the National Assessment Governing Board (NAGB) to help interpret student performance on NAEP. The three NAEP achievement levels, from lowest to highest, are

**Basic**—denotes partial mastery of the knowledge and skills that are fundamental for proficient work at each grade.

**Proficient**—represents solid academic performance. Students reaching this level have demonstrated competency over challenging subject matter.

**Advanced**—signifies superior performance.

Detailed descriptions of the NAEP achievement levels for each subject can be found on the NAGB website (http://www.nagb.org/pubs/pubs.html).
Characteristics of Students in Private Schools

Comparison of the characteristics of students in different types of private schools can indicate the extent to which they are serving different student populations. Characteristics of public school students are also shown for comparison. The student groups highlighted here include those defined by students’ race/ethnicity, by the level of their parents’ education, by their eligibility to receive free or reduced-price school lunch, and by whether they have been identified as having a disability or as English language learners. The figures on these pages display the data from the 2003 grade 8 reading assessment, and are representative of findings across the subjects and grades. For student demographics in other grades and subjects, see the data appendix and the NAEP data tool, http://nces.ed.gov/nationsreportcard/ndel/.

**Race/Ethnicity**

At all grades and in all subject assessments included in this report, private schools overall had a higher percentage of White students than public schools, and public schools had higher percentages of Black and Hispanic students than private schools. Within the different types of private schools, Lutheran schools generally enrolled a higher percentage of White students than Catholic schools, and Catholic schools enrolled a higher percentage of Hispanic students than Lutheran schools. Figure 1 shows the racial/ethnic distributions for the 2003 grade 8 reading assessment. The racial/ethnic categories shown—White, Black, Hispanic, and Asian/Pacific Islander—are mutually exclusive. Pacific Islander includes Native Hawaiian, and Hispanic includes Latino. Race categories exclude Hispanic origin unless specified. Data for American Indian/Alaska Native students are included in the total but are not broken out separately due to small sample sizes.

**Parents’ Highest Level of Education**

When eighth-grade students were asked to report their parents’ highest level of education, a greater percentage of students in private schools compared to public schools reported that at least one parent had graduated from college. No statistically significant differences in the highest level of education reported for at least one parent were found for any school category among the three types of private schools. Figure 2 provides data at grade 8 from the 2003 reading assessment.

---

**Figure 1.** Percentage distribution of students who participated in reading assessment, by race/ethnicity and type of school, grade 8: 2003

**Figure 2.** Percentage distribution of students who participated in reading assessment, by student-reported parents’ highest level of education and type of school, grade 8: 2003

---

*a* Significantly different from Catholic schools.

*b* Significantly different from Lutheran schools.

*c* Significantly different from public schools.

**NOTE:** At each grade, approximately 1 percent of public school students were classified as American Indian/Alaska Native, while the proportion of students of the same race/ethnicity in private schools rounds to zero. Results are not shown for students whose race/ethnicity was “other.” Data for Other Private schools are included in the overall Private data but not reported separately.

Free and Reduced-Price School Lunch

A student’s eligibility for free or reduced-price school lunch, which depends on family income, is often used as a proxy for a measure of socioeconomic status. In the 2003 reading assessment, 6 percent of students in all private schools combined were reported as eligible for free or reduced-price lunch at grade 4, and 11 percent at grade 8. Approximately 9 percent of fourth-graders in both Catholic and Lutheran schools were eligible. In contrast, public schools reported that 44 percent of their fourth-graders were eligible for free or reduced-price lunch.

However, these data should be interpreted with caution because of the high percentages of students in private schools for whom information was not available. Information was not available for about half of the fourth-grade students in Catholic and Lutheran schools. As seen in figure 3, the percentages of private school students for whom lunch-eligibility information was not available are also high for grade 8. In contrast, information was not available for only 6 percent of the eighth-grade students in public schools. As a result, this report does not present performance data by groups defined by eligibility for free and reduced-price school lunch.

Figure 3. Percentage distribution of students in reading, by students’ eligibility for free/reduced-price school lunch and type of school, grade 8: 2003

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Private</td>
<td>11(^a)</td>
</tr>
<tr>
<td></td>
<td>29(^d)</td>
</tr>
<tr>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Catholic</td>
<td>13(^c,d)</td>
</tr>
<tr>
<td></td>
<td>39(^c,d)</td>
</tr>
<tr>
<td></td>
<td>48(^c,d)</td>
</tr>
<tr>
<td>Lutheran</td>
<td>10(^c,d)</td>
</tr>
<tr>
<td></td>
<td>39(^c,d)</td>
</tr>
<tr>
<td></td>
<td>50(^c,d)</td>
</tr>
<tr>
<td>Conservative</td>
<td>12(^h,d)</td>
</tr>
<tr>
<td></td>
<td>87(^a,h,d)</td>
</tr>
</tbody>
</table>

- \(^a\) Significantly different from Catholic schools.
- \(^b\) Significantly different from Lutheran schools.
- \(^c\) Significantly different from Conservative Christian schools.
- \(^d\) Significantly different from public schools.

\(^1\) For Conservative Christian schools, the percentage of students who were eligible for free/reduced-price lunch rounds to zero. This percentage is significantly different from the percentages in Catholic, Lutheran, and public schools.

NOTE: Detail may not sum to totals because of rounding. Data for Other Private schools are included in the overall Private data but not reported separately.

Students With Disabilities (SD) and English Language Learners (ELL)

Two other demographic variables that were analyzed for this report are the percentage of students with disabilities and the percentage of students identified as English language learners in the different types of schools. Figure 4 shows these percentages for eighth-graders participating in the reading assessment in 2003. Overall, less than 3 percent of grade 8 students enrolled in any private school were identified as SD or ELL. No significant differences in this percentage were found among the three private school types discussed here. Public schools enroll a larger percentage of students who are SD, ELL, or both—15 percent.

Figure 4. Percentage distribution of students in reading, by students with disabilities and English language learners, and by type of school, grade 8: 2003

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Grade 8 SD(^1)</th>
<th>Grade 8 ELL(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Private</td>
<td>3(^d)</td>
<td>100(^d)</td>
</tr>
<tr>
<td>Catholic</td>
<td>2(^d)</td>
<td>100(^d)</td>
</tr>
<tr>
<td>Lutheran</td>
<td>2(^d)</td>
<td>100(^d)</td>
</tr>
<tr>
<td>Conservative Christian</td>
<td>3(^d)</td>
<td>1(^d)</td>
</tr>
</tbody>
</table>

\(^1\) Significantly different from public schools.
\(^2\) English language learners.
\(^3\) The percentages of students who are English language learners round to zero in private schools overall and in Catholic and Lutheran schools. For private schools overall and for Catholic schools, these percentages are significantly different from those for public schools.

NOTE: Detail may not sum to totals because of rounding. Data for Other Private schools are included in the overall Private data but not reported separately.

Few differences in performance were found among the three types of private schools. Where differences were found, students in Lutheran schools generally outperformed their counterparts in Conservative Christian schools. With some exceptions, no significant differences were found between the performance of students in Lutheran schools and Catholic schools. Students in Catholic schools outperformed students in Conservative Christian schools in three of the grade/subject combinations.

For each assessment discussed in this report at each grade, students as a whole in every category of private schools had a higher average score than their counterparts in public schools. In addition, higher percentages of students in most categories of private schools performed at or above Proficient and Basic compared to the percentages of public school students.

**Reading Performance**

As shown in figure 5, among the three types of private schools, there were no significant differences in reading performance at grade 4 in 2003. The average scale scores in 2003 for students at grade 8 in both Catholic and Lutheran schools were higher than the average score for those in Conservative Christian schools.

Students in all types of private schools at both grades 4 and 8 had higher average reading scores in 2003 than students at the same grade in public schools. Also, higher percentages of students at grades 4 and 8 performed at or above Proficient and at or above Basic in reading in all private schools combined, as well as in all the subcategories of private schools, than in public schools. No significant differences in these percentages were found among the three types of private schools at either grade.

### Figure 5.

Average scale scores and achievement-level results in reading, by type of school, grades 4 and 8: 2003

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>216</td>
<td>261</td>
</tr>
<tr>
<td>Private</td>
<td>235$d$</td>
<td>282$d$</td>
</tr>
<tr>
<td>Catholic</td>
<td>235$d$</td>
<td>281$t,d$</td>
</tr>
<tr>
<td>Lutheran</td>
<td>232$d$</td>
<td>281$t,d$</td>
</tr>
<tr>
<td>% at or above Basic</td>
<td>62</td>
<td>72</td>
</tr>
<tr>
<td>% at or above Proficient</td>
<td>30</td>
<td>53$e$</td>
</tr>
</tbody>
</table>

*a* Significantly different from Catholic schools.

*b* Significantly different from Lutheran schools.

*c* Significantly different from Conservative Christian schools.

*d* Significantly different from public schools.

**NOTE:** The NAEP reading scale ranges from 0 to 500. Data for Other Private schools are included in the overall Private data but not reported separately.

**SOURCE:** U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.
Mathematics Performance
Figure 6 shows that at grade 4 there were no significant differences in mathematics performance between students in Catholic and Lutheran schools in 2003. At grade 8, students in Lutheran schools scored higher on average than students in Catholic and Conservative Christian schools in 2003. A greater percentage of eighth-graders in Lutheran schools also performed at or above Proficient in 2003 than their counterparts in Catholic and Conservative Christian schools. There were no significant differences in average scores among the three types of private schools at grade 12 in 2000.

In 2003, students at grades 4 and 8 in private schools overall, and in all types of private schools with reportable data, had higher average scores in mathematics than students in public schools. Similar results were seen for grade 12 in 2000 (the most recent mathematics assessment for that grade). Also, higher percentages of students at all three grades in private schools overall, and in Catholic and Lutheran schools, performed at or above Proficient and at or above Basic, compared to their counterparts in public schools.

Science Performance
Figure 7 shows results from the most recent science assessment in 2000. Fourth-graders in Lutheran schools scored 8 points higher, on average, than their counterparts in Conservative Christian schools. No significant differences were seen among eighth-grade students in the three private school types. The average score of twelfth-graders in Catholic schools was 6 points higher than the average score of twelfth-graders in Lutheran schools. Also, a greater percentage of Catholic school students at grade 12 than their peers in Lutheran schools performed at or above Basic. These twelfth-grade science results are the only overall results in which students in Catholic schools outperformed their Lutheran school counterparts.

In 2000, students at grades 4, 8, and 12 in all categories of private schools with reportable data had higher average scores in science than students in public schools. Higher percentages of students at grades 4 and 8 in all private school categories performed at or above Proficient in comparison to their counterparts in public schools. At grade 12, higher percentages of students in private schools overall and in Catholic schools than students in public schools performed at or above Proficient.

Figure 6. Average scale scores and achievement-level results in mathematics, by type of school, grades 4, 8, and 12: 2000 and 2003

<table>
<thead>
<tr>
<th>Grade 4 2003</th>
<th>Grade 8 2003</th>
<th>Grade 12 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale score</td>
<td>Scale score</td>
<td>Scale score</td>
</tr>
<tr>
<td>Public</td>
<td>Private</td>
<td>Catholic</td>
</tr>
<tr>
<td>234</td>
<td>244</td>
<td>245</td>
</tr>
<tr>
<td>% at or above Basic</td>
<td>% at or above Proficient</td>
<td>% at or above Basic</td>
</tr>
<tr>
<td>76</td>
<td>31</td>
<td>67</td>
</tr>
<tr>
<td>88</td>
<td>44</td>
<td>82</td>
</tr>
<tr>
<td>88</td>
<td>43</td>
<td>81</td>
</tr>
<tr>
<td>90</td>
<td>48</td>
<td>78</td>
</tr>
</tbody>
</table>

Type of school: Public, Private, Catholic, Lutheran, Conservative Christian.

a Significantly different from Catholic schools.
b Significantly different from Lutheran schools.
c Significantly different from Conservative Christian schools.
d Significantly different from public schools.

NOTE: The NAEP mathematics scale ranges from 0 to 500. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.

Writing Performance

Figure 8 shows results from the most recent writing assessment in 2002. Average scores for fourth-graders in Catholic and Lutheran schools were higher than the average score for fourth-graders in Conservative Christian schools. The percentages of students at grade 4 who scored at or above Proficient were higher in Catholic and Lutheran schools than in Conservative Christian schools. There were no significant differences in performance among the three private school types at grade 8.

In 2002, students at grades 4 and 8 in private schools overall, and in Catholic and Lutheran schools, had higher average scores in writing than students in the same grades in public schools. Students at grade 4 in Conservative Christian schools scored higher on average than their counterparts in public schools. Higher percentages of fourth- and eighth-grade students in private schools overall, in Catholic, and in Lutheran schools performed at or above Proficient in writing in comparison to their counterparts in public schools.

Figure 7. Average scale scores and achievement-level results in science, by type of school, grades 4, 8, and 12: 2000

Figure 8. Average scale scores and achievement-level results in writing, by type of school, grades 4 and 8: 2002
Changes in Scores for Public and Private Schools

Changes in performance between 2000 and 2003 can be examined for reading at grade 4 and mathematics at grades 4 and 8. Studying changes over time will help to determine which types of schools are making strides in improving student achievement as well as provide an indicator of changing demographics. Changes in overall average scale scores are shown in this section. Additional results, including changes in achievement-level performance and changes for particular student groups, can be found in the data appendix.

Grade 4 Reading Changes

As shown in figure 9, the apparent increase between 2000 and 2003 in the average reading score for private school students was not statistically significant. Catholic schools were the only private school type with reliable data in both 2000 and 2003. The apparent increase of 6 points between those two years was also not statistically significant. Public school students showed an improvement of 5 points in the average reading score during the same time. There was no statistically significant change in the reading score gap between public and private school fourth-graders from 2000, when it was 20 points, to 2003, when it was 18 points.

Grade 4 Mathematics Changes

At grade 4, improvements in average mathematics scores between 2000 and 2003 were seen in almost all types of schools, including private overall, Catholic, Lutheran, and public schools (see figure 10). Average scores in Catholic and Lutheran schools increased by 7 and 4 points, respectively, while average scores for all private schools combined increased 6 points. Average scores in public schools increased by 10 points in this same time period. The public-private score gap decreased from 14 points in 2000 to 10 points in 2003.

Grade 8 Mathematics Changes

As shown in figure 11, average mathematics scores increased between 2000 and 2003 for eighth-grade students in all private schools combined, in Catholic schools, and in public schools. The average score for Catholic schools increased by 6 points between 2000 and 2003. The apparent increase during that same time in average score for Lutheran schools was not statistically significant. The difference between average mathematics scores of eighth-grade students in public and private schools did not change significantly between 2000, when it was 15 points, and 2003, when it was 16 points.

2005 Private School Results

The 2005 results for private school students overall are not presented because the participation rates for this group were too low to produce valid and reliable estimates. Results are, however, available for students who attended two types of private schools: Catholic and Lutheran. Only these two private school types had sufficient participation rates to produce valid and reliable results.

In reading, the average scores for Catholic and Lutheran students were not significantly different in 2005 than in any previous assessment year in both grades 4 and 8. Likewise, the average mathematics scores for both types of schools did not differ significantly from those in previous years in either grade. No differences between 2005 and previous assessment years were seen in the percentage of students performing at or above Proficient or at or above Basic at either grade 4 or 8 in either subject. No differences were seen in 2005 between the two types of private schools on any measure.

These data and other 2005 data are available through the NAEP data tool (http://nces.ed.gov/nationsreportcard/nde/.)

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Average scale score</th>
<th>Percentage of students</th>
<th>Average scale score</th>
<th>Percentage of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catholic</td>
<td>234</td>
<td>80</td>
<td>244</td>
<td>88</td>
</tr>
<tr>
<td>Lutheran</td>
<td>231</td>
<td>77</td>
<td>245</td>
<td>89</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catholic</td>
<td>280</td>
<td>90</td>
<td>290</td>
<td>81</td>
</tr>
<tr>
<td>Lutheran</td>
<td>280</td>
<td>89</td>
<td>293</td>
<td>84</td>
</tr>
</tbody>
</table>

Figure 9. Average reading scale scores, by type of school, grade 4: 2000 and 2003

<table>
<thead>
<tr>
<th>School Type</th>
<th>2000</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>211</td>
<td>216</td>
</tr>
<tr>
<td>Private</td>
<td>231</td>
<td>235</td>
</tr>
<tr>
<td>Catholic</td>
<td>229</td>
<td>235</td>
</tr>
</tbody>
</table>

* Significantly different from 2003.

NOTE: The NAEP reading scale ranges from 0 to 500. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.


Figure 10. Average mathematics scale scores, by type of school, grade 4: 2000 and 2003

<table>
<thead>
<tr>
<th>School Type</th>
<th>2000</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>224</td>
<td>234</td>
</tr>
<tr>
<td>Private</td>
<td>238</td>
<td>244</td>
</tr>
<tr>
<td>Catholic</td>
<td>237</td>
<td>244</td>
</tr>
<tr>
<td>Lutheran</td>
<td>241</td>
<td>245</td>
</tr>
</tbody>
</table>

* Significantly different from 2003.

NOTE: The NAEP mathematics scale ranges from 0 to 500. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.


Figure 11. Average mathematics scale scores, by type of school, grade 8: 2000 and 2003

<table>
<thead>
<tr>
<th>School Type</th>
<th>2000</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>272</td>
<td>276</td>
</tr>
<tr>
<td>Private</td>
<td>266</td>
<td>292</td>
</tr>
<tr>
<td>Catholic</td>
<td>284</td>
<td>289</td>
</tr>
<tr>
<td>Lutheran</td>
<td>292</td>
<td>296</td>
</tr>
</tbody>
</table>

* Significantly different from 2003.

NOTE: The NAEP mathematics scale ranges from 0 to 500. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.

NAEP gathers information related to academic achievement from school records and from questionnaires administered to the students it assesses. The groups highlighted in this report include those defined by students’ race/ethnicity and by level of parents’ education. Further results for these groups, and information on results by gender, can be found in the data appendix for reading and mathematics and in the data tool for writing and science (http://nces.ed.gov/nationsreportcard/nde/).

**Race/Ethnicity**

No significant differences were found in the performance of Black students or Hispanic students in the three types of private schools. In several instances, White students in Catholic and Lutheran schools had higher average scores than White students at the same grade in Conservative Christian schools. For instance, at grade 4, White students in Catholic schools had higher average scores on the 2002 writing assessment than White students in Conservative Christian schools. At grade 8, White students in Catholic schools scored higher, on average, on the 2003 reading assessment than White students in Conservative Christian schools. At grade 8, White students in Catholic schools scored higher, on average, on the 2003 reading assessment than White students in Conservative Christian schools. White eighth-grade students in Lutheran schools also had higher average mathematics scores in 2003 than their counterparts in Catholic schools.

In private schools overall, in every case but one, students from each racial/ethnic category had higher average scores than their counterparts in public schools in the most recent assessment year. The one exception was in grade 4 mathematics (shown in figure 12), where there was no statistically significant difference between the average scores of Asian/Pacific Islander students in public schools (246) and in private schools overall (249) in 2003.

White students and Hispanic students in Catholic schools scored higher on average than public school students of the same race/ethnicity in every subject and grade where the sample size was sufficient to produce a reliable estimate. Black students in Catholic schools had higher average scores than Black students in public schools in all subjects and grades, except in grade 4 mathematics and grade 4 writing, where the apparent differences were not statistically significant.

Again, as with the average scores, generally a greater percentage of students in private schools performed at or above Proficient than their counterparts in public schools, regardless of their racial/ethnic background. However, there were some exceptions. For example, there was no statistically significant difference between the percentages of Asian/Pacific Islander students performing at or above Proficient in mathematics and writing in private schools overall compared to public schools at grade 4.

**Parents’ Highest Level of Education**

Within each level of parents’ highest educational attainment, there were almost no differences in the performance of students among the three types of private schools. Figure 13 illustrates these findings with grade 8 data from the 2003 reading assessment. Additional data are provided in the data appendix. The exceptions were in grade 8 mathematics and grade 12 science. In 2003, the average mathematics score of eighth-graders in Lutheran schools who reported that at least one parent graduated from college was higher than the average scores of their counterparts in Catholic and Conservative Christian schools. In 2000, the average science score of twelfth-graders in Catholic schools who reported that one parent received some education after high school was higher than the average score of students in Lutheran schools reporting the same parental education level.

In almost all cases, the average scores of students in all types of private schools were higher than those of their counterparts in public schools for each reported level of their parents’ education. Exceptions to this pattern occurred in mathematics and science at grades 8 and 12, where there were no significant differences between the average scores of students in public and private schools for certain categories of parents’ education.

In both eighth and twelfth grades and in all subjects included in this report, a greater percentage of students reporting that at least one parent graduated from college than their counterparts in public schools performed at or above Proficient.
Figure 12.  Average scale scores and achievement-level results in mathematics, by race/ethnicity and type of school, grade 4: 2003

Interpret data with caution. The nature of the sample does not allow accurate determination of the variability of the statistic.

Significantly different from public schools.

Data for Hispanic and Asian/Pacific Islander students attending Lutheran schools are not shown because sample size is insufficient to permit a reliable estimate.

NOTE: The NAEP mathematics scale ranges from 0 to 500. Data for Other Private schools are included in the overall Private data but not reported separately.


Figure 13.  Average scale scores and achievement-level results in reading, by student-reported parents’ highest level of education and type of school, grade 8: 2003

Significantly different from public schools.

Data for Catholic, Lutheran, and Conservative Christian are not shown because sample sizes are insufficient to permit reliable estimates of these categories for students whose parents’ reported education level is less than high school.

NOTE: The NAEP reading scale ranges from 0 to 500. Data for Other Private schools are included in the overall Private data but not reported separately.

Changes in Average Scores for Racial/Ethnic Groups

Because of the strong national emphasis on reducing the achievement gaps between minority and nonminority students, it is informative to examine the changes in average scores by race/ethnicity (see tables 2 and 3). This section examines the White – Black and White – Hispanic score gaps in private schools. Because the sample sizes for Black and Hispanic students in private schools were small, resulting in large standard errors, tests of statistical significance may show no difference in cases where the differences appear large.

Grade 4 Mathematics
The average mathematics scores in private schools overall increased for White, Black, and Hispanic fourth-graders between 2000 and 2003. Average scores for White and Hispanic fourth-graders in Catholic schools also increased between 2000 and 2003. In Lutheran schools, the average score for White students increased by 6 points from 2000 to 2003. The apparent changes in the White – Black and White – Hispanic score gaps between 2000 and 2003 were not statistically significant in any of the types of private schools. In contrast, the score gaps between White and Black students and between White and Hispanic students in public schools decreased between 2000 and 2003.

Grade 8 Mathematics
Average scores for White eighth-graders increased between 2000 and 2003 in private schools overall and in Catholic schools. In public schools, average mathematics scores for White, Black, and Hispanic eighth-graders increased between 2000 and 2003, by 4, 8, and 6 points, respectively. The White – Black score gap decreased between 2000 and 2003 in public schools only. No statistically significant changes in the score gaps between White and Hispanic students were detected for any type of school.
### Table 2. Average scale scores and score gaps for White and Black students in reading and mathematics, by type of school, grades 4 and 8: 2000 and 2003

<table>
<thead>
<tr>
<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><strong>Grade 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>223</td>
<td>227</td>
<td>4</td>
<td>189</td>
<td>197</td>
<td>9</td>
<td>34</td>
<td>30</td>
<td>-4</td>
</tr>
<tr>
<td>Private</td>
<td>236</td>
<td>239</td>
<td>3</td>
<td>213</td>
<td>210</td>
<td>-3</td>
<td>24</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>Catholic</td>
<td>236</td>
<td>240</td>
<td>4</td>
<td>209</td>
<td>211</td>
<td>1</td>
<td>27</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>233</td>
<td>243</td>
<td>9</td>
<td>203</td>
<td>216</td>
<td>13</td>
<td>30</td>
<td>27</td>
<td>-4</td>
</tr>
<tr>
<td>Private</td>
<td>241</td>
<td>248</td>
<td>7</td>
<td>213</td>
<td>221</td>
<td>8</td>
<td>29</td>
<td>27</td>
<td>-2</td>
</tr>
<tr>
<td>Catholic</td>
<td>241</td>
<td>248</td>
<td>7</td>
<td>210</td>
<td>222</td>
<td>12</td>
<td>31</td>
<td>26</td>
<td>-5</td>
</tr>
<tr>
<td><strong>Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>283</td>
<td>287</td>
<td>4</td>
<td>243</td>
<td>252</td>
<td>8</td>
<td>40</td>
<td>35</td>
<td>-5</td>
</tr>
<tr>
<td>Private</td>
<td>291</td>
<td>297</td>
<td>6</td>
<td>258</td>
<td>260</td>
<td>2</td>
<td>33</td>
<td>37</td>
<td>4</td>
</tr>
<tr>
<td>Catholic</td>
<td>289</td>
<td>296</td>
<td>7</td>
<td>254</td>
<td>260</td>
<td>6</td>
<td>35</td>
<td>36</td>
<td>1</td>
</tr>
</tbody>
</table>

* Interpret data with caution. The nature of the sample does not allow accurate determination of the variability of the statistic.

**NOTE:** The NAEP reading and mathematics scales each range from 0 to 500. Score differences are calculated based on differences between unrounded average scale scores. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.


### Table 3. Average scale scores and score gaps for White and Hispanic students in reading and mathematics, by type of school, grades 4 and 8: 2000 and 2003

<table>
<thead>
<tr>
<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><strong>Grade 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>223</td>
<td>227</td>
<td>4</td>
<td>188</td>
<td>199</td>
<td>11</td>
<td>35</td>
<td>28</td>
<td>-7</td>
</tr>
<tr>
<td>Private</td>
<td>236</td>
<td>239</td>
<td>3</td>
<td>215</td>
<td>220</td>
<td>5</td>
<td>21</td>
<td>19</td>
<td>-2</td>
</tr>
<tr>
<td>Catholic</td>
<td>236</td>
<td>240</td>
<td>4</td>
<td>211</td>
<td>219</td>
<td>8</td>
<td>25</td>
<td>21</td>
<td>-4</td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>233</td>
<td>243</td>
<td>9</td>
<td>207</td>
<td>221</td>
<td>14</td>
<td>26</td>
<td>21</td>
<td>-5</td>
</tr>
<tr>
<td>Private</td>
<td>241</td>
<td>248</td>
<td>7</td>
<td>220</td>
<td>231</td>
<td>11</td>
<td>21</td>
<td>17</td>
<td>-4</td>
</tr>
<tr>
<td>Catholic</td>
<td>241</td>
<td>248</td>
<td>7</td>
<td>217</td>
<td>229</td>
<td>11</td>
<td>24</td>
<td>20</td>
<td>-4</td>
</tr>
<tr>
<td><strong>Grade 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>283</td>
<td>287</td>
<td>4</td>
<td>252</td>
<td>258</td>
<td>6</td>
<td>31</td>
<td>28</td>
<td>-3</td>
</tr>
<tr>
<td>Private</td>
<td>291</td>
<td>297</td>
<td>6</td>
<td>273</td>
<td>274</td>
<td>1</td>
<td>18</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>Catholic</td>
<td>289</td>
<td>296</td>
<td>7</td>
<td>271</td>
<td>272</td>
<td>1</td>
<td>18</td>
<td>24</td>
<td>6</td>
</tr>
</tbody>
</table>

* Interpret data with caution. The nature of the sample does not allow accurate determination of the variability of the statistic.

**NOTE:** The NAEP reading and mathematics scales each range from 0 to 500. Score differences are calculated based on differences between unrounded average scale scores. Data for Nonsectarian and Other Religious schools for 2000 and data for Other Private schools for 2003 are included in the overall Private data for those years but not reported separately.

The Nation’s Report Card: 2011 Mathematics and Reading, TUDA Grades 4 and 8
Release Event Debrief Report

Overview
The public release of *The Nation’s Report Card: 2011 Mathematics and Reading, TUDA Grades 4 and 8* took place on December 7, 2011 at 10 a.m. EST as a live and webcast event at City Springs Elementary/Middle School in Baltimore, MD. Governing Board member Tonya Miles served as the moderator and the release panel included Governing Board member Andrés Alonso, National Center for Education Statistics (NCES) Commissioner Jack Buckley, and Michael Casserly, executive director of the Council of the Great City Schools (CGCS). Reingold, in partnership with Ogilvy PR Worldwide, conducted all release activities in cooperation with the Governing Board staff, NCES staff, and contractors. For this release the participants included a total of 66 in-person release attendees (all internal staff and contractors were removed from the registrant list) and 252 webcast participants (all internal staff and contractors were removed from the registrant list).

Webcast and Live Event Attendees – 318 total (excluding internal staff and contractors)

Social Media Outreach Results
The release was promoted across social media sites by many stakeholder groups and public school districts. In addition, as a result of online channels, many other stakeholders posted or
reposted content promoting the release. The Governing Board’s increased engagement with Twitter users leading up to the release helped raise awareness and generate interest in the event.

1 **On December 7, there were over 550 mentions** of the release on social media; primarily on Twitter, blogs, and Facebook. Multimedia was also posted on YouTube and Flickr.
2 During the webinar release hour (10 am – 11 am), there were over 175 on-topic social media mentions.
3 There were over 100 visits to the event splash page from social media sites, mobile device applications, and other online sources, and the splash page was shared 75 times through social media.
4 Several reporters and media outlets promoted the release via social media, including: Huffington Post Education, Shanker Blog, Dorie Turner (Associated Press), and Erik Robelen and Catherine Gewertz (Education Week).
5 Several individuals and groups, including media, wrote replies, re-tweeted, and shared the Board’s content. Stakeholders that promoted the release included:
   - AFT Teach
   - Atlanta Public Schools
   - American Federation of Teachers
   - Baltimore City Public Schools
   - Baltimore County Chamber of Commerce
   - Baltimore Curriculum Project
   - Baltimore Education Coalition
   - Boston Public Schools
   - Council of the Great City Schools
   - Dallas Independent School District
   - Detroit Public Schools
   - Drexel Math Forum
   - Ed Trust
   - Ed Sector
   - Fritzwire
   - Grantmakers for Children
   - Meck Ed
   - National Association of State Boards of Education
   - New York Academy of Sciences
   - National Council of Teachers of Mathematics
   - PIE Network
   - School Choice Kentucky
   - Southern Education Desk
   - Supporting Public Schools of Choice (Baltimore)
   - U.S. Department of Education
   - Young Education Professionals

**Traditional Media Outreach Results**
Advisories on the release were sent to thousands of print, broadcast and online journalists nationally, and access to an embargoed media site was established 48 hours prior to the event.
The Board also convened an embargoed media call the day before the release to answer journalists’ questions, with participation from Board executive director Cornelia Orr, Commissioner Jack Buckley and CGCS executive director Michael Casserly.

1. The release received news coverage through **258** print, broadcast and online media stories with **117** local news stories in outlets targeting TUDA districts.
2. The combined coverage of this release has the potential to reach **222.6 million people**.
   - Independent coverage in **19** daily newspapers, including those in 11 TUDA districts reached a daily print circulation of **4.8 million readers**.
   - Online placements of **154** stories in major online newspapers and consumer news websites like the *Huffington Post*, *Education Week*, and *Slate* will reach approximately **160 million unique visitors per month**.
   - Broadcast media outlets published 54 news stories posted in the week following the release allowing coverage to reach an estimated total of **20.8 million unique visitors per month**. This figure includes coverage from TodayShow.com, MSNBC.com, CNN.com and CBS.com.
   - Twenty-nine well-reputed blogs, including the Philadelphia Public Schools Notebook, Gotham Schools, Joanne Jacobs, Washington Post and the Education Gadfly discussed the release. These blogs combine for more than **37 million unique visitors per month**.
3. The TUDA press release was picked up by **239** news sites that carry wire stories.
The Nation’s Report Card:
2011 Trial Urban District Assessment, Mathematics and Reading

Select Media Clips

Associated Press
Austin American-Statesman
Baltimore Sun
Christian Science Monitor
CNN
Education Week
Miami Herald
Michigan Radio
New York Times
Slate Magazine
The Today Show/Charlotte Observer
Voices of San Diego
Wall Street Journal
Washington Post
WBAL-TV (NBC affiliate in Baltimore)
Associated Press
Report: Atlanta students improving despite scandal
By Dorie Turner

ATLANTA (AP) — Atlanta students are performing better than ever before on national math and reading tests — and they're not cheating to get there, federal officials said Wednesday.

The city's fourth- and eighth-graders have made substantial gains on the National Assessment of Educational Progress — called the "Nation's Report Card" — since 2002, a report released by the U.S. Department of Education shows. A federal probe determined that Atlanta schools did not cheat on the national test, as they did on state standardized exams.

In July, Georgia investigators found widespread cheating in nearly half of Atlanta's 100 schools on state standardized tests dating back to 2001. Tens of thousands of students were affected by what experts say is the largest test cheating scandal in U.S. history. The national test is administered by independent officials rather than by the district.

Atlanta is one of 21 urban districts that volunteered to be part of the federal testing program, which is congressionally mandated to gauge how students are performing using a uniform measure.

Federal officials warned against comparing the urban districts that participated in the national test because they vary widely in student makeup, teacher experience and culture. Still, the urban districts' results mirror results released in a national report last month — students made progress in math but their reading scores have mostly remained stagnant in the last two years.

Since 2002, though, reading scores have climbed steadily in most participating cities for fourth- and eighth-graders.

"Urban schools in general are getting better. But we are determined to make them better still," said Michael Casserly, executive director of the Council of the Great City Schools. "We are not satisfied but we believe that we are on the right track — and the new NAEP data bolster our confidence."

In Atlanta, 24 percent of fourth-graders are proficient in reading, compared to 11 percent in 2002 and 21 percent two years ago, according to the report. Those scores outpace the national average for urban districts of 23 percent proficiency.

Eighth-grade math students hit 16 percent proficiency, up from just 6 percent in 2003 and 11 percent two years ago. But that trails the urban district average of 26 percent.

"The travesty of all of this is there are more and more indicators that suggest the system did not have to cheat," Atlanta schools Superintendent Erroll Davis told The Associated Press. "The educational achievement levels are not near where we want them to be, but we are continuing adding value at rates faster than other systems are adding value."
Elsewhere:
— In Boston, 27 percent of fourth-graders passed muster in reading, compared to 15 percent in 2002. And 33 percent of eighth-graders were proficient in math, compared to 18 percent in 2003.

— Chicago saw 20 percent of fourth-graders score at the proficient level in math, compared to 10 percent in 2003, while 20 percent of eighth-graders passed muster, compared to just 9 percent eight years prior.

— For Detroit, where the troubled school district is being run by the state, there were small gains since 2009 — the first year the district participated in NAEP — but the numbers are still lagging behind many other districts: 69 percent of fourth-graders scored below basic in reading, and just four percent of eighth-graders passed muster in math.

In the Atlanta cheating scandal, investigators said nearly 180 educators gave answers to students, changed answers on tests after students had turned them in or ordered subordinates to cheat. Teachers who tried to report the cheating were retaliated against and punished, creating a culture of "fear and intimidation" in the district, investigators reported.

The educators face possible criminal charges and could lose their teaching licenses. So far, eight teachers and three school administrators have lost their certification with the state. Many of the educators had resigned or retired when the report was released over the summer, but the ones remaining have been placed on leave and are in the process of being fired.

But the national test scores show that most Atlanta students were learning despite the cheating on the state test, experts said.

"The NAEP results represent the district as a whole, not this school or that school where there might have been cheating occurring," said Cornelia Orr, executive director of the National Assessment Governing Board, which administers the test. "I don't think you can assume that because it went on with a certain group of students in certain schools, gains were not across the board."

The national test doesn't come with the same pressure as the state tests, which are used to determine whether a school meets federal benchmarks, experts said. And the students who take the test are chosen by federal officials to create a sample that represents the entire district.

The testing problems in Atlanta schools first came to light after The Atlanta Journal-Constitution reported that some scores were statistically improbable. The state released audits of test results after the newspaper published its analysis.

A state probe also has led to an investigation by the U.S. education department's Office of Inspector General and the Georgia Department of Education, which says the district could owe thousands in federal money for low-income schools that have high test scores.
Austin American Statesman
Austin fourth grade math performance goes up
By Laura Heinauer

Austin fourth grade students’ performance increased in math in the past two years, according to a report released today.

The report looked specifically at students in large urban school districts who took the National Assessment of Educational Progress exam this year.

The report showed Austin students’ average score continue to be higher than other urban districts in 2011, and that the average math score was higher for Austin fourth graders in 2011 compared to what they were in 2009. The scores for fourth grade reading and for eighth grade math and reading not significantly different from what they were in 2009, the report said.

“Results from the Nation’s Report Card show that the AISD fourth and eighth grade students’ scores in mathematics and reading were among the highest in the nation,” said Meria Carstarphen in a statement. “We are extremely proud of our students.”
Baltimore Sun
Baltimore students remain in bottom third on test vs. other cities
By Liz Bowie

Baltimore’s scores on a rigorous national math and reading test were in the bottom third of large urban school districts across the country, though educators highlighted some progress in math and a promising trend of better-than-average results among some low-income black students.

Overall achievement was poor on the National Assessment of Educational Progress, a test Congress mandated be given to a sampling of students across the nation every two years. The results released Wednesday showed that the city’s children in fourth and eighth grades are scoring better than those in Detroit, Washington and Cleveland but behind those in New York, Boston and Atlanta.

"If you look at the absolute numbers, the nature of the challenge is huge," Baltimore schools CEO Andrés Alonso said. He said the results show the "urgency of the work" to be done in the school system. But deeper in the data, Alonso and others also found some reason for optimism.

Baltimore students "compared very well with African-American students in other districts," said Michael Casserly, executive director of the Council of Great City Schools, a group that represents urban districts. "We want to see more progress. I remain convinced they are on the right track with their reforms."

NAEP, also called the Nation's Report Card, began in the late 1960s and is the longest-running national assessment of basic skills. In recent years, a group of urban school districts have allowed a larger sample of their students to be tested so that the results could be compared among districts with similar demographics. Alonso decided several years ago that Baltimore would join the other districts so that the city’s progress could be charted against its peers.

This year's results showed that only 11 percent of fourth graders are considered proficient or advanced readers, and only 17 percent are proficient or advanced at math. Twelve percent of eighth graders were proficient or advanced in reading, and 13 percent ranked in those categories in math.

Those results come more than a decade after public schools in Baltimore undertook a systemwide reform, and illustrate that the city still lags behind statewide scores. According to NEAP data released earlier this year, 40 percent to 48 percent of Maryland students in math and reading are considered proficient or advanced.

Alonso points out that below the discouraging data lie some positive trends. Baltimore is doing as well as many school districts with low-income black students, who comprise more than 85 percent of the student population. When the white, Asian, and middle-income students are stripped out of the data, Baltimore’s scores look better.
For instance, Alonso said when he compared low-income students who are African-American males in the city to students in that same demographic in the other 21 urban school districts, the city ranked eighth, ahead of Atlanta, Los Angeles and Chicago — cities that are considered to have had success with education reforms.

Casserly said that African-American students in Baltimore are performing "at levels higher than what you would expect statistically" and so more analysis should be done to understand what reforms in Baltimore could be replicated elsewhere.

Baltimore also made progress in fourth- and eighth-grade math since the test was last given in 2009. The city's fourth graders made the third greatest improvement in math of the 21 urban districts. Alonso noted that those teachers are among the least experienced compared to other districts.

"So, maybe we need to stop talking about the experience of our teachers… and about what's going on in the classroom," Alonso said.

Eighth graders made the fifth largest gains in math, although that improvement isn't considered statistically significant. Scores dropped in fourth-grade reading and stayed constant in eighth-grade reading. NAEP is considered more difficult than the annual state tests, including the Maryland School Assessments, on which Baltimore students score far better.

Alonso said he agreed to join the other urban school districts in NEAP testing because he wanted to change the discussion among teachers and administrators from how to meet the targets of the No Child Left Behind law to what students in the city needed to learn. And he said he wants the system aims for a bar "much higher than the bar that has been set before us."

The Nation's Report Card has charted significant progress in school districts where the test has been given since 2003. Atlanta, for instance, has made some of the largest gains in both math and reading in both grades. Boston, Chicago, Charlotte and Washington also have made gains in some areas. In addition, the urban districts that take part have been improving at a faster rate than schools in the rest of the nation.

"It is clear that the nation's urban public schools are not only improving but are catching up," Casserly said.

Math scores rose in the urban districts in 2011, while reading scores were stagnant. The reason, educators believe, is that math is taught almost entirely at school while many of the skills needed to become a good reader are also taught at home.

The city has been more focused on improving math, including providing more training for teachers and a Saturday school for students who need extra help. "Reading is where we're going to have to show that we are ready for that kind of transformation in the classroom," Alonso said.
The NAEP results include a wealth of data apart from subject scores, including information on how often students read for pleasure. Surprisingly, fourth graders who were asked to read aloud frequently in school performed worse on the NAEP reading test than those who weren't often asked to do so. And 80 percent of students tested in Baltimore are reading aloud in class almost every day. The results also show that students who say they read for fun score better on the reading portion of the test.

"When children have access to a novel or literature that they're passionate about, they do better," said Maura Roberts, a fourth grade teacher at City Springs Elementary and Middle School who argues that the curriculum should require teachers to use more good books rather than short excerpts in the teaching of reading in the elementary grades. In addition, she said, so many students come with a limited vocabulary.

"So we really have to be explicit in how we talk to and challenge them," she said. "We use as many words as possible. We speak to them as if we're speaking to a professional."

At City Springs, which hosted the press conference to announce the NEAP results in Baltimore, students are taught reading through a prescribed program that includes having students tap to help them learn to stop at commas and periods. Fourth grader Khlil Lowther said he likes it because it helps him focus.

"I want to get better in my reading, and tapping," he said. "It makes more sense."

_Baltimore Sun reporter Erica Green contributed to this article._
Students in big-city schools show gains in latest NAEP 'report card'

By Amanda Paulson

Students in America’s largest cities are making gains in math, in many cases faster than students in the nation as a whole. Reading scores in those large cities – just as in the nation – have largely remained flat for the past two years.

And in some cities – including Atlanta, Boston, Los Angeles, and Houston – students have made particularly striking gains over the past eight years, while in other cities progress has lagged.

Most notably, the gap between national scores and large-city scores is narrowing. That’s the good news in the latest report from the National Assessment of Educational Progress (NAEP), better known as the Nation’s Report Card.

The release Wednesday provided detailed scores for students in 21 large cities – a voluntary subset that participates in NAEP’s Trial Urban District Assessment (TUDA). Ten of those cities – Los Angeles, Houston, Atlanta, Washington, New York, Chicago, San Diego, Charlotte, N.C., Boston, and Cleveland – have participated at least since 2003, giving a decent picture of how their students have fared in that time.

“We’re now down to less than a 10 scale-point difference between [large cities] and the country in reading and math for both fourth and eighth grade,” says Michael Casserly, executive director of the Council of Great City Schools, noting that that gap has closed by 25 to 35 percent in the past eight years depending on which subject and grade are examined.

“It’s clear we’re improving the numbers of kids at a proficient level and decreasing the numbers at the below-basic level – maybe not as fast as we would like, but it’s a convincing set of trend lines that tells us we’re heading in the right direction.”

The data also show wide differences in how cities perform. One fourth-grade mathematics problem, for instance, asked students to do a four-digit subtraction problem. The percentage of students answering it correctly ranged from 41 percent in Detroit to 77 percent in Austin, Texas.

The average score for fourth-grade math ranges from 203 in Detroit to 247 in Charlotte (on a 500-point scale), with a national average of 240. And in Charlotte, 48 percent of fourth-graders performed at proficient or advanced, compared with 3 percent in Detroit. The student populations also vary drastically by city.

In terms of racial makeup, for instance, about 16 percent of fourth-graders nationally are African-American, while in the 21 TUDA districts the averages range from 2 percent in Albuquerque, N.M., to 87 percent in Baltimore.

A few cities particularly shine in certain areas. In Austin and Charlotte, both fourth- and eighth-graders outperformed their peers in math in both large cities and the nation. As with the nation, reading scores from the big-city districts were largely stagnant.
Charlotte was the only district that posted an increase since 2009, for Grade 8. But in math, four districts improved their scores at Grade 4 since 2009 (Atlanta, Austin, Baltimore, and Philadelphia). Six districts did so for Grade 8 (Atlanta, Charlotte, Chicago, Detroit, Washington, and Jefferson County, Ky.).

The hope for all this data – and the reason the TUDA project was started – is to gain clues as to what’s working in certain districts, and how policymakers can continue to make a difference for often-disadvantaged urban students. NAEP results never show the cause of gains or declines, but Mr. Casserly’s organization recently completed a lengthy review of all the data through 2009, combined with case studies of what different cities were doing, and came to a few conclusions.

The study focused in particular on Atlanta, where students were making reading gains more than three times larger than in other cities or in the nation; on Boston, where students have been making similarly large gains in math; on Charlotte, where students outperform all other TUDA districts in reading and math, even after adjusting for demographics; and on Cleveland, which was the only city to not show consistent gains. In the end, six key areas seemed to make the difference:
• Stable reform-focused leadership.
• Clear goals and mechanisms for holding staff accountable.
• A common, high-quality curriculum.
• High-quality, strategic professional development.
• Good support and oversight for reform efforts.
• Data systems used to inform those reforms.

“The differences for the districts that really moved were clustered in those six areas,” says Casserly.

In Baltimore, where the TUDA results were released Wednesday, both fourth- and eighth-graders made improvements in math since 2009, though only the fourth-grade change was considered statistically significant.

In prepared remarks Wednesday, Andrés Alonso, CEO of Baltimore City Public Schools, credited numerous reforms in the past two years for the improvements, including overhauling the math curriculum, creating extended learning opportunities for students, providing parents and students with more options, and giving schools more autonomy over resources.

“The TUDA results show us where we are making progress and where we need to focus more attention moving forward,” Superintendent Alonso said, noting that the flat results in reading are unsurprising given that the emphasis on literacy has been more recent. He also praised the wealth of data available in the report.

“I urge my fellow TUDA superintendents to look not just at their scores, but to go deep with the data, use it to direct change, and share our successes and our disappointments,” Alonso said.
Math scores in some urban school districts are improving faster than the nation as a whole according to a report released Wednesday by the National Assessment Governing Board that compares math and reading scores for fourth and eighth graders.

"We continue to narrow the differences between urban school districts and the nation at large," said Michael Casserly of the Council of the Great City Schools.

Large city school districts tend to have higher numbers of students considered to be at risk than their suburban counterparts because the students often come from lower-income households or are from black or Hispanic families, groups that traditionally score lower than whites on standardized tests.

"Despite their distinct challenges, many of these districts are making steady progress in math. But, like school districts nationwide, they need to find ways to raise student achievement in reading," said David Driscoll of the National Assessment Governing Board.

Atlanta; Austin, Texas; Baltimore and Philadelphia improved their scores in grade four mathematics, while Atlanta; Charlotte, North Carolina; Chicago; Detroit; the District of Columbia and Jefferson County, Kentucky, improved their grade eight mathematics scores when compared to the last assessment two years ago. Charlotte also had an increase in its eighth grade reading scores compared to the 2009 assessment.

In a call with reporters before the release of the study, Casserly was asked about the numbers for Atlanta in light of the reports of teachers changing student responses on the Georgia state standardized tests.

"There appeared because of the state investigative report that there was cheating by some teachers on the state test but at the same time what you saw by way of reform in the school district was real," said Casserly. "There's no reason to believe that the results on the national assessment should be called into question at all."

Some large school districts have shown improvement over the long run while others have not. Analyzing data from 2009 and before for Atlanta, Boston, Charlotte-Mecklenburg and Cleveland, the Council for Great City Schools found that the instructional practices of the first three districts were quite different than those for Cleveland, which showed lower performance and lower gains.

Differences were found in areas including leadership and reform vision, accountability, curriculum and instruction, teacher quality and use of data and assessments. Not all large cities take part in the comparisons.
Twenty-one districts volunteered to have their scores compared in the 2011 assessment, including Atlanta; Austin; Baltimore; Boston; Charlotte-Mecklenburg; Chicago; Cleveland; Detroit; District of Columbia; Fresno, California; Houston; Jefferson County (Louisville, Kentucky); Los Angeles; Miami-Dade County; Milwaukee; New York; San Diego; Philadelphia and, for the first time, Albuquerque, New Mexico; Dallas and Hillsborough County, Florida.
Urban Districts Post Gains in NAEP Math
By Erik W. Robelen

Reading achievement in a set of large urban districts has stayed mostly flat since 2009, based on new national test results, while in mathematics, half the school systems saw some growth over the past two years, including Atlanta, the district at the center of a recent, high-profile cheating scandal.

In math, four out of 18 big-city districts posted statistically significant 4th grade gains from 2009 to 2011, while six out of 18 made progress at 8th grade, according to data released today from the Trial Urban District Assessment, which tests representative samples of students on the National Assessment of Educational Progress, known as “the nation’s report card.”

Atlanta was the only district to make math gains at both grade levels since 2009. In reading, meanwhile, the Charlotte-Mecklenburg, N.C., school district was the only participant to see reading gains of statistical significance since 2009, and those were only at the 8th grade level.

The longer historical view tells a more hopeful story, however. Nearly all the districts to participate in TUDA since the early 2000s have made gains in both subjects. For instance, all six urban systems that took part in the 4th grade reading exam back in 2002, when TUDA began, posted higher scores this round. And in math, nine of 10 districts gained ground since 2003 both at the 4th and 8th grades. (Cleveland saw no statistical changes.)

Atlanta’s academic growth since 2009, and over the long haul, seems especially noteworthy, given that the system has been reeling from a state investigation that found widespread cheating on state assessments for a number of years.

In a conference call with reporters early this week, a federal official emphasized that, based on a separate federal investigation, there was no reason to believe the cheating extended to NAEP.

“The short story is we couldn’t find any evidence [of cheating],” said Jack Buckley, the commissioner of the National Center for Education Statistics, which administers NAEP. “What happened is virtually impossible on NAEP” because of its format and other factors, he added. Atlanta’s strongest gains were in 8th grade math, where the district climbed 6 points on the 500-point scale since 2009, and 22 points since 2003.

Measured another way, the proportion of Atlanta students scoring “basic” or above climbed from 30 percent in 2003 to 54 percent in 2011. In reading, Atlanta saw no statistically significant changes since 2009, but climbed by 16 points since 2002 at 4th grade and by 17 points at 8th grade.

“There was cheating by some teachers on the state tests, but at the same time, what you saw by way of reform in the district was real,” said Michael Casserly, the executive
director of the Washington-based Council of the Great City Schools, who took part in
the conference call.

**Narrowing the Gap**
The latest TUDA results come about a month after a new round of NAEP data in
reading and math for the nation was released. That data found average reading scores
flat in the 4th grade, but a slight increase in 8th grade reading as well as math at both
grade levels.

The new TUDA report provides a close look not only at the set of participating districts,
but also highlights average achievement for students across all large U.S. cities and
suggests the achievement trend here largely mirrors the nation as a whole. However,
the pace of improvement for large cities appears to be eclipsing that of the nation.

For example, 8th grade math scores for students from big cities climbed 3 points from
2009, compared with 1 point for the nation. In taking a longer view, large cities gained
12 points since 2003, compared with 7 for the nation.

In 4th grade reading, there was no significant change since 2009 for the nation or for
large cities, but urban students posted a 9-point increase since 2002, compared with a
3-point rise in the national average.

“We’ve continued to narrow the differences between urban school districts and the
nation at large,” said Mr. Casserly. He noted that the gap for the first time was less than
10 points in all four categories—that is, in both grade levels and subjects. Mr. Casserly
said his organization earlier this fall issued a report on the factors that help to explain
why some urban districts have shown greater gains than others.

The report identifies some common threads among urban districts that have shown
strong improvements over time. They include strong and stable leadership and setting
clear, systemwide goals and a “culture of accountability” for meeting them, as well as a
coherent, well-articulated program of curriculum and instruction and regular use of data
to gauge learning, modify practice, and target resources.

**Detroit Makes Headway**
Although Atlanta was the only TUDA district to make gains in math at both the 4th and
8th grades since 2009, half the participants did see significant improvement in one or the
other grade. Those to do so at the 4th grade were Austin, Texas; Baltimore; and
Philadelphia. At the 8th grade, the other districts to improve were Charlotte-
Mecklenburg, Chicago, Detroit, the District of Columbia, and Jefferson County, Ky.

The largest gains posted in 8th grade math were for Detroit, which climbed 8 points
since 2009. But even with those gains, its score of 246 was the lowest of all 21 districts
to participate in TUDA this year. Measured another way, 71 percent of Detroit’s 8th
graders scored below basic in the subject. And only 4 percent were “proficient” or
“advanced.” This year, three school districts in urban locations participated in TUDA for
the first time: Albuquerque, N.M.; Dallas; and Hillsborough County, Fla. They join 18
others that have taken part more than once.
Miami-Dade students continue to outperform peers in national test
By Laura Figueroa

National test scores released Wednesday show that Miami-Dade students outperform students in the nation’s top four largest school districts.

Miami-Dade may be the nation’s fourth-largest school district, but when stacked against its larger counterparts — New York, Los Angeles and Chicago — the district tops the list in national reading and math scores, according to testing data released Wednesday.

The results of the 2011 National Assessment in Education Progress, dubbed the “nation’s report card,” show that Miami-Dade has largely been able to score higher than other large urban school districts and remain on par with its peers nationwide.

Held as the gold standard for comparison between districts, the NAEP exams are given every two years to a sample of fourth- and eighth-grade students throughout the nation to compare student achievement from state-to-state.

The U.S. Department of Education also pulls data from 21 urban school districts, including Miami-Dade, where the majority of students are minorities or classified as poor, to determine how those students measure up against their peers nationwide.

In 2009, Miami-Dade’s first time participating in the sample, the district received national praise for its scores among Hispanic and black students. The district managed to keep those scores steady, only dropping by 1 or 2 percent in most categories.

“Recognizing a number of factors, including growing poverty in our community, the fact that we were able to hold the ground from our 2009 scores, that we remain a national leader in the student achievement, I was fairly gratified,” Miami-Dade Superintendent Alberto Carvalho said.

Fourth-grade reading scores show that 67 percent of test takers scored at or above grade reading level, which is just above the national average of 66 percent, but one point below the district’s score in 2009. Math scores for fourth-graders show that 79 percent of students had basic to advanced skills, one point below the district’s number in 2009 and just two points below the national average of 81 percent.

Nearly 70 percent of Miami-Dade’s eighth-graders scored at or above grade reading level, two points below the district’s score in 2009, and five points below the national average. Eighth-grade math scores indicate that 62 percent of students scored basic or higher on their exams, some 11 points below the national average.

“It speaks well for South Florida, particularly for Miami-Dade, that a diverse urban district can perform to the national level and in many cases outperform other large urban school districts,” Carvalho said.
Though Broward is the nation’s sixth-largest district, it is not included in the urban measure, because not enough students meet the poverty criteria. Miami-Dade was initially the only Florida district included in the urban sample, but this year Hillsborough County was added. The Tampa-area district posted scores that were on average slightly higher than Miami-Dade’s. Both districts and their superintendents were heralded by Florida Education Commissioner Gerald Robinson.

“The commitment and leadership in these districts represent Florida’s focused mission to reduce the academic achievement gap while supporting our students in their pursuit of excellence,” Robinson said in a prepared statement.

Overall, test administrators say urban school districts have steadily increased their scores since 2003, closing the gap between large city districts and smaller suburban districts.

“We know that we’re still behind, but we seem to be narrowing the gaps,” said Michael Casserly, executive director of the Council of the Great City Schools.
Detroit students show small gains on national standardized test
By Sarah Cwiek

Students in Detroit Public Schools showed slight improvements in the latest round of a benchmark standardized test. But Detroit students still posted the worst scores of any district in the country on the National Assessment of Educational Progress test.

The NAEP exam tests fourth and eighth-graders in reading and math. When Detroit students took the test for the first time in 2009, they produced the worst scores in the test’s history. In 2011:
- Math, 4th grade: proficiency up from 31% to 34%
- Math, 8th grade: proficiency up from 22% to 29%
- Reading, 4th grade: proficiency up from 27% to 31%
- Reading, 8th grade: proficiency up from 41% to 43%

Some experts question whether the posted gains are even statistically significant. But Detroit schools’ emergency manager Roy Roberts says the important thing is a positive trend.

“Detroit had the highest gains of any city in any subject in mathematics,” Roberts says. “Detroit also exceeded the state in gains in reading.”

“Like the budget deficit, it will not be eliminated overnight. But we have demonstrated real progress.” Roberts says the biggest thing hindering Detroit students’ academic performance has been “instability” in the district.

He says the district will announce in January how many more schools to close, charter, or move to the Education Achievement System, a new statewide district for the lowest-performing schools.
New York Times

New York’s Math Scores Dip on U.S. Student Tests, Diverging From Trend in Other Big Cities,
By Winnie Hu

New York City students scored slightly lower on federal math tests this year compared with two years ago, according to results released on Wednesday, even as test scores of their counterparts in other big cities inched upward.

The results from the National Assessment of Educational Progress, also known as the nation’s report card, showed that the city’s fourth-grade math average dropped three points to 234 (on a scale of 500) from 2009, the last time the exams were taken.

Federal education officials cautioned that the changes were too small to be significant, but the dip diverged from the trend nationally and for other large cities. In 2011, the average fourth-grade math score rose by one point nationally and two points for cities with 250,000 or more people.

For the city’s eighth graders, the average math score dropped one point from 2009, to 272; nationally, the average eighth-grade math score rose by one point, and three points for students in large cities.

On reading tests, the city’s fourth-grade average score remained flat at 216; the national score also stayed flat at 220. The average eighth-grade reading score rose by two points in New York, to 254, a positive turn. Nationally, the eighth-grade average score was up a point.

The results for the city’s students generally matched what happened across New York State on fourth- and eighth-grade math scores, which declined, and on eighth-grade reading scores, which rose. But while the statewide fourth-grade reading score declined, the city’s held steady. And even with the recent declines, the city’s fourth- and eighth-grade math scores are still up eight points and six points, respectively, since 2003.

Allison Horowitz, a policy analyst with the Education Trust, an advocacy group promoting academic achievement, said New York City’s overall averages had not improved as much as in other large cities, though there had been “good progress” in moving students up from the lowest achievement level since 2003.

“Parents in New York City have the same high aspirations for their kids as other parents across the country,” she said. “And we’re not going to get all students where they need to be if all we’re doing is lifting the floor.”

Indeed, the 2011 results show the city has made little, if any, progress in reaching the lowest-performing students since 2009. The percentage of students performing below basic achievement levels actually grew by one point in fourth-grade reading and eighth
grade math, and three points in fourth-grade math. Once again the only improvement came in eighth-grade reading, with a three-point drop.

At the top end, the percentage of students performing at or above proficiency level shrank by two points each in fourth-grade and eighth-grade math from 2009. Among fourth graders, 33 percent are at or above the proficient level in math in 2011, and 76 percent are at or above the basic level — both below national percentages, but better than those of other large cities. There was a two-point increase in the percentage of students at or above proficiency in eighth-grade reading, while fourth-grade reading was unchanged.

But New York City showed significant improvement in narrowing the achievement gap for poor students over a decade. Eighth-grade students eligible for free and reduced lunch scored 14 points lower on reading than those who were not eligible in 2011, compared with 30 points lower in 2003.

The achievement gap for blacks also appeared to shrink slightly in the city. In 2011, black students averaged 26 points lower than white students on reading tests in fourth grade, compared with 29 points lower in 2002. In math, they averaged 22 points lower in fourth-grade and 30 points lower in eighth grade, compared with 25 points and 36 points lower in 2003.

Girls generally did better on reading tests, averaging nine points higher in fourth-grade, eight points higher in eighth-grade. On math tests, there was little difference between sexes.
Slate.com
Urban NAEP Scores Show Slowly Brightening Education Picture
By Matthew Yglesias

Today the latest round of Trial Urban District Assessment data about American school performance came out. On the reading front it showed no statistically significant change over the past two years in fourth graders in any of the participating cities, and statistically significant (and positive) change for eighth graders in Charlotte while everyone else was insignificant.

When you combine all the cities into a large sample, it shows a modest overall statistically significant improvement. In math we fourth graders improving in Atlanta, Austin, Baltimore, and Philadelphia and eighth graders improving in Atlanta, Charlotte, Chicago, Detroit, DC, and Louisville. There are no statistically significant declines anywhere.

I don't think you can draw any sweeping conclusions from this, but I do think it's worth laying it out there as a baseline. Most of the commentary I read about public education in America is very negative but to the best we can tell things are gradually improving even during a very difficult social and economic period for the country.

The math gains in Atlanta seem especially noteworthy since a lot of attention has been paid to a big cheating scandal on the Georgia state tests. It looks, however, like that cheating was occurring against a background of real learning gains for the city's kids.

The thing I do worry about, however, is that lots of cities don't participate in TUDA. Policy varies greatly across the participating cities, but one thing they have in common is that they all want to know how they're doing, which I think is the baseline beginning for improvement.

But some very large school districts -- most notably Clark County in Nevada, a bunch of non-Miami Florida districts, and Dallas, TX -- seem to prefer to wallow in ignorance. There's really no excuse for this.
CHARLOTTE, N.C. -- Charlotte-Mecklenburg Schools topped most other big cities on this year's national reading and math tests, according to results released this morning.

CMS is among 21 urban districts reporting fourth- and eighth-grade reading and math scores on the National Assessment of Educational Progress, also known as NAEP or the nation's report card. CMS rated first or second in all areas.

The strong comparative showing affirms the recognition CMS earned this fall, when it claimed the Broad Prize for Urban Education. But the results also reinforce a theme sounded at that ceremony: Academic progress has been painfully slow for poor and minority students, even in the most successful districts.

Fewer than 20 percent of low-income and black CMS eighth-graders were rated proficient in either subject, falling far below white and middle-class counterparts. For instance, 16 percent of African American eighth-graders in CMS scored proficient in math, compared with 66 percent of white classmates. Yet Charlotte's black eighth-graders topped the national average and outperformed counterparts in most of the urban districts.

CMS officials are holding a news conference this morning to discuss the NAEP results. The national report repeatedly highlights Charlotte's success. The NAEP was launched in 1969 to gauge national progress and compare state results. In 2002 it started breaking out results for a handful of urban districts, and CMS joined that group in 2003.

CMS; Austin, Texas; and Hillsborough County (Tampa), Fla., juggled the top three spots in each category. With 52 percent of students qualifying for federal lunch aid, CMS has the lowest poverty level of all the urban districts tallied, matching the national average for all public schools. Austin and Tampa are next-lowest. Districts at the bottom of the rankings, such as Detroit, have very few white and non-poor students.

But the edge at CMS and the other high-ranking districts doesn't come solely from having more of the students who traditionally do well on tests. They also fare better than average on same-group comparisons.

NAEP doesn't report results for individual schools, and not all students take the exams. A representative sampling is tested in participating districts. The reading and math tests do not include high schools. Read details at www.nationsreportcard.gov (see "Trial Urban District Assessments").
Voices of San Diego
Math Gap Grows Between Black and White Students
By Emily Alpert

The gap between math scores for black and white students in San Diego Unified grew wider on a national exam despite growing attention to math in the school district. It's a troubling change that school officials are still trying to understand.

While 58 percent of white eighth graders scored proficient or above on the math tests, only 8 percent of black students did. White students had improved continuously over years; black students improved and then dropped this year.

The math gap between poor students and their better-off classmates also grew. The results are especially disappointing because San Diego Unified devoted more attention to math last year, after Superintendent Bill Kowba announced that it would be a major focus for improvement.

It also just launched a new plan to improve African American student achievement, though it hasn't put any money behind that plan so far. These results underscore why it is paying special attention.

Deputy Superintendent Nellie Meyer said they are still analyzing the results, but she fears budget cuts are part of the problem. The school district has shrunk its day-to-day spending over the last four years, cutting back on summer school, reducing tutoring and paring back school site budgets.

"We have a safety net for students who need it and we're slowly cutting it away," Meyer said.

San Diego students take a smattering of different exams, but what makes this one unique is that it provides a common yardstick to compare San Diego Unified to other urban school systems across the country. State tests differ from state to state; this exam gives a window into how students do nationally.

The national exam is given every other year to a sample group of students in fourth and eighth grade in a smattering of urban school districts. It gauges math and reading skills.

San Diego Unified has improved significantly since it started taking the national tests eight years ago. It tends to perform well on these tests compared to the average urban district, perhaps partly because it has fewer poor children than other districts that take the exam. It also outperforms the California average despite having more poor children and more English learners.

But the recent gains are so slight that they didn't make a statistical blip. And math scores for eighth graders actually dropped, although the drop was so small it doesn't count statistically either.
The small improvements are puzzling because in the same years, San Diego Unified has made notable strides on state tests. The same gap showed up last year, when San Diego Unified showed only slight growth on the national exam yet surged on state tests.

Ron Rode, who oversees assessments, believes that the national exam measures more critical thinking and problem-solving than the state tests. The slow growth makes him wonder whether the school district is doing enough to cultivate skills that go beyond the simpler questions on state tests. Beefing up critical thinking is another thing Kowba promised to push; these results suggest there is more left to do.
Large urban school districts have made steady progress on national elementary school math and reading exams over the past nine years but continue to score far below national averages, according to federal data released Wednesday.

Results of the 2011 National Assessment of Educational Progress show that scores in urban districts rose slightly or remained flat since the exams were last given in 2009—similar to the national performance.

But a more promising picture emerges when trend lines are extended back to the early 2000s. Students in cities such as Chicago, Atlanta and Houston posted double-digit gains on several exams since 2002, helping close the chasm between their performance and that of districts nationwide.

The results will likely fuel the debate over how to improve the lowest-performing schools, many of which are in cities. The data also come amid an intensifying battle between congressional Republicans and the Obama administration over how to revamp No Child Left Behind, the federal education law that some say propelled the test score gains in inner cities.

The Council of Great City Schools, a research and policy group that represents large districts, commissioned a study that found districts making the most progress had stable leadership, high academic goals for students, quality professional development for teachers and data analysis that helped alter teaching. The study was funded by the Bill & Melinda Gates Foundation.

By law, the U.S. Department of Education gives the math and reading tests to a representative sample of students in every state biennially. Ten cities have volunteered for the comparative study in math since the exam was first given in 2003.

All of them, except Cleveland, posted statistically significant gains, with most surpassing the growth made nationwide. Boston schools, for example, scored 237 out of 500 on fourth-grade math, a 17-point jump since 2003. Washington, D.C., schools saw a 17-point bump to 222, and Chicago a 10-point increase to 224.

Nationwide, the average math score increased six points to 240 in the period. In eighth grade math, nine of the 10 districts equaled or surpassed the average national improvement of seven points—only Cleveland didn't.

Still, in most of the 21 urban districts tested this year, 75% of their students did not score "proficient" in math. Proficient means students have a solid grasp of the material.
In fourth-grade reading, six urban districts have participated since the exam was first given in 2002 and each posted gains of at least twice the national average of three points.

In eighth grade, the national reading score hasn’t budged since 2002, but students in Atlanta, Houston and Los Angeles posted gains, while the other cities remained steady. Atlanta Public Schools, jolted by a state-test cheating scandal this year, made the most improvement. Federal officials said they have analyzed the scores and feel confident cheating didn't take place in the U.S. exams, noting the exams are administered and scored by federal contractors.

"The real travesty of the cheating scandal is that it clearly didn't need to take place," said Erroll Davis Jr., who became superintendent of Atlanta schools when his predecessor resigned amid the cheating probe. "As you can see from the data, our teachers continue to add spectacular value to our students."

The federal study also showed that Detroit and Cleveland—Rust Belt cities losing population and economic resources—continue to rank at the bottom. In Detroit, the ailing district has been under the control of a state-appointed emergency manager for almost three years. Roy Roberts, appointed by Republican Gov. Rick Snyder of Michigan to oversee the schools, said he anticipates improvement as the district stabilizes its academics and finances.

LaMarr Mitchell, an eighth grader at low-income City Springs Elementary Middle School in Baltimore, which hosted the national press conference on the test scores, attributed his school's success to teachers who "make school fun, so it's easy to learn." He said he has written plays in his language-arts class and competes in school math tournaments that pit girls against boys.

—Matthew Dolan contributed to this article
D.C. public schools have the largest achievement gap between black and white students among the nation’s major urban school systems, a distinction laid bare in a federal study released Wednesday.

The District also has the widest achievement gap between white and Hispanic students, the study found. The study is based on the 2011 National Assessment of Educational Progress, federal reading and math exams taken this year by fourth- and eighth-graders across the country.

The tests are the only continuing and nationally representative assessment of what students know. State-by-state results were released last month, but large cities have agreed to have their own results published separately since 2002, with 21 participating this year.

Generally speaking, the results in large cities mirror national trends: Students are showing some improvement in math, but progress in reading is stagnating. In reading and math, the gap in scores between black and white students was widest in D.C. schools compared with those in 20 other urban systems, including New York City, Los Angeles and Miami.

The D.C. gap was also greater than the national average and the average for cities with populations of 250,000 or more, according to the study. On the fourth-grade math test, for example, black students in the District scored an average of 223 points out of a possible 500 while their white classmates averaged 272, or 49 points higher. That difference is more than twice the national achievement gap for that test.

The achievement gap has proved to be a stubborn problem, and one of growing concern among educators, policy makers and civic leaders. With enactment of the No Child Left Behind law in 2002, the federal government made closing the gap a priority and a reason for increased accountability in public education. A host of strategies have been deployed by schools across the country to attack the gap, but few have resulted in substantial progress.

All 21 cities in the study displayed a difference in performance between whites and blacks and between whites and Hispanics. But in every case, their variations were less than in the District — in some cases five times smaller. In the fourth grade math example, for instance, Cleveland’s black and white students were separated by a 14-point gap.

The District’s racial gap is really an income divide, said Michael Casserly, executive director of the Council of the Great City Schools, which represents the largest urban school systems.
“You’ve got relatively more well-to-do whites in Upper Northwest quadrants, particularly Ward 3, which score higher than white students nationally and you’re comparing it with poor, African American students largely in Wards 7 and 8,” Casserly said. “There are extreme income disparities.”

Atlanta, which also has a sizeable achievement gap, presents the same problem, Casserly said. “You’ve got white students in the Buckhead section scoring off the charts but when you compare with poor African Americans that make up most of Atlanta, you’re looking at an achievement gap that is similar to the District’s,” he said.

And although Cleveland appears to have a narrow racial gap, the small difference between test scores of black and white students is linked to the fact that both groups are relatively low-income, Casserly said.

“You’ve got poor Appalachian whites in Cleveland and poor African American students,” he said. “You’re looking at a wealth factor.”

The District’s racial achievement gap is a long-standing pattern, he said. But it’s difficult to say whether that gap has changed over time relative to the other cities because for much of the past decade, there haven’t been enough white students in D.C. taking the test to reliably draw conclusions, according to the National Assessment Governing Board, which sets policy for the test.

Currently, 46,191 students are enrolled in D.C. public schools, with about 79 percent African American, 12 percent Hispanic, 7 percent white and 2 percent self-declared “other.”

The new study did not include test scores of students who attend public charter schools in the District, which now educate about 40 percent of public schoolchildren. An analysis of the test scores of D.C. public charter students in 2011 showed that black students attending charters scored higher in math and reading tests in both fourth and eighth grades than their counterparts in traditional District schools. The number of white students attending public charters in the District was too small to draw a conclusion about their academic performance compared with any other group.

Overall, the District placed at or near the bottom of the 21 cities in the study in terms of scores for math and reading in fourth and eighth grades; D.C. tied with Detroit for last place in eighth grade reading.

The school systems that consistently scored at the top of the heap were Charlotte, which was either No. 1 or 2 in every category; Hillsborough County, Fla.; and Austin.

Staff writer Bill Turque contributed to this report.
WBAL-TV
4th-Grade Math Results On Rise In City: NAEP Test Results Released

BALTIMORE -- Maryland fourth-graders are making progress in math, according to results from a national test that were released Wednesday.

11 News education reporter Tim Tooten said the results are from the NAEP test -- or National Assessment of Educational Progress. It’s a test that compares how well Baltimore students are doing with students in 20 other major urban cities, Tooten said.

Tooten on Wednesday visited City Springs Elementary-Middle School, where the students have a love for math.

“I don't think math is hard, because math isn't hard if you pay attention and get through it,” said fourth-grader Chelsea Gilmer.

Fourth-grade math scores were up slightly in the NAEP, which compares city students to those in cities like Philadelphia, Washington, D.C., and New York. Tooten said that over a two-year period, fourth-grade math scores were up 3.4 percent. There was also an increase for eighth-graders, but it was below the nation’s best.

“I think these are gratifying results,” said Baltimore City Schools CEO Andres Alonso. “Results that underline the urgency around how our kids are doing, and really useful information in terms of what to do about our core mission.” Alonso was part of a national panel which released the NAEP, Tooten said.

“For we are not interested in reflecting or perpetuating the inequities under which too many of our children suffer, our job is to overcome them, so poverty, race and language never define our kids' chances for a brighter future,” said Michael Casserly, executive director of Great City Schools.

Baltimore school officials said there is plenty of room for improvement across the board in reading and math, but they seemed most impressed with the progress of African-American male students, Tooten said.

Their math scores increased 1.8 percent for fourth-graders, and 3.5 percent for eighth-graders. Educators said the progress has a lot to do with what's taking place in the classroom.

“We try to approach math as a science, so the children are constantly discovering new things, and they are coming with these theories on their own,” said fourth-grade teacher Maura Roberts. “And they're learning that's how it works, that's how it applies.”

The results represent only a sampling of Baltimore fourth- and eighth-graders, Tooten said.
## Upcoming NAEP Reports as of March 2012

<table>
<thead>
<tr>
<th>Report</th>
<th>Expected Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial NAEP Releases</strong></td>
<td></td>
</tr>
<tr>
<td>2005 High School Transcript Study: Mathematics Course Content Analysis</td>
<td>April 2012</td>
</tr>
<tr>
<td>2009 Science Hands-On Tasks and Interactive Computer Tasks: Grades 4, 8, and 12</td>
<td>May 2012</td>
</tr>
<tr>
<td>2011 Science Report Card: Grade 8</td>
<td>May 2012</td>
</tr>
<tr>
<td>2011 Writing Report Card: Grades 8 and 12</td>
<td>July 2012</td>
</tr>
<tr>
<td>2011 National Indian Education Study: Grades 4 and 8</td>
<td>July 2012</td>
</tr>
<tr>
<td><strong>Other NAEP Reports</strong></td>
<td></td>
</tr>
<tr>
<td>2011 Meaning Vocabulary: Grades 4 and 8</td>
<td>July 2012</td>
</tr>
<tr>
<td>Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade</td>
<td>December 2012</td>
</tr>
<tr>
<td><strong>NAGB Reports</strong></td>
<td></td>
</tr>
<tr>
<td>Mega-States Report: Grades 4, 8, and 12</td>
<td>July 2012</td>
</tr>
<tr>
<td><strong>Other Related Reports from NCES</strong></td>
<td></td>
</tr>
<tr>
<td>Reading, Mathematics, and Science Achievement of Language-Minority Students in Grade 8</td>
<td>February/March 2012</td>
</tr>
<tr>
<td>Digest of Education Statistics, 2011</td>
<td>February/March 2012</td>
</tr>
</tbody>
</table>
NATIONAL ASSESSMENT GOVERNING BOARD
RELEASE PLAN FOR
NAEP SCIENCE 2011 REPORT

The Nation’s Report Card in Science 2011

The Nation’s Report Card in Science 2011 will be released to the general public during May 2012. Following 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. Full accompanying data will be posted on the Internet at the scheduled time of release.

The Report Card presents results of the National Assessment of Educational Progress (NAEP) from a representative sample of about 122,000 8th graders at the national and state levels. Results will be reported in terms of scale scores and the percentage of students at or above achievement levels. In addition to results for the nation as a whole, the report will include national and state level NAEP results for various demographic groups. Information about the Science Framework will be included, along with examples of questions and student responses.

This assessment was purposely scheduled for 2011 so that the data could be linked with the Trends in International Mathematics and Science Study (TIMSS) that was also administered in 2011. The NAEP-TIMSS linking report will be issued following release of the international TIMSS results in December 2012.

DATE AND LOCATION

The release event for the media and the public will occur in May 2012. The exact date and location 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 member of the National Assessment Governing Board
- Data presentation by the Commissioner of Education Statistics
- Comments by at least one Governing Board member
- Questions from members of the press and then the general audience
- Program will last approximately 60 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.
EMBARGOED ACTIVITIES BEFORE RELEASE

In the days preceding the release, the Governing Board and NCES will offer embargoed briefings to U.S. Congressional staff in Washington, DC. Representatives of governors, state education agencies, and appropriate media will have access to a special website with embargoed data after signing the Governing Board’s embargo agreement.

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 various other resources, will 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, publications 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, key graphics, and audio and/or video material related to the event.

ACTIVITIES AFTER THE RELEASE

The Governing Board’s communications contractor, Reingold-Ogilvy, will work with Board staff to coordinate an in-person or online event designed to extend the life of the NAEP Science results by featuring current topics that would be of great interest and relevance to stakeholders. The event will be designed for organizations, officials, and individuals in the fields of education and policy who have an interest in science education and assessment.
NOTE TO Reporting and Dissemination Committee on NAEP Testing and Reporting of SD and ELL Students

Based on recommendations by two expert panels, the Governing Board adopted a new policy in March 2010 on NAEP Testing and Reporting on Students with Disabilities (SD) and English Language Learners (ELL). The policy covers how such students should be tested by NAEP and how NAEP reporting should take into account their participation or exclusion from state and district samples. The policy also contains a research and development agenda for additional improvements.

At the March 2 meeting NCES will present an update on implementation of the policy, including changes in the 2011 Report Cards, research on targeted testing at different performance levels, and plans for the introduction of new rules for testing SD and ELL students in 2013.

Information on these activities as well as the text of the Board’s SD and ELL policy is included in this attachment.

The attachment also includes a letter from Florida Education Commissioner Gerard Robinson, requesting further action because of the continued wide differences in SD and ELL participation rates for the states and urban districts in NAEP. Commissioner Robinson said the variations “call into question the validity” of NAEP’s state comparisons. He proposes that NAEP not report data for jurisdictions falling below the inclusion goals set by the Board—95 percent of all students selected for a NAEP sample and 85 percent of students in the sample who are identified as either SD or ELL. Starting in 2011, the goals have been included in NAEP Report Cards along with information on which jurisdictions met them and which did not. Robinson proposes that “states not meeting the minimum standards should face funding sanctions.”

The attachment includes responses to his letter by Board Chairman David Driscoll and NCES Associate Commissioner Peggy Carr plus several news articles.

Some issues to consider:

Should the goals and targets in the Board policy for SD and ELL student participation become the standards for reporting state and district data on NAEP?

Should some other, lower rate be adopted as the minimum for reporting, such as the 70 percent school participation rate used for publishing private school results?

Should the issue of SD and ELL reporting standards be studied by an expert panel?

Should consideration of any changes in reporting criteria be deferred until after implementation of the new testing rules in 2013?
Governing Board Policy Implementation in 2011

Policy Principle 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.

- NAEP 2012 includes a pilot test of the new decision tree for including students with disabilities and English language learners.
- Other special studies have been completed to determine possible accommodations for NAEP (Calculator Booklet Study, Inclusion Booklet (KaSA) Study).

Policy Principle 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.

- Each state develops state-specific guidelines for including students with disabilities and English language learners on NAEP. The guidelines include the allowable accommodation for NAEP and spell out the state’s expectation for including students on NAEP. The guidelines are included with the preassessment visit packet with the SD/ELL worksheets.
- NAEP state and TUDA coordinators monitor exclusions during the preassessment window and contact schools that are not following the state’s expectations for including students on NAEP and encourage these students to be included on NAEP.

Policy Principles 3 and 4

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.

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.

- Since the inclusion goals were implemented in the reporting of NAEP 2011, most states and districts worked to increase inclusion on NAEP.
Policy Principle 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.

- See the attached Inclusion Fact Sheet. This was developed through the NAEP 2012 Field Publications meetings with input from NCES and NAEP contractors.

- The Inclusion Fact sheet was included in the preassessment visit packet to provide context for including students in NAEP. SD/ELL worksheets are completed by school staff that are most knowledgeable about how each student should be included in assessments and determine how the student should participate in NAEP.

Policy Principle 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.

- In fall 2010, prior to the administration of NAEP 2011, NCES and the NAEP State Service Center hosted Inclusion Workshops for state and district officials to explain the new Governing Board policy on inclusion. NAEP state and TUDA coordinators attended with a team of state and district representatives (assessment directors, exceptional children specialist, English acquisition specialist, and accommodation specialist) to develop strategies for increasing inclusion on NAEP and plans to communicate the inclusion expectations with school-level decision makers.

- The Inclusion Workshop will be held again on September 11-12, 2012 to prepare for the administration of NAEP 2013.
2011 Knowledge and Skills Appropriate (KaSA) Study

Overview of the Study

NAEP has had difficulties measuring the abilities of lower-performing students nationally, as well as in comparatively low-performing jurisdictions. In an effort to obtain more information on what low-performing students know and can do, new fourth- and eighth-grade mathematics blocks of items were developed to be more knowledge and skills appropriate (KaSA) for such students.

These blocks of KaSA items were administered, along with the regular operational math blocks, in the mainland U.S. and in Puerto Rico to a small special study sample in 2011. The target sample size for each grade was roughly 9,000 students (5,000 in Puerto Rico and 4,000 in the mainland U.S.).

One goal of the analysis is to determine whether the addition of KaSA items does in fact provide better measurement of the knowledge and skills of Puerto Rico’s students, for whom it has been difficult to obtain reliable results in previous NAEP administrations. Although the 2011 KaSA study focused on the use of these items in Puerto Rico, the results have implications for the possible use of KaSA items to aid in the measurement of lower-performing students throughout the nation.

The Mathematics Computer-Based Study

In 2011, NAEP administered the Mathematics Computer-Based Study (MCBS) to a sample of 8th-grade students as an exploration of the feasibility and potential advantages of an adaptive testing model in the NAEP context. The primary goals of the study were to administer items better targeted at students’ ability levels in order to:

- Improve student engagement in the NAEP mathematics assessment; and
- Reduce measurement error across a wider range of ability levels than is possible with the current NAEP administration model.

As the goal of the study was to understand the potential impact of this alternative design in NAEP, and to provide a roadmap for moving forward with an adaptive approach, the study had both experimental and control conditions. In the experimental condition, students were randomly assigned one of two routing blocks, which were scored by the computer. Based on their performance in the routing block, a decision was then made to assign students an easy, medium, or difficult block. In the control condition, students were also randomly assigned one of the two routing blocks, but the assignment of the second block was not based on performance in the first block. Instead, students were randomly assigned to the easy, medium, or difficult blocks.

The blocks used in this study were constructed of items in the existing item pool, and adapted for administration on the computer. As such, this was not considered a study of new or innovative technology-
enhanced items. The primary focus was on the potential implementation of an adaptive model for NAEP. Some of the research questions for this study include:

1. Does multi-staged testing improve students’ engagement?
2. What are the implications of this administration model for scaling?
3. Is measurement error reduced, especially in the lower range of ability?
4. How effective and efficient were the routing decisions that were made?

NCES is currently analyzing the data from this study, and will share the results as they become available. We anticipate that the results of this study will help design a next generation of assessment that will improve our ability to measure a wider range of student performance effectively and efficiently.
2011 and 2012 NAEP Studies on Testing and Reporting SD and ELL Students

2011 Mathematics Inclusion Study

The 2011 mathematics inclusion study is designed to determine if excluded students can meaningfully participate in NAEP with one of two special booklets, thereby increasing inclusion rates. The two types of special booklets used in the study were (1) a calculator booklet, and (2) a less demanding (KaSA) booklet.

The study included two passes through the SD decision tree with the first pass conducted in the standard NAEP fashion. After all inclusion decisions had been made via the first pass, NAEP field staff reviewed with school coordinators the list of excluded students and offered the special inclusion booklets as appropriate. Excluded students who use a calculator as an accommodation on the state math assessment were offered the calculator special study booklet. Excluded students who do not use a calculator as an accommodation on the state math assessment were offered the KaSA booklet. These students were counted as excluded in the 2011 assessments, and their responses were not part of the results that were published.

One purpose of this study is to see how inclusion rates would have changed if these students had been considered “assessed” instead of excluded. The second component of analysis is to examine the effect on student performance had these students been included in the results. To accomplish these analyses, the entire operational analysis will be repeated with the special study students included. The results (i.e., average scale scores and achievement level percentages for the nation and states) will be compared to the published results.

2012 SD and ELL Decision-Tree Pilot Studies

Goals of the Governing Board Policy “NAEP Testing and Reporting on Students with Disabilities and English Language Learners” include (1) ensuring that NAEP is fully representative of SD and ELL students and (2) maximizing student participation in NAEP.

All students sampled to participate in NAEP who are identified as SD and/or ELL are individually evaluated via a “decision tree” with the help of knowledgeable school staff to determine if they can participate in the assessment and if an accommodation is appropriate given their IEP/504 and ELL status. The current operational decision trees have been modified and are being pilot-tested in 50 schools in 2012 at 4th and 8th grade. The main changes to the SD decision tree include (1) general encouragement that the students take the assessment even if the required accommodation(s) is not offered and (2) provision for exclusion only if the student’s IEP calls for them to be tested with an alternative assessment with alternate achievement standards.

The primary change to the ELL decision tree was a provision to exclude students only if they have been enrolled in US schools for less than 1 academic year.

The primary goals of the special pilot are to (1) provide information from the field about language refinements needed to make the tree more user-friendly and (2) provide a comparison sample to gauge the impact of the new decision tree on inclusion rates. Data collection is currently ongoing.
Inclusion on NAEP

WHAT IS NAEP?

- The National Assessment of Educational Progress (NAEP) is the largest continuing and nationally representative assessment of what our nation’s students know and can do in core subjects such as mathematics and reading.

- NAEP is congressionally mandated, and was first administered in 1969 to measure student achievement nationally. Teachers, principals, parents, policymakers, and researchers all use NAEP results to assess progress and develop ways to improve education in the United States.

Students with Disabilities and English Language Learners

To ensure that the National Assessment of Educational Progress (NAEP) reflects the educational progress of all students, students with disabilities and English language learners must be included to the fullest extent possible.

The Importance of Inclusion

The responses of students with disabilities and English language learners on NAEP 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. Such best practices are essential to ensuring an inclusive assessment that yields meaningful NAEP results for students with disabilities and English language learners. 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.

Selection

Students with disabilities and English language learners are selected to participate in NAEP just like any other student. NAEP is administered to a sample of students who represent the student population of the nation as a whole and of individual states and districts participating in the Trial Urban District Assessment (TUDA). 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 chosen.

**Participation**

Many students with disabilities and English language learners are able to participate in the assessment administration alongside their peers. Participating in NAEP allows students with disabilities and English language learners to experience a large-scale assessment administration without high-stakes consequences. Since NAEP does not produce scores for individual students or results for schools, participation in NAEP is not tied to grades or evaluations of students, teachers, or schools.

**Accommodations**

Students with disabilities and English language learners are provided with testing accommodations so that they can demonstrate their content knowledge and skills on NAEP. NAEP offers a comprehensive set of accommodations to increase access to testing for students with disabilities and English language learners. To ensure that appropriate accommodations are determined for individual students, NAEP relies on school staff to make inclusion and accommodations decisions for those students selected for the assessment.

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.

Your NAEP State Coordinator has developed state-specific guidelines for including students with disabilities and English language learners on NAEP and provided them to schools selected for the assessment. These instructions will highlight any differences between NAEP and your state’s policies and provide guidance on how to include and accommodate sampled students.

**For More Information**

Contact your NAEP State Coordinator with specific questions about including students with disabilities and English language learners. Details of the National Assessment Governing Board’s policy on testing and reporting results on students with disabilities and English language learners on NAEP are available at [http://www.nagb.org/publications/inclusion-special-population-naep.pdf](http://www.nagb.org/publications/inclusion-special-population-naep.pdf). Current NAEP accommodations and inclusion policies are posted on the website at [http://nces.ed.gov/nationsreportcard/about/inclusion.asp](http://nces.ed.gov/nationsreportcard/about/inclusion.asp).
January 10, 2012

Chair David Driscoll
National Assessment Governing Board
800 North Capitol Street, NW
Suite 825
Washington, DC 20002

Dear Chair Driscoll:

As the Chair of the Executive Board of the National Assessment Governing Board (NAGB), we know you understand how important it is to include as many students with disabilities (SDs) and English language learners (ELLs) as possible in the NAEP sample. Unfortunately, in recent years, there have been substantial variations in inclusion rates among states that have led us to be concerned about the validity of state-level comparisons to the nation, as well as other states.

On March 6, 2010, NAGB adopted a Policy Statement based on input from panels of experts in the field as well as consideration of the views expressed by a wide range of public comments and in the detailed analyses provided by the National Center for Education Statistics (NCES), the agency responsible for conducting NAEP under the policy guidance of NAGB. As the policy states, the goal is to maximize participation of sampled students in NAEP, reduce variation in inclusion rates for SD and ELL students across states and districts, develop uniform rules for including students in NAEP, and ensure that NAEP is fully representative of SD and ELL students.

The Policy Statement defines specific inclusion goals for NAEP samples at the national, state, and district levels: 95 percent of all students selected for NAEP samples and 85 percent of those in a NAEP sample who are identified as SD or ELL.
In 2011, the percentage of identified SD and ELL students participating in the grade 4 NAEP reading sample for the nation was 77 percent. Similarly, the grade 8 NAEP reading inclusion rate for the nation was 76 percent. Of course, the national percentage is reflective of a large variation in state inclusion rates, with Maryland having the lowest inclusion rates—including only 31 percent of their identified SD students in grade 4, and 30 percent of their identified SD students in grade 8. While the national inclusion rates of identified ELL students met the minimum standards, there still remained much variation at the state level.

For many years, NAEP has been identified as the “gold standard” to which all state-level results should be compared. Because of the importance of NAEP results, it is imperative that NAGB seek ways to ensure all states meet the minimum requirements. To continue reporting results, despite significant variations in state inclusion rates, calls into question the validity of any conclusions drawn from state-level comparisons to the nation or to other states.

As you prepare for the Future of NAEP Summit, as well as any potential Congressional actions, I propose that you consider a policy of only reporting or using state-level results if the minimum standards are met. This would ensure the validity of the reported results for the nation and for the participating states. States not meeting the minimum standards should face funding sanctions.

Thank you for your consideration of this important issue. I look forward to your response.

Sincerely,

Gerard Robinson

GR/KE/sk
January 30, 2012

Gerard Robinson
Commissioner of Education
Florida Department of Education
325 W. Gaines Street
Tallahassee, FL 32399

Dear Commissioner Robinson:

Thank you very much for your letter about the variations in state exclusion rates on the National Assessment of Educational Progress. The concerns you raise are indeed serious ones that the Governing Board has been trying to respond to for some time.

As you note, the Board adopted a new policy in March 2010 on NAEP testing and reporting on students with disabilities (SD) and English-language learners (ELL). Several key features of the policy that deal with public reporting of assessment results were implemented in the reports on the 2011 NAEP assessments in reading and mathematics. The key aspects related to student testing are scheduled for implementation in 2013. These include limiting the grounds on which students may automatically be exempted from NAEP and revising the “decision-trees” for determining exclusions and accommodations.

All of the changes are being made with two goals in mind: increasing participation in the National Assessment and reducing the state-to-state and district-to-district variations in exclusion and accommodation rates, which are still much too large.

At its meeting on March 2-3, the Board will receive a briefing on implementation of the SD and ELL policy from the National Center on Education Statistics (NCES). We will also be discussing your letter. In addition, NCES will provide you shortly with information about studies they conducted to better understand the impact of differing exclusion rates.

I appreciate your interest in assuring that comparisons reported by NAEP are full and fair.

Sincerely,

David P. Driscoll
Chairman
February 7, 2012

Commissioner Gerard Robinson
Florida Department of Education
325 West Gaines Street
Tallahassee, Florida 32399-0400

Dear Commissioner Robinson:

Your letter to the Honorable David Driscoll, Chair of the National Assessment Governing Board, was forwarded to me for response. I share your concern about the variation among states in the inclusion of students with disabilities (SD) and English language learners (ELL) on the National Assessment of Education Progress (NAEP). We have devoted considerable research beginning in the early 1990s, and we continue to develop new procedures to increase the participation of these two groups in our assessments.

In the last 20 years, the percentage of students identified as either disabled or ELL has increased considerably. At the same time, we have made considerable progress in reducing exclusion rates in NAEP. In 1992, for example, 10 percent of all 4th grade students in the NAEP reading assessment were identified as SD or ELL, compared with 22 percent in 2011. During the same period, our exclusion rate decreased from 6 percent to 4 percent of all students. Thus, in 1992, 60 percent of disabled students and 64 percent of ELLs were excluded, compared with 23 percent of disabled students and 11 percent of ELLs in 2011.

While this improvement is encouraging, we have endeavored to do more. One feature of the Governing Board’s new policy on the testing and reporting of SD and ELL students is a more inclusive decision tree (a tool provided to the participating schools to guide officials in deciding whether a sampled SD or ELL student may be excluded from the NAEP assessment). This new decision tree should result in lower exclusion rates in all jurisdictions.

As one example, the new decision tree states that the only SDs eligible for exclusion are those who take an alternate state test with alternate achievement standards. The current decision tree (the one used in 2011) allows for exclusion of any student who has an accommodation on their IEP that NAEP does not allow. Maryland (which you mentioned in your letter) allows the “read aloud” accommodation on their state reading test and NAEP does not; Maryland thus excluded from NAEP many of the students who had this accommodation on their IEP. When the new decision tree is fully implemented, NAEP guidelines will indicate that these students are to be assessed.
The exception would be that those students who take the alternate state assessment with alternate achievement standards could still be excluded. This is expected to be a very small percentage of students (i.e. 1% or less). This very important change was made to help standardize the exclusion decision among jurisdictions participating in NAEP.

While we anticipate continued decline in exclusion rates in NAEP, we are confident that the variations that do remain have little to no impact on the validity of the conclusions drawn from the comparisons between jurisdictions and the nation. This is a pattern we have documented for more than a decade as part of our ongoing investigation into the impact of exclusion rates on NAEP scores. These data show a small to negligible non-consistent pattern between state exclusion rates and state scores on NAEP. These findings indicate that it is not possible to reliably predict NAEP state scores based on state exclusion rates. In any given year, how states score on NAEP is not related to how many students they exclude.

More consistent with the inferences in your memo, however, we have found a moderate relationship between changes in exclusion rates and changes in NAEP scores, but typically only for reading at grade 4. Changes in state exclusion rates over time have been found to be moderately related to changes in NAEP scores. In 2011, we found a significant correlation of .45 for reading at grade 4. When exclusion rates went up over time, a state’s average 4th grade NAEP reading score tended to increase. For other grade/subject combinations, such relationships are typically found to be negligible or nonexistent.

Because of these latter findings, for several years NCES has closely monitored the impact of changes in exclusion rates on NAEP trends using what we call “full population estimates” (FPE; http://nces.ed.gov/nationsreportcard/about/inclusion.asp#research). These FPEs allow us to see how each state’s published NAEP score compares to what the score would have been if the exclusion rate had been zero. In 2011, for 4th grade reading, there were no significant differences between the published NAEP scores and the full population estimates. However, there are 4 jurisdictions in which that difference (one as low as .1) was enough to convert the 2009 to 2011 published NAEP score from a significant gain to a non-significant one using the FPE results.

Collectively, such research provides no technical basis for NCES to withhold reporting state scores if they exceed a certain exclusion rate. Absolute state exclusion rates in any given year are not correlated with NAEP scores, and changes in exclusion rates over time, for any given state, don’t reliably predict state trends on NAEP. (The exception, as noted above, has been 4th grade reading to a moderate extent.) In fact, it is not uncommon for jurisdictions to decrease exclusion while also increasing their absolute scores. While we are unable to account for all of the reasons why a common standard is not technically feasible, our data indicate that states differ in the prevalence of students with disabilities and English language learners, the degree of severity of those disabilities or lack of English fluency, and differences in state policies regarding testing accommodations. In large part, this variation is why we have seen little or no predictability in scores using a common standard. In an NCES study entitled Measuring Status and Change in Exclusion Rates of Students with Disabilities, available at http://nces.ed.gov/nationsreportcard/pubs/studies/2011457.asp, the analysis underscores the differences in the characteristics of the disabled population among states and how those differences should result in differential expectations for exclusion rates for states.
Thank you for taking the time to suggest ways in which we can improve NAEP. We are confident that differences in exclusion rates among states will continue to decline as we implement our new policy in 2013. If you should have any further concerns or questions, please let me know.

Sincerely,

Peggy Carr, PhD
Associate Commissioner of Assessment

Cc: David Driscoll, PhD, Chair of the National Assessment Governing Board
    Cornelia Orr, PhD, Executive Director of the National Assessment Governing Board
    Jack Buckley, PhD, Commissioner of the National Center for Education Statistics
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.
NAEP Inclusion

It is important for NAEP to assess as many students selected to participate as possible. Assessing representative samples of students, including students with disabilities (SD) and English language learners (ELL), helps to ensure that NAEP results accurately reflect the educational performance of all students in the target population and can continue to serve as a meaningful measure of U.S. students’ academic achievement over time.

The National Assessment Governing Board, which sets policy for NAEP, has been exploring ways to ensure that NAEP continues to appropriately include as many students as possible and to do so in a consistent manner for all jurisdictions assessed and reported. In March 2010, the Governing Board adopted a new policy, *NAEP Testing and Reporting on Students with Disabilities and English Language Learners*. This policy was the culmination of work with experts in testing and curriculum, and those who work with exceptional children and students learning to speak English. The policy aims to

- maximize participation of sampled students in NAEP,
- reduce variation in exclusion rates for SD and ELL students across states and districts,
- develop uniform national rules for including students in NAEP, and
- ensure that NAEP is fully representative of SD and ELL students.

The policy defines specific inclusion goals for NAEP samples. At the national, state, and district levels, the goal is to include 95 percent of all students selected for the NAEP samples, and 85 percent of those in the NAEP sample who are identified as SD or ELL.

Students are selected to participate in NAEP based on a sampling procedure designed to yield a sample of students that is representative of students in all schools nationwide and in public schools within each state. First, schools are selected, and then students are sampled from within those schools without regard to disability or English language proficiency. Once students are selected, those previously identified as SD or ELL may be offered accommodations or excluded.

States and jurisdictions vary in their proportions of special-needs students and in their policies on inclusion and the use of accommodations. Despite the increasing identification of SD and ELL students in some states, in particular of ELL students at grade 4, NAEP inclusion rates have generally remained steady or increased since 2003. Only a small number of states included a smaller percentage of students in the 2011 NAEP reading assessments than in 2009. At grade 4, inclusion rates increased by more than 1 percentage point for 28 of 52 jurisdictions and decreased by more than 1 percentage point for only 2 states. At grade 8, the inclusion rates increased by more than 1 percentage point for 19 jurisdictions, and no jurisdictions saw a decline of more than 1 percentage point. This reflects efforts on the part of states and jurisdictions to include all students who can meaningfully participate in the NAEP assessments. The new NAEP inclusion policy is an effort to ensure that this trend continues.

Determining whether each jurisdiction has met the NAEP inclusion goals involves looking at three different inclusion rates—an overall inclusion rate, an inclusion rate for SD students, and an inclusion rate for ELL students. Each inclusion rate is calculated as the percentage of sampled students who were included in the assessment (i.e., were not excluded).

Inclusion rate percentages are estimates because they are based on representative samples of students rather than on the entire population of students. As such, the inclusion rates are associated with a margin of error. The margin of error for each jurisdiction’s inclusion rate was taken into
account when comparing it to the corresponding inclusion goal. For example, if the point estimate of a state’s overall inclusion rate was 93 percent and had a margin of error of plus or minus 3 percentage points, the state was considered to have met the 95 percent inclusion goal because the 95 percent goal falls within the margin of error, which ranges from 90 percent to 96 percent. Refer to the Technical Notes for more details about how the margin of error was used in these calculations.

Forty-one of the states/jurisdictions participating in the 2011 reading assessment met the 95 percent inclusion goal at both grades 4 and 8 (figure 33). See appendix table A-4 for the inclusion rates as a percentage of all students in each state/jurisdiction, and table A-5 for the rates as a percentage of the SD or ELL students.

**Figure 33. States and jurisdictions meeting the 95 percent inclusion rate goal in NAEP reading at grades 4 and 8: 2011**

| State met 95 percent inclusion goal at both grades 4 and 8 in 2011. |
| State met 95 percent inclusion goal at grade 4 but not at grade 8 in 2011. |
| State met 95 percent inclusion goal at grade 8 but not at grade 4 in 2011. |
| State did not meet 95 percent inclusion goal at both grades 4 and 8 in 2011. |

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

**Inclusion Policy**

See the National Assessment Governing Board’s policy on *NAEP Testing and Reporting on Students with Disabilities and English Language Learners* at http://www.nagb.org/policies/PoliciesPDFs/Reporting%20and%20Dissemination/naep_testandreport_studentswithdisabilities.pdf.

**SOURCE:** U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Reading Assessment.
Table A-4. Inclusion rate and confidence interval in NAEP reading for fourth- and eighth-grade public school students, as a percentage of all students, by state/jurisdiction: 2011

<table>
<thead>
<tr>
<th>State/jurisdiction</th>
<th>Grade 4</th>
<th></th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inclusion rate</td>
<td>95% confidence interval</td>
<td></td>
<td></td>
<td>Inclusion rate</td>
<td>95% confidence interval</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
</tr>
<tr>
<td><strong>Nation (public)</strong></td>
<td>96¹</td>
<td>95.9</td>
<td>96.3</td>
<td></td>
<td>97¹</td>
<td>96.4</td>
<td>96.7</td>
</tr>
<tr>
<td>Alabama</td>
<td>98¹</td>
<td>97.0</td>
<td>98.3</td>
<td></td>
<td>98¹</td>
<td>97.1</td>
<td>98.5</td>
</tr>
<tr>
<td>Alaska</td>
<td>98¹</td>
<td>97.4</td>
<td>98.4</td>
<td></td>
<td>98¹</td>
<td>97.6</td>
<td>98.6</td>
</tr>
<tr>
<td>Arizona</td>
<td>99¹</td>
<td>97.9</td>
<td>99.0</td>
<td></td>
<td>99¹</td>
<td>98.3</td>
<td>99.2</td>
</tr>
<tr>
<td>Arkansas</td>
<td>99¹</td>
<td>98.3</td>
<td>99.1</td>
<td></td>
<td>99¹</td>
<td>98.0</td>
<td>98.9</td>
</tr>
<tr>
<td>California</td>
<td>98¹</td>
<td>96.8</td>
<td>98.5</td>
<td></td>
<td>98¹</td>
<td>97.1</td>
<td>98.4</td>
</tr>
<tr>
<td>Colorado</td>
<td>99¹</td>
<td>98.1</td>
<td>99.0</td>
<td></td>
<td>98¹</td>
<td>97.8</td>
<td>98.9</td>
</tr>
<tr>
<td>Connecticut</td>
<td>98¹</td>
<td>96.7</td>
<td>98.5</td>
<td></td>
<td>98¹</td>
<td>96.8</td>
<td>98.4</td>
</tr>
<tr>
<td>Delaware</td>
<td>93</td>
<td>92.1</td>
<td>93.8</td>
<td></td>
<td>95¹</td>
<td>94.0</td>
<td>95.4</td>
</tr>
<tr>
<td>Florida</td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
<td></td>
<td>98¹</td>
<td>97.0</td>
<td>98.2</td>
</tr>
<tr>
<td>Georgia</td>
<td>94</td>
<td>92.4</td>
<td>94.8</td>
<td></td>
<td>96¹</td>
<td>94.7</td>
<td>96.4</td>
</tr>
<tr>
<td>Hawaii</td>
<td>98¹</td>
<td>97.1</td>
<td>98.2</td>
<td></td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
</tr>
<tr>
<td>Idaho</td>
<td>98¹</td>
<td>97.5</td>
<td>98.7</td>
<td></td>
<td>98¹</td>
<td>97.7</td>
<td>98.7</td>
</tr>
<tr>
<td>Illinois</td>
<td>98¹</td>
<td>97.6</td>
<td>98.9</td>
<td></td>
<td>98¹</td>
<td>97.8</td>
<td>98.8</td>
</tr>
<tr>
<td>Indiana</td>
<td>98¹</td>
<td>98.3</td>
<td>99.1</td>
<td></td>
<td>98¹</td>
<td>97.0</td>
<td>98.5</td>
</tr>
<tr>
<td>Iowa</td>
<td>98¹</td>
<td>98.4</td>
<td>99.4</td>
<td></td>
<td>98¹</td>
<td>98.8</td>
<td>99.5</td>
</tr>
<tr>
<td>Kansas</td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
<td></td>
<td>98¹</td>
<td>97.4</td>
<td>98.6</td>
</tr>
<tr>
<td>Kentucky</td>
<td>91</td>
<td>90.2</td>
<td>92.2</td>
<td></td>
<td>93</td>
<td>92.1</td>
<td>93.4</td>
</tr>
<tr>
<td>Louisiana</td>
<td>99¹</td>
<td>98.1</td>
<td>99.1</td>
<td></td>
<td>99¹</td>
<td>98.5</td>
<td>99.3</td>
</tr>
<tr>
<td>Maine</td>
<td>98¹</td>
<td>98.0</td>
<td>98.8</td>
<td></td>
<td>98¹</td>
<td>97.6</td>
<td>98.8</td>
</tr>
<tr>
<td>Maryland</td>
<td>90</td>
<td>88.6</td>
<td>90.6</td>
<td></td>
<td>92</td>
<td>90.5</td>
<td>92.5</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>94¹</td>
<td>93.3</td>
<td>95.2</td>
<td></td>
<td>94</td>
<td>92.5</td>
<td>94.7</td>
</tr>
<tr>
<td>Michigan</td>
<td>96¹</td>
<td>95.5</td>
<td>97.2</td>
<td></td>
<td>95¹</td>
<td>94.2</td>
<td>96.0</td>
</tr>
<tr>
<td>Minnesota</td>
<td>98¹</td>
<td>97.8</td>
<td>98.9</td>
<td></td>
<td>97¹</td>
<td>96.3</td>
<td>97.8</td>
</tr>
<tr>
<td>Mississippi</td>
<td>99¹</td>
<td>98.4</td>
<td>99.3</td>
<td></td>
<td>99¹</td>
<td>98.5</td>
<td>99.3</td>
</tr>
<tr>
<td>Missouri</td>
<td>98¹</td>
<td>97.8</td>
<td>98.8</td>
<td></td>
<td>99¹</td>
<td>98.0</td>
<td>99.0</td>
</tr>
<tr>
<td>Montana</td>
<td>96¹</td>
<td>94.9</td>
<td>96.5</td>
<td></td>
<td>96¹</td>
<td>95.2</td>
<td>96.6</td>
</tr>
<tr>
<td>Nebraska</td>
<td>96¹</td>
<td>94.4</td>
<td>96.7</td>
<td></td>
<td>95¹</td>
<td>94.6</td>
<td>95.9</td>
</tr>
<tr>
<td>Nevada</td>
<td>99¹</td>
<td>98.4</td>
<td>99.2</td>
<td></td>
<td>98¹</td>
<td>97.3</td>
<td>98.6</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>97¹</td>
<td>96.3</td>
<td>97.9</td>
<td></td>
<td>96¹</td>
<td>94.9</td>
<td>96.6</td>
</tr>
<tr>
<td>New Jersey</td>
<td>91</td>
<td>90.8</td>
<td>92.4</td>
<td></td>
<td>93</td>
<td>91.2</td>
<td>94.3</td>
</tr>
<tr>
<td>New Mexico</td>
<td>94¹</td>
<td>92.9</td>
<td>95.4</td>
<td></td>
<td>94</td>
<td>93.6</td>
<td>94.9</td>
</tr>
<tr>
<td>New York</td>
<td>97¹</td>
<td>96.2</td>
<td>98.3</td>
<td></td>
<td>97¹</td>
<td>96.0</td>
<td>97.6</td>
</tr>
<tr>
<td>North Carolina</td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
<td></td>
<td>98¹</td>
<td>97.2</td>
<td>98.5</td>
</tr>
<tr>
<td>North Dakota</td>
<td>94</td>
<td>92.6</td>
<td>94.3</td>
<td></td>
<td>92</td>
<td>91.2</td>
<td>92.9</td>
</tr>
<tr>
<td>Ohio</td>
<td>94¹</td>
<td>92.5</td>
<td>95.6</td>
<td></td>
<td>94¹</td>
<td>93.1</td>
<td>95.2</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>95¹</td>
<td>93.9</td>
<td>96.0</td>
<td></td>
<td>96¹</td>
<td>94.7</td>
<td>96.5</td>
</tr>
<tr>
<td>Oregon</td>
<td>97¹</td>
<td>96.7</td>
<td>97.9</td>
<td></td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>97¹</td>
<td>96.2</td>
<td>97.8</td>
<td></td>
<td>97¹</td>
<td>95.6</td>
<td>97.7</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>98¹</td>
<td>97.3</td>
<td>98.4</td>
<td></td>
<td>99¹</td>
<td>98.4</td>
<td>99.1</td>
</tr>
<tr>
<td>South Carolina</td>
<td>97¹</td>
<td>96.2</td>
<td>98.0</td>
<td></td>
<td>95¹</td>
<td>93.6</td>
<td>95.6</td>
</tr>
<tr>
<td>South Dakota</td>
<td>97¹</td>
<td>96.1</td>
<td>97.4</td>
<td></td>
<td>97¹</td>
<td>96.2</td>
<td>97.3</td>
</tr>
<tr>
<td>Tennessee</td>
<td>93</td>
<td>91.7</td>
<td>94.0</td>
<td></td>
<td>94</td>
<td>92.6</td>
<td>94.6</td>
</tr>
<tr>
<td>Texas</td>
<td>90</td>
<td>88.4</td>
<td>91.5</td>
<td></td>
<td>94</td>
<td>92.7</td>
<td>95.0</td>
</tr>
<tr>
<td>Utah</td>
<td>96¹</td>
<td>94.7</td>
<td>96.8</td>
<td></td>
<td>96¹</td>
<td>95.4</td>
<td>97.0</td>
</tr>
<tr>
<td>Vermont</td>
<td>98¹</td>
<td>96.9</td>
<td>98.2</td>
<td></td>
<td>97¹</td>
<td>96.7</td>
<td>97.7</td>
</tr>
<tr>
<td>Virginia</td>
<td>97¹</td>
<td>96.3</td>
<td>97.9</td>
<td></td>
<td>96¹</td>
<td>95.4</td>
<td>97.2</td>
</tr>
<tr>
<td>Washington</td>
<td>97¹</td>
<td>96.4</td>
<td>97.8</td>
<td></td>
<td>98¹</td>
<td>97.3</td>
<td>98.3</td>
</tr>
<tr>
<td>West Virginia</td>
<td>98¹</td>
<td>97.7</td>
<td>98.7</td>
<td></td>
<td>99¹</td>
<td>98.0</td>
<td>98.9</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>98¹</td>
<td>97.6</td>
<td>98.6</td>
<td></td>
<td>98¹</td>
<td>97.2</td>
<td>98.3</td>
</tr>
<tr>
<td>Wyoming</td>
<td>98¹</td>
<td>97.5</td>
<td>98.5</td>
<td></td>
<td>98¹</td>
<td>97.4</td>
<td>98.5</td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td>District of Columbia</td>
<td>97¹</td>
<td>95.9</td>
<td>96.4</td>
<td></td>
<td>97¹</td>
<td>96.4</td>
</tr>
<tr>
<td>DoDEA²</td>
<td>93</td>
<td>92.4</td>
<td>94.0</td>
<td></td>
<td>97</td>
<td>95.9</td>
<td>97.4</td>
</tr>
</tbody>
</table>

¹ The state/jurisdiction’s inclusion rate is higher than or not significantly different from the National Assessment Governing Board goal of 95 percent.
² Department of Defense Education Activity (overseas and domestic schools).

Table A-5. Inclusion rate and standard error in NAEP reading for fourth- and eighth-grade public school students with disabilities (SD) and English language learners (ELL), as a percentage of identified SD or ELL students, by state/jurisdiction: 2011

<table>
<thead>
<tr>
<th>State/jurisdiction</th>
<th>SD Inclusion rate</th>
<th>SD SE</th>
<th>ELL Inclusion rate</th>
<th>ELL SE</th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nation (public)</td>
<td>77 0.5</td>
<td></td>
<td>89¹ 0.7</td>
<td></td>
<td>76 0.5</td>
<td>86¹ 0.8</td>
</tr>
<tr>
<td>Alabama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>77 3.5</td>
<td></td>
<td>95¹ 3.4</td>
<td></td>
<td>82¹ 2.9</td>
<td>96¹ 1.1</td>
</tr>
<tr>
<td>Alaska</td>
<td>92¹ 1.4</td>
<td></td>
<td>92¹ 1.5</td>
<td></td>
<td>88¹ 1.7</td>
<td>96¹ 1.1</td>
</tr>
<tr>
<td>Arizona</td>
<td>88¹ 2.2</td>
<td></td>
<td>99¹ 0.6</td>
<td></td>
<td>89¹ 2.2</td>
<td>96¹ 1.1</td>
</tr>
<tr>
<td>Arkansas</td>
<td>89¹ 1.5</td>
<td></td>
<td>98¹ 0.9</td>
<td></td>
<td>87¹ 2.1</td>
<td>97¹ 1.6</td>
</tr>
<tr>
<td>California</td>
<td>80¹ 3.3</td>
<td></td>
<td>96¹ 0.8</td>
<td></td>
<td>78 3.3</td>
<td>95¹ 1.1</td>
</tr>
<tr>
<td>Colorado</td>
<td>89¹ 1.8</td>
<td></td>
<td>98¹ 0.7</td>
<td></td>
<td>87¹ 2.1</td>
<td>92¹ 2.2</td>
</tr>
<tr>
<td>Connecticut</td>
<td>88¹ 2.0</td>
<td></td>
<td>84¹ 4.8</td>
<td></td>
<td>87¹ 2.4</td>
<td>77¹ 5.9</td>
</tr>
<tr>
<td>Delaware</td>
<td>60 2.6</td>
<td></td>
<td>63 4.4</td>
<td></td>
<td>67 2.2</td>
<td>91¹ 1.1</td>
</tr>
<tr>
<td>Florida</td>
<td>89¹ 1.5</td>
<td></td>
<td>92¹ 1.5</td>
<td></td>
<td>87¹ 1.9</td>
<td>83¹ 2.8</td>
</tr>
<tr>
<td>Georgia</td>
<td>54 3.4</td>
<td></td>
<td>69 7.3</td>
<td></td>
<td>62 3.1</td>
<td>60 8.9</td>
</tr>
<tr>
<td>Hawaii</td>
<td>87¹ 2.1</td>
<td></td>
<td>89¹ 2.4</td>
<td></td>
<td>93¹ 1.6</td>
<td>84¹ 1.9</td>
</tr>
<tr>
<td>Idaho</td>
<td>84¹ 2.5</td>
<td></td>
<td>94¹ 2.4</td>
<td></td>
<td>82¹ 3.0</td>
<td>87¹ 2.7</td>
</tr>
<tr>
<td>Illinois</td>
<td>91¹ 1.6</td>
<td></td>
<td>92¹ 2.3</td>
<td></td>
<td>90¹ 1.6</td>
<td>91¹ 2.7</td>
</tr>
<tr>
<td>Indiana</td>
<td>93¹ 1.2</td>
<td></td>
<td>98¹ 0.8</td>
<td></td>
<td>86¹ 2.2</td>
<td>90¹ 3.7</td>
</tr>
<tr>
<td>Iowa</td>
<td>93¹ 1.6</td>
<td></td>
<td>98¹ 1.1</td>
<td></td>
<td>95¹ 1.1</td>
<td>99¹ 0.9</td>
</tr>
<tr>
<td>Kansas</td>
<td>87¹ 1.5</td>
<td></td>
<td>94¹ 1.6</td>
<td></td>
<td>84¹ 2.6</td>
<td>98¹ 1.4</td>
</tr>
<tr>
<td>Kentucky</td>
<td>45 2.3</td>
<td></td>
<td>37 5.2</td>
<td></td>
<td>39 2.6</td>
<td>59 6.7</td>
</tr>
<tr>
<td>Louisiana</td>
<td>89¹ 1.9</td>
<td></td>
<td>100¹ †</td>
<td></td>
<td>92¹ 1.9</td>
<td>59 6.7</td>
</tr>
<tr>
<td>Maine</td>
<td>91¹ 1.2</td>
<td></td>
<td>98¹ 1.6</td>
<td></td>
<td>90¹ 1.5</td>
<td>59 6.7</td>
</tr>
<tr>
<td>Maryland</td>
<td>31 2.2</td>
<td></td>
<td>52 4.3</td>
<td></td>
<td>30 3.3</td>
<td>45 6.6</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>71 2.6</td>
<td></td>
<td>82¹ 3.2</td>
<td></td>
<td>69 2.9</td>
<td>70 4.8</td>
</tr>
<tr>
<td>Michigan</td>
<td>75 3.1</td>
<td></td>
<td>93¹ 2.4</td>
<td></td>
<td>63 3.3</td>
<td>79¹ 4.5</td>
</tr>
<tr>
<td>Minnesota</td>
<td>90¹ 1.8</td>
<td></td>
<td>98¹ 0.8</td>
<td></td>
<td>78 2.7</td>
<td>94¹ 2.3</td>
</tr>
<tr>
<td>Mississippi</td>
<td>90¹ 2.2</td>
<td></td>
<td>† †</td>
<td></td>
<td>88¹ 2.2</td>
<td>† †</td>
</tr>
<tr>
<td>Missouri</td>
<td>88¹ 1.7</td>
<td></td>
<td>97¹ 1.7</td>
<td></td>
<td>90¹ 1.7</td>
<td>† †</td>
</tr>
<tr>
<td>Montana</td>
<td>64 3.5</td>
<td></td>
<td>87¹ 4.0</td>
<td></td>
<td>68 2.8</td>
<td>† †</td>
</tr>
<tr>
<td>Nebraska</td>
<td>80 2.0</td>
<td></td>
<td>84¹ 5.0</td>
<td></td>
<td>70 2.2</td>
<td>† †</td>
</tr>
<tr>
<td>Nevada</td>
<td>90¹ 1.8</td>
<td></td>
<td>99¹ 0.3</td>
<td></td>
<td>83¹ 2.4</td>
<td>94¹ 1.8</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>83¹ 2.1</td>
<td></td>
<td>90¹ 3.3</td>
<td></td>
<td>77 2.2</td>
<td>† †</td>
</tr>
<tr>
<td>New Jersey</td>
<td>50 3.9</td>
<td></td>
<td>55 8.8</td>
<td></td>
<td>64 3.7</td>
<td>† †</td>
</tr>
<tr>
<td>New Mexico</td>
<td>72 2.9</td>
<td></td>
<td>82¹ 2.4</td>
<td></td>
<td>66 2.4</td>
<td>80 1.8</td>
</tr>
<tr>
<td>New York</td>
<td>90¹ 2.4</td>
<td></td>
<td>86¹ 2.6</td>
<td></td>
<td>85¹ 2.1</td>
<td>79¹ 4.1</td>
</tr>
<tr>
<td>North Carolina</td>
<td>84¹ 2.0</td>
<td></td>
<td>96¹ 1.1</td>
<td></td>
<td>85¹ 2.2</td>
<td>91¹ 2.7</td>
</tr>
<tr>
<td>North Dakota</td>
<td>58 2.4</td>
<td></td>
<td>64 4.7</td>
<td></td>
<td>48 2.8</td>
<td>40 5.4</td>
</tr>
<tr>
<td>Ohio</td>
<td>59 3.9</td>
<td></td>
<td>83¹ 5.7</td>
<td></td>
<td>62 3.4</td>
<td>73 6.6</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>74 2.9</td>
<td></td>
<td>80¹ 5.3</td>
<td></td>
<td>75 2.7</td>
<td>76¹ 6.5</td>
</tr>
<tr>
<td>Oregon</td>
<td>84¹ 1.8</td>
<td></td>
<td>95¹ 1.2</td>
<td></td>
<td>85¹ 1.9</td>
<td>94¹ 2.0</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>85¹ 1.9</td>
<td></td>
<td>74 4.9</td>
<td></td>
<td>84¹ 2.5</td>
<td>66 10.1</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>88¹ 1.6</td>
<td></td>
<td>91¹ 2.6</td>
<td></td>
<td>95¹ 0.8</td>
<td>88¹ 3.3</td>
</tr>
<tr>
<td>South Carolina</td>
<td>82¹ 3.0</td>
<td></td>
<td>91¹ 2.4</td>
<td></td>
<td>57 3.8</td>
<td>80¹ 5.6</td>
</tr>
<tr>
<td>South Dakota</td>
<td>82 1.6</td>
<td></td>
<td>87¹ 3.1</td>
<td></td>
<td>74 2.5</td>
<td>71 5.1</td>
</tr>
<tr>
<td>Tennessee</td>
<td>50 4.1</td>
<td></td>
<td>82¹ 3.5</td>
<td></td>
<td>47 4.3</td>
<td>† †</td>
</tr>
<tr>
<td>Texas</td>
<td>48 3.8</td>
<td></td>
<td>75 2.7</td>
<td></td>
<td>53 3.6</td>
<td>80¹ 3.7</td>
</tr>
<tr>
<td>Utah</td>
<td>72 3.5</td>
<td></td>
<td>85¹ 3.0</td>
<td></td>
<td>70 3.0</td>
<td>75 4.7</td>
</tr>
<tr>
<td>Vermont</td>
<td>86¹ 1.9</td>
<td></td>
<td>92¹ 3.3</td>
<td></td>
<td>85¹ 1.8</td>
<td>† †</td>
</tr>
<tr>
<td>Virginia</td>
<td>81¹ 2.5</td>
<td></td>
<td>92¹ 1.9</td>
<td></td>
<td>77 3.0</td>
<td>78 3.9</td>
</tr>
<tr>
<td>Washington</td>
<td>82¹ 2.3</td>
<td></td>
<td>94¹ 1.1</td>
<td></td>
<td>86¹ 1.7</td>
<td>88¹ 2.9</td>
</tr>
<tr>
<td>West Virginia</td>
<td>90¹ 1.4</td>
<td></td>
<td>† †</td>
<td></td>
<td>89¹ 1.8</td>
<td>† †</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>88¹ 1.4</td>
<td></td>
<td>95¹ 1.7</td>
<td></td>
<td>86¹ 1.8</td>
<td>92¹ 2.5</td>
</tr>
<tr>
<td>Wyoming</td>
<td>89¹ 1.4</td>
<td></td>
<td>91¹ 2.4</td>
<td></td>
<td>88¹ 1.9</td>
<td>† †</td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>83¹ 2.2</td>
<td></td>
<td>88¹ 2.2</td>
<td></td>
<td>87¹ 1.6</td>
<td>83¹ 2.9</td>
</tr>
<tr>
<td>DoDEA²</td>
<td>55 2.3</td>
<td></td>
<td>69 3.4</td>
<td></td>
<td>72 3.8</td>
<td>75 4.8</td>
</tr>
</tbody>
</table>

¹ Not applicable. Standard error estimate cannot be accurately determined.
² Reporting standards not met. Sample size insufficient to permit a reliable estimate.
¹ The state/jurisdiction’s inclusion rate is higher than or not significantly different from the National Assessment Governing Board goal of 85 percent.
³ Department of Defense Education Activity (overseas and domestic schools).

NOTE: SD includes students identified as having an Individualized Education Program but excludes other students protected under Section 504 of the Rehabilitation Act of 1973. SE = Standard error.

Table A-6. Percentage of fourth- and eighth-grade public school students with disabilities (SD) and English language learners (ELL) identified, excluded, and accommodated in NAEP reading, as a percentage of all students, by state/jurisdiction: 2011

<table>
<thead>
<tr>
<th>State/jurisdiction</th>
<th>Grade 4 SD</th>
<th>ELL</th>
<th>Grade 8 SD</th>
<th>ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall excluded</td>
<td></td>
<td>Identified</td>
<td>Excluded</td>
</tr>
<tr>
<td>Nation (public)</td>
<td>4  13  3  7</td>
<td>11  1  4</td>
<td>3  13  3  8</td>
<td>6  1  2</td>
</tr>
<tr>
<td>Alabama</td>
<td>2  10  2  3</td>
<td>2  #  #</td>
<td>2  10  2  4</td>
<td>2  #  #</td>
</tr>
<tr>
<td>Alaska</td>
<td>2  16  1  12</td>
<td>14  1  10</td>
<td>2  13  1  11</td>
<td>11  1  #</td>
</tr>
<tr>
<td>Arizona</td>
<td>1  12  1  8</td>
<td>12  #  6</td>
<td>1  11  1  8</td>
<td>2  #  1</td>
</tr>
<tr>
<td>Arkansas</td>
<td>1  13  1  9</td>
<td>8  #  5</td>
<td>1  11  1  9</td>
<td>5  #  3</td>
</tr>
<tr>
<td>California</td>
<td>2  10  2  5</td>
<td>32  1  3</td>
<td>2  10  2  5</td>
<td>17  1  3</td>
</tr>
<tr>
<td>Colorado</td>
<td>1  11  1  8</td>
<td>16  #  7</td>
<td>2  10  1  8</td>
<td>7  1  3</td>
</tr>
<tr>
<td>Connecticut</td>
<td>2  14  1  11</td>
<td>6  1  5</td>
<td>2  12  1  10</td>
<td>4  1  3</td>
</tr>
<tr>
<td>Delaware</td>
<td>7  16  6  7</td>
<td>4  1  5</td>
<td>5  14  5  9</td>
<td>2  1  1</td>
</tr>
<tr>
<td>Florida</td>
<td>2  16  2  11</td>
<td>9  1  8</td>
<td>2  14  2  12</td>
<td>5  1  4</td>
</tr>
<tr>
<td>Georgia</td>
<td>6  12  5  4</td>
<td>5  2  2</td>
<td>4  10  4  5</td>
<td>2  1  1</td>
</tr>
<tr>
<td>Hawaii</td>
<td>2  10  1  7</td>
<td>11  1  4</td>
<td>2  11  1  8</td>
<td>9  1  3</td>
</tr>
<tr>
<td>Idaho</td>
<td>2  11  2  6</td>
<td>5  #  2</td>
<td>2  8  1  5</td>
<td>4  #  1</td>
</tr>
<tr>
<td>Illinois</td>
<td>2  14  1  9</td>
<td>8  1  6</td>
<td>2  14  1  11</td>
<td>4  #  2</td>
</tr>
<tr>
<td>Indiana</td>
<td>1  16  1  10</td>
<td>7  #  5</td>
<td>2  14  2  11</td>
<td>3  #  2</td>
</tr>
<tr>
<td>Iowa</td>
<td>1  15  1  12</td>
<td>6  #  4</td>
<td>1  15  1  12</td>
<td>3  #  2</td>
</tr>
<tr>
<td>Kansas</td>
<td>2  14  2  8</td>
<td>11  1  4</td>
<td>2  12  2  8</td>
<td>6  #  1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>9  15  8  3</td>
<td>2  1  1</td>
<td>7  12  7  4</td>
<td>1  1  #</td>
</tr>
<tr>
<td>Louisiana</td>
<td>1  20  1  16</td>
<td>2  #  1</td>
<td>1  14  1  13</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Maine</td>
<td>2  17  2  14</td>
<td>3  #  2</td>
<td>2  18  2  13</td>
<td>2  #  1</td>
</tr>
<tr>
<td>Maryland</td>
<td>10  14  8  4</td>
<td>6  3  3</td>
<td>8  11  7  3</td>
<td>3  2  1</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>6  18  5  12</td>
<td>8  1  1</td>
<td>6  19  5  12</td>
<td>4  1  1</td>
</tr>
<tr>
<td>Michigan</td>
<td>4  13  3  7</td>
<td>9  #  5</td>
<td>5  12  4  6</td>
<td>2  #  1</td>
</tr>
<tr>
<td>Minnesota</td>
<td>2  15  1  8</td>
<td>10  #  3</td>
<td>3  13  3  7</td>
<td>5  #  1</td>
</tr>
<tr>
<td>Mississippi</td>
<td>1  9  1  5</td>
<td>2  #  1</td>
<td>1  7  1  5</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Missouri</td>
<td>2  13  2  8</td>
<td>3  #  2</td>
<td>1  13  1  10</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Montana</td>
<td>4  12  4  5</td>
<td>2  #  1</td>
<td>4  12  4  6</td>
<td>2  #  1</td>
</tr>
<tr>
<td>Nebraska</td>
<td>4  17  3  8</td>
<td>8  1  3</td>
<td>5  14  4  7</td>
<td>3  1  1</td>
</tr>
<tr>
<td>Nevada</td>
<td>1  11  1  7</td>
<td>27  #  13</td>
<td>2  10  2  6</td>
<td>10  1  4</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>3  17  3  13</td>
<td>3  #  2</td>
<td>4  18  4  11</td>
<td>2  1  #</td>
</tr>
<tr>
<td>New Jersey</td>
<td>9  17  8  7</td>
<td>3  1  2</td>
<td>7  17  6  10</td>
<td>2  1  1</td>
</tr>
<tr>
<td>New Mexico</td>
<td>6  13  4  7</td>
<td>17  3  5</td>
<td>6  12  4  5</td>
<td>12  2  2</td>
</tr>
<tr>
<td>New York</td>
<td>3  16  2  13</td>
<td>9  1  8</td>
<td>3  16  2  13</td>
<td>6  1  4</td>
</tr>
<tr>
<td>North Carolina</td>
<td>2  15  2  10</td>
<td>7  #  3</td>
<td>2  14  2  10</td>
<td>5  #  2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>6  15  6  6</td>
<td>3  1  8</td>
<td>4  17  4  7</td>
<td>5  2  1</td>
</tr>
<tr>
<td>Ohio</td>
<td>6  14  5  7</td>
<td>3  1  3</td>
<td>6  15  5  8</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>5  15  4  8</td>
<td>6  1  2</td>
<td>4  16  4  9</td>
<td>3  1  1</td>
</tr>
<tr>
<td>Oregon</td>
<td>3  15  2  8</td>
<td>14  1  5</td>
<td>2  13  2  8</td>
<td>6  #  2</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>3  15  2  10</td>
<td>3  1  2</td>
<td>3  16  2  12</td>
<td>2  1  1</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>2  14  2  11</td>
<td>6  1  2</td>
<td>1  16  1  12</td>
<td>3  #  2</td>
</tr>
<tr>
<td>South Dakota</td>
<td>3  14  2  7</td>
<td>5  #  1</td>
<td>5  11  5  4</td>
<td>5  1  1</td>
</tr>
<tr>
<td>South Dakota 1</td>
<td>3  16  3  6</td>
<td>4  1  2</td>
<td>3  11  3  5</td>
<td>2  1  #</td>
</tr>
<tr>
<td>Tennessee</td>
<td>7  14  7  4</td>
<td>3  1  3</td>
<td>6  12  6  4</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Texas</td>
<td>10  10  6  3</td>
<td>22  5  1</td>
<td>6  11  5  3</td>
<td>9  2  1</td>
</tr>
<tr>
<td>Utah</td>
<td>4  13  4  6</td>
<td>7  1  2</td>
<td>4  10  3  5</td>
<td>5  1  1</td>
</tr>
<tr>
<td>Vermont</td>
<td>2  17  2  13</td>
<td>2  #  1</td>
<td>3  18  2  13</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Virginia</td>
<td>3  13  2  7</td>
<td>7  1  3</td>
<td>4  13  3  7</td>
<td>6  1  1</td>
</tr>
<tr>
<td>Washington</td>
<td>3  14  2  7</td>
<td>11  1  6</td>
<td>2  12  2  8</td>
<td>5  1  2</td>
</tr>
<tr>
<td>West Virginia</td>
<td>2  17  2  8</td>
<td>1  #  1</td>
<td>1  14  1  7</td>
<td>1  #  1</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>2  14  2  11</td>
<td>8  #  6</td>
<td>2  14  2  11</td>
<td>5  #  3</td>
</tr>
<tr>
<td>Wyoming</td>
<td>2  16  2  11</td>
<td>4  #  2</td>
<td>2  13  2  11</td>
<td>2  1  1</td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>3  15  3  12</td>
<td>7  1  3</td>
<td>3  17  2  14</td>
<td>6  1  4</td>
</tr>
<tr>
<td>DoDEA1</td>
<td>7  13  5  5</td>
<td>7  2  2</td>
<td>3  10  3  7</td>
<td>5  1  1</td>
</tr>
</tbody>
</table>

Rounds to zero.

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

NOTE: Students identified as both SD and ELL were counted only once in overall, 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.

Table A-7. Percentage of fourth- and eighth-grade public school students identified as students with disabilities (SD) and/or English language learners (ELL) excluded in NAEP reading, as a percentage of all students, by state/jurisdiction: Various years, 1992–2011

<table>
<thead>
<tr>
<th>State/jurisdiction</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Alaska</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Arizona</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Arkansas</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>California</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Colorado</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>—</td>
<td>3</td>
</tr>
<tr>
<td>Connecticut</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Delaware</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Florida</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Georgia</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Hawaii</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Idaho</td>
<td>4</td>
<td>5</td>
<td>—</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Illinois</td>
<td>—</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Indiana</td>
<td>4</td>
<td>5</td>
<td>—</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Iowa</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Kansas</td>
<td>—</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Kentucky</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Louisiana</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Maine</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Maryland</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Michigan</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Minnesota</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Mississippi</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Missouri</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Montana</td>
<td>—</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Nebraska</td>
<td>4</td>
<td>4</td>
<td>—</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Nevada</td>
<td>—</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>—</td>
<td>4</td>
</tr>
<tr>
<td>New Jersey</td>
<td>6</td>
<td>6</td>
<td>—</td>
<td>—</td>
<td>5</td>
</tr>
<tr>
<td>New Mexico</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>New York</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>North Carolina</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>North Dakota</td>
<td>2</td>
<td>2</td>
<td>—</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Ohio</td>
<td>6</td>
<td>—</td>
<td>—</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Oregon</td>
<td>—</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>4</td>
<td>6</td>
<td>—</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>South Carolina</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>South Dakota</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Tennessee</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Texas</td>
<td>8</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Utah</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Vermont</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Virginia</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Washington</td>
<td>—</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>West Virginia</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Wyoming</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>DoDEA2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

— Not available. The state/jurisdiction did not participate or did not meet the minimum participation guidelines for reporting.
1Accommodations not permitted.
2Department of Defense Education Activity (overseas and domestic schools).

<table>
<thead>
<tr>
<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>Nation (public)</td>
<td>60</td>
<td>44</td>
<td>42</td>
<td>40</td>
<td>34</td>
<td>36</td>
<td>34</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Alabama</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>22</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>Alaska</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>61</td>
<td>43</td>
<td>49</td>
<td>41</td>
<td>45</td>
<td>34</td>
<td>31</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Arkansas</td>
<td>51</td>
<td>51</td>
<td>43</td>
<td>36</td>
<td>37</td>
<td>49</td>
<td>45</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>California</td>
<td>49</td>
<td>49</td>
<td>60</td>
<td>40</td>
<td>25</td>
<td>29</td>
<td>26</td>
<td>27</td>
<td>20</td>
</tr>
<tr>
<td>Colorado</td>
<td>59</td>
<td>52</td>
<td>26</td>
<td></td>
<td>20</td>
<td>24</td>
<td>27</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>Connecticut</td>
<td>39</td>
<td>43</td>
<td>51</td>
<td>29</td>
<td>29</td>
<td>23</td>
<td>18</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>Delaware</td>
<td>49</td>
<td>40</td>
<td>9</td>
<td>63</td>
<td>72</td>
<td>55</td>
<td>49</td>
<td>38</td>
<td>13</td>
</tr>
<tr>
<td>Florida</td>
<td>54</td>
<td>50</td>
<td>34</td>
<td>27</td>
<td>19</td>
<td>25</td>
<td>25</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Georgia</td>
<td>59</td>
<td>49</td>
<td>40</td>
<td>30</td>
<td>24</td>
<td>40</td>
<td>38</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>Hawaii</td>
<td>42</td>
<td>52</td>
<td>35</td>
<td>25</td>
<td>19</td>
<td>24</td>
<td>13</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>Idaho</td>
<td>43</td>
<td>43</td>
<td>31</td>
<td>23</td>
<td>27</td>
<td>25</td>
<td>26</td>
<td>15</td>
<td>29</td>
</tr>
<tr>
<td>Illinois</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>59</td>
<td>45</td>
<td>36</td>
<td>27</td>
<td>25</td>
<td>24</td>
<td>27</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>Iowa</td>
<td>42</td>
<td>41</td>
<td>34</td>
<td>45</td>
<td>45</td>
<td>36</td>
<td>30</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>49</td>
<td>51</td>
<td>58</td>
<td>69</td>
<td>59</td>
<td>56</td>
<td>48</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>Louisiana</td>
<td>56</td>
<td>56</td>
<td>49</td>
<td>59</td>
<td>29</td>
<td>60</td>
<td>21</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Maine</td>
<td>47</td>
<td>60</td>
<td>50</td>
<td>36</td>
<td>37</td>
<td>35</td>
<td>31</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>Maryland</td>
<td>47</td>
<td>48</td>
<td>45</td>
<td>49</td>
<td>46</td>
<td>40</td>
<td>51</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>40</td>
<td>35</td>
<td>23</td>
<td>27</td>
<td>16</td>
<td>35</td>
<td>29</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>Michigan</td>
<td>70</td>
<td>62</td>
<td>60</td>
<td>62</td>
<td>57</td>
<td>48</td>
<td>33</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>Minnesota</td>
<td>47</td>
<td>34</td>
<td>22</td>
<td>28</td>
<td>18</td>
<td>24</td>
<td>15</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Mississippi</td>
<td>77</td>
<td>63</td>
<td>60</td>
<td>62</td>
<td>59</td>
<td>34</td>
<td>23</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Missouri</td>
<td>42</td>
<td>42</td>
<td>45</td>
<td>54</td>
<td>45</td>
<td>46</td>
<td>22</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Montana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td>31</td>
<td>25</td>
<td>26</td>
<td>24</td>
<td>27</td>
<td>29</td>
<td>23</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>37</td>
<td>38</td>
<td>22</td>
<td></td>
<td>20</td>
<td>18</td>
<td>21</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>New Jersey</td>
<td>51</td>
<td>47</td>
<td>26</td>
<td>29</td>
<td>39</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>60</td>
<td>43</td>
<td>51</td>
<td>42</td>
<td>24</td>
<td>43</td>
<td>50</td>
<td>36</td>
<td>28</td>
</tr>
<tr>
<td>New York</td>
<td>55</td>
<td>58</td>
<td>42</td>
<td>45</td>
<td>37</td>
<td>28</td>
<td>29</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>North Carolina</td>
<td>34</td>
<td>36</td>
<td>45</td>
<td>62</td>
<td>38</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>North Dakota</td>
<td>20</td>
<td>20</td>
<td>31</td>
<td>25</td>
<td>36</td>
<td>56</td>
<td>44</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Ohio</td>
<td>63</td>
<td>63</td>
<td>62</td>
<td>60</td>
<td>50</td>
<td>44</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>69</td>
<td>65</td>
<td>26</td>
<td>30</td>
<td>29</td>
<td>43</td>
<td>45</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>44</td>
<td>54</td>
<td>30</td>
<td>23</td>
<td>27</td>
<td>30</td>
<td>19</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>37</td>
<td>35</td>
<td>32</td>
<td>18</td>
<td>17</td>
<td>13</td>
<td>18</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>South Carolina</td>
<td>54</td>
<td>51</td>
<td>48</td>
<td>28</td>
<td>45</td>
<td>39</td>
<td>27</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>South Dakota</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>40</td>
<td>48</td>
<td>24</td>
<td>28</td>
<td>62</td>
<td>64</td>
<td>62</td>
<td>49</td>
<td>39</td>
</tr>
<tr>
<td>Texas</td>
<td>55</td>
<td>54</td>
<td>47</td>
<td>56</td>
<td>52</td>
<td>49</td>
<td>52</td>
<td>49</td>
<td>53</td>
</tr>
<tr>
<td>Utah</td>
<td>42</td>
<td>43</td>
<td>43</td>
<td>37</td>
<td>25</td>
<td>28</td>
<td>39</td>
<td>42</td>
<td>28</td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>53</td>
<td>52</td>
<td>42</td>
<td>59</td>
<td>57</td>
<td>64</td>
<td>47</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>Washington</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>60</td>
<td>57</td>
<td>70</td>
<td>65</td>
<td>61</td>
<td>31</td>
<td>10</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>63</td>
<td>61</td>
<td>53</td>
<td>43</td>
<td>32</td>
<td>31</td>
<td>31</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Wyoming</td>
<td>37</td>
<td>37</td>
<td>24</td>
<td>17</td>
<td>11</td>
<td>10</td>
<td>23</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>82</td>
<td>79</td>
<td>59</td>
<td>50</td>
<td>39</td>
<td>44</td>
<td>74</td>
<td>68</td>
<td>17</td>
</tr>
<tr>
<td>DoDEA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Accommodations not permitted.
2 DoDEA = Department of Defense Education Activity (overseas and domestic schools).

NOTE: SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973.

Table A-9. Percentage of fourth- and eighth-grade public school English language learners (ELL) excluded in NAEP reading, as a percentage of identified ELL students, by state/jurisdiction: Various years, 1992-2011

<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>Nation (public)</td>
<td>64</td>
<td>39</td>
<td>38</td>
<td>26</td>
<td>24</td>
<td>22</td>
<td>20</td>
<td>16</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaska</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>54</td>
<td>40</td>
<td>45</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>70</td>
<td>43</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>72</td>
<td>73</td>
<td>75</td>
<td>43</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>52</td>
<td>40</td>
<td>27</td>
<td>31</td>
<td>23</td>
<td>30</td>
<td>29</td>
<td>21</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>49</td>
<td>25</td>
<td>32</td>
<td>27</td>
<td>29</td>
<td>14</td>
<td>17</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>58</td>
<td>70</td>
<td>43</td>
<td>50</td>
<td>36</td>
<td>31</td>
<td>30</td>
<td>18</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>64</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>53</td>
<td>45</td>
<td>24</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>34</td>
<td>27</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>37</td>
<td>46</td>
<td>56</td>
<td>51</td>
<td>35</td>
<td>27</td>
<td>19</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>59</td>
<td>35</td>
<td>40</td>
<td>37</td>
<td>26</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>38</td>
<td>40</td>
<td>52</td>
<td>34</td>
<td>33</td>
<td>40</td>
<td>34</td>
<td>28</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>Utah</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Not available. The state/jurisdiction did not participate or did not meet the minimum participation guidelines for reporting.
- Rounds to zero.
- Reporting standards not met. Sample size insufficient to permit a reliable estimate.
- Accommodations not permitted.
- Department of Defense Education Activity (overseas and domestic schools).

Table A-10. Percentage of fourth-grade public school students identified as students with disabilities (SD) and/or English language learners (ELL)
excluded and assessed in NAEP reading, as a percentage of identified SD and/or ELL students, by state/jurisdiction: 2011
Percentage of identified SD and/or ELL students
SD

SD and/or ELL
Assessed

State/jurisdiction
Excluded
Nation (public)
17
Alabama
19
Alaska
7
Arizona
7
Arkansas
6
California
6
Colorado
6
Connecticut
12
Delaware
37
Florida
9
Georgia
39
Hawaii
11
Idaho
12
Illinois
8
Indiana
5
Iowa
5
Kansas
9
Kentucky
54
Louisiana
6
Maine
8
Maryland
54
Massachusetts
23
Michigan
21
Minnesota
7
Mississippi
9
Missouri
10
Montana
31
Nebraska
19
Nevada
3
New Hampshire
15
New Jersey
46
New Mexico
21
New York
11
North Carolina
10
North Dakota
38
Ohio
33
Oklahoma
24
Oregon
9
Pennsylvania
16
Rhode Island
11
South Carolina
15
South Dakota
16
Tennessee
42
Texas
33
Utah
22
Vermont
12
Virginia
15
Washington
13
West Virginia
9
Wisconsin
9
Wyoming
10
Other jurisdictions
District of Columbia
15
DoDEA1
36

Total
83
81
93
93
94
94
94
88
63
91
61
89
88
92
95
95
91
46
94
92
46
77
79
93
91
90
69
81
97
85
54
79
89
90
62
67
76
91
84
89
85
84
58
67
78
88
85
87
91
91
90

Without
accommodations
40
51
21
34
25
78
40
9
23
13
26
31
37
28
30
17
43
23
16
20
10
27
35
48
41
33
31
35
46
13
8
42
4
32
24
12
31
45
21
24
46
44
17
57
36
17
35
35
46
15
23

85
64

8
28

ELL

Assessed
With
accommodations Excluded
43
23
30
23
72
8
59
12
68
9
16
20
54
11
79
10
40
38
77
11
35
44
57
13
51
15
64
9
65
7
78
6
48
13
23
53
78
7
73
9
35
59
50
27
43
25
45
10
50
10
57
12
38
35
47
20
50
10
72
15
45
47
37
28
85
10
57
14
37
40
55
38
45
26
46
16
63
15
65
11
39
18
39
18
41
49
10
53
41
28
71
13
51
18
52
18
44
10
76
12
66
11
77
36

17
42

Total
77
77
92
88
91
80
89
90
62
89
56
87
85
91
93
94
87
47
93
91
41
73
75
90
90
88
65
80
90
85
53
72
90
86
60
62
74
84
85
89
82
82
51
47
72
87
82
82
90
88
89

Without
accommodations
21
46
17
22
17
30
15
9
18
18
23
10
28
29
29
15
28
24
13
13
11
8
26
34
38
29
23
30
30
10
9
22
5
20
18
12
25
29
22
8
34
42
19
20
27
14
26
29
46
14
19

83
58

4
19

Assessed
With
accommodations Excluded
56
11
31
5
75
8
66
1
73
2
51
4
75
2
81
16
44
37
71
8
33
31
77
11
57
6
62
8
64
2
79
2
59
6
23
63
80
#
78
2
30
48
65
18
49
7
56
2
52
‡
59
3
42
13
50
16
61
1
74
10
44
45
49
18
85
14
66
4
41
36
49
17
49
20
55
5
63
26
81
9
48
9
40
13
32
18
27
25
45
15
74
8
56
8
53
6
44
‡
75
5
70
9
80
39

12
31

Total
89
95
92
99
98
96
98
84
63
92
69
89
94
92
98
98
94
37
100
98
52
82
93
98
‡
97
87
84
99
90
55
82
86
96
64
83
80
95
74
91
91
87
82
75
85
92
92
94
‡
95
91

Without
accommodations
58
67
23
43
37
86
54
9
40
3
33
47
55
23
29
22
59
12
42
52
7
65
70
67
‡
48
66
43
50
29
5
51
2
53
49
9
45
57
13
58
73
49
10
71
51
38
46
39
‡
15
43

With
accommodations
31
28
69
55
61
11
44
75
23
89
36
42
40
69
69
76
35
25
58
46
44
17
23
31
‡
49
21
42
49
61
50
31
84
43
15
75
35
39
60
33
18
38
72
4
35
54
46
54
‡
80
48

88
69

16
43

72
26

# Rounds to zero.
‡ Reporting standards not met. Sample size insufficient to permit a reliable estimate.
1
Department of Defense Education Activity (overseas and domestic schools).
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. Detail may not sum to totals because of rounding.
119 of Educational Progress (NAEP), 2011 Reading Assessment.

READING 2011

81


Table A-11. Percentage of eighth-grade public school students identified as students with disabilities (SD) and/or English language learners (ELL) excluded and assessed in NAEP reading, as a percentage of identified SD and/or ELL students, by state/jurisdiction: 2011

<table>
<thead>
<tr>
<th>State/jurisdiction</th>
<th>SD and/or ELL</th>
<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></td>
<td>Excluded</td>
<td>Total</td>
<td>Assessed</td>
<td>Without</td>
<td>With</td>
<td>Excluded</td>
<td>Total</td>
<td>Assessed</td>
<td>Without</td>
<td>With</td>
<td>Excluded</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>accommodations</td>
<td>accommodations</td>
<td></td>
<td></td>
<td></td>
<td>accommodations</td>
<td>accommodations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nation (public)</td>
<td>20</td>
<td>80</td>
<td>29</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>18</td>
<td>82</td>
<td>50</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaska</td>
<td>9</td>
<td>91</td>
<td>18</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>10</td>
<td>90</td>
<td>19</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>9</td>
<td>91</td>
<td>18</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>9</td>
<td>91</td>
<td>64</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>10</td>
<td>90</td>
<td>28</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>14</td>
<td>86</td>
<td>10</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>33</td>
<td>67</td>
<td>12</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>12</td>
<td>88</td>
<td>7</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>37</td>
<td>63</td>
<td>10</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>11</td>
<td>89</td>
<td>38</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td>15</td>
<td>85</td>
<td>32</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>10</td>
<td>90</td>
<td>20</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>12</td>
<td>88</td>
<td>14</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>4</td>
<td>96</td>
<td>15</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>11</td>
<td>89</td>
<td>41</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>56</td>
<td>44</td>
<td>11</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>6</td>
<td>94</td>
<td>7</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>9</td>
<td>91</td>
<td>21</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>61</td>
<td>39</td>
<td>10</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>29</td>
<td>71</td>
<td>15</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>33</td>
<td>67</td>
<td>21</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>17</td>
<td>83</td>
<td>39</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td>12</td>
<td>88</td>
<td>21</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td>10</td>
<td>90</td>
<td>13</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td>30</td>
<td>70</td>
<td>21</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td>29</td>
<td>71</td>
<td>26</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>11</td>
<td>89</td>
<td>40</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>21</td>
<td>79</td>
<td>21</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>37</td>
<td>63</td>
<td>8</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>26</td>
<td>74</td>
<td>45</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>15</td>
<td>85</td>
<td>2</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>12</td>
<td>88</td>
<td>22</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td>50</td>
<td>50</td>
<td>14</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>36</td>
<td>64</td>
<td>10</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>24</td>
<td>76</td>
<td>24</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>12</td>
<td>88</td>
<td>37</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>18</td>
<td>82</td>
<td>9</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>6</td>
<td>94</td>
<td>23</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>35</td>
<td>65</td>
<td>38</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>26</td>
<td>74</td>
<td>29</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>49</td>
<td>51</td>
<td>16</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>33</td>
<td>67</td>
<td>51</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>26</td>
<td>74</td>
<td>31</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>14</td>
<td>86</td>
<td>20</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>20</td>
<td>80</td>
<td>36</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>13</td>
<td>87</td>
<td>28</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>10</td>
<td>90</td>
<td>38</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>12</td>
<td>88</td>
<td>13</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>14</td>
<td>86</td>
<td>12</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other jurisdictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>13</td>
<td>87</td>
<td>9</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DoDEA†</td>
<td>23</td>
<td>77</td>
<td>22</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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. Detail may not sum to totals because of rounding.

† Reporting standards not met. Sample size insufficient to permit a reliable estimate.

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

Florida education Commissioner Gerard Robinson has asked the National Assessment Governing Board to consider setting standards for the numbers of students with disabilities and English-language learners that states exclude from taking national assessments in reading and math. In a letter to David Driscoll, the chairman of NAGB, which sets policy for the National Assessment of Educational Progress, Mr. Robinson wrote that differences in inclusion rates among states raise concerns about state-level comparisons of the test results.

The numbers of 4th and 8th grade students who took NAEP and were identified as having disabilities or being English-language learners have risen for more than a decade, since NAEP first allowed students to use accommodations on the tests. Recent NAEP results showed Florida's reading and math gains have stalled after years of steep increases.

On the most recent NAEP administration, 40 states plus the District of Columbia met the goal of including 95 percent of all 4th and 8th graders in the reading assessment samples.

*The Tampa Bay Times* reported that Cornelia Orr, the executive director of NAGB and a former testing honcho in Florida, said that while the percentages of excluded students in some states may appear large, the raw numbers of students are small, so it doesn't make a large difference in the overall scores.

**By Nirvi Shah**

Vol. 31, Issue 18, Page 4
Following a push to make "the nation's report card" better reflect the academic performance of all children in America's schools, most states boosted the numbers of students with disabilities and English-language learners who participated in the 2011 reading and math tests that are part of the National Assessment of Educational Progress.

But many states still have far to go to reach the inclusion targets set for them last year by federal policymakers.

Overall, the numbers of 4th and 8th grade students who took NAEP and were identified as having disabilities or being English-language learners rose in 2011, continuing a longer-term trend that began more than a decade ago when NAEP first allowed students to use accommodations, such as additional time, when taking the exams.

But to further drive up inclusion rates for students with disabilities and English-learners, especially in states and school districts that continue to exclude large numbers of such students, the National Assessment Governing Board—the independent body that makes policy for NAEP—set inclusion targets last year for states to meet in the 2011 reading and math exams.

Forty states, plus the District of Columbia, met the goal of including 95 percent of all students in the original testing sample for the reading assessment for grades 4 and 8. On the math assessment, Oklahoma was the only state to fall short of the 95 percent inclusion goal for both grades, while Maryland did so in grade 8.

"We think states and school districts are taking this seriously and we think there will continue to be improvement," said David P. Driscoll, NAGB's chairman. "We want to see an end to these exclusion rates bouncing all over the place."

Far fewer states measured up when it came to meeting NAGB's goal of including 85 percent of the students identified as having disabilities or being English-learners targeted for testing. And some continue to exclude large numbers of students from the exams.

To ensure that "the nation's report card" is a nationally representative sample of students, the federal testing program selects potential test-takers from a state's entire population at each grade level. State and district educators then may exclude students whose language difficulties or disabilities make test-taking impractical. A state's exclusion rate is the percentage of students from these categorical groups that are removed from testing.

**Left Out**

The rates at which students with disabilities and English-language learners are being excluded from National Assessment of Education Progress in reading and mathematics have declined overall since at least 2003. The goal is for the tests to be administered to 95 percent of all students and 85 percent of special education students or English-language learners who are in the initial sample of students targeted for testing.

**Jumps in Exclusions**

In Kentucky, which has one of the highest exclusion rates, 63 percent of the 4th graders who were identified as English-learners in the state's testing sample were excluded from the reading assessment.
in 2011, up 20 percentage points from two years ago. Among the targeted test-takers in Oklahoma, 60 percent of the 8th graders identified as having disabilities were excluded from the math exam, as were 51 percent of 4th graders. The Oklahoma numbers represent double-digit increases over the exclusion rates in 2009.

Keeping those percentages consistent across states is important because scores could rise and fall with changes in the population of test-takers with learning challenges. The stakes will get higher for states in 2013 when additional rules kick in to further limit who can be excluded from the tests.

Only students with significant cognitive disabilities who take alternate state assessments may be excluded, said Grady Wilburn, an associate research scientist at the National Center for Education Statistics, which oversees the design and administration of NAEP. And for English-learners, school districts will have to include all such students who have been in a U.S. school more than one year. That rule is technically in effect already, but school districts have found ways around it, he said.

NAGB board member Andrew Porter, the dean of the University of Pennsylvania's graduate school of education, zeroed in on exclusions when he joined the board six years ago.

"What we want is fair and valid comparisons over time," he said. "It was obviously an area we needed to look at. I wouldn't have raised it as an issue if I didn't think it was clouding some of the results.

"There's also just the credibility factor: If you have a lot of exclusions, it raises questions," he said. "We want NAEP to be the gold standard."

Even before NAGB approved its policy last year to minimize exclusions from NAEP, the objective had long been to include more students with disabilities and English-learners. That had mostly been done by allowing for a range of testing accommodations for students who needed them. For example, an English learner can take the math NAEP using a bilingual test booklet.

That's why even as the number of English-learners has grown markedly, the percentage of such students participating in NAEP has also increased, said Arnold Goldstein, the director for design, analysis, and reporting at the NCES.

"Allowing for those accommodations has really been a major vehicle to getting those students participating in the test," Mr. Goldstein said.

There are still a few accommodations that NAEP doesn't allow, Mr. Goldstein said. Students can't have someone read aloud to them during the NAEP in reading, for one. Another is giving a test over multiple days, both of which some states allow on their assessments.

But in some states the NAEP accommodations don't appear yet to have had a major effect on inclusion rates for English-learners. Kentucky is one.

The state has roughly 15,500 English-learners in its public schools, out of a total enrollment of 645,000, said Lisa Y. Gross, a spokeswoman for the Kentucky education department.

Ms. Gross said the department had not yet analyzed the 2011 NAEP exclusion rates for English-learners to understand why the percentage had jumped from 43 percent in 4th grade reading to 63 percent. In contrast, the rate dropped in 8th grade reading from 68 percent in 2009 to 41 percent this year. Ms. Gross said one possible explanation for the 4th grade spike is growth in the English-learner population since newcomers don't take NAEP.

Kentucky also allows readers for all parts of its state reading exams. Any student who has received that accommodation on the state test may have automatically been excluded from taking NAEP, Ms. Gross
said. But that accommodation window will shrink when the state's new **assessment and accountability system** takes effect in the spring, she said.

**Opportunity to Improve**

Laura Kaloi, the public-policy director for the New York-based **National Center for Learning Disabilities**, wondered if the lack of dramatic progress in national NAEP scores in 2011—performance improved in math and 8th grade reading but was flat at the 4th grade level—is in part because more special education students were included in NAEP this year.

"It shines a light on why students with disabilities weren't being included in NAEP," she said. In a nutshell, schools and districts fear, as they do with other tests, that special education students will bring down scores. Instead, she said, educators should view expanding their testing populations as an opportunity to improve instruction.

The District of Columbia, which had a detailed plan for including more students with disabilities in NAEP, excluded 17 percent of 4th graders and 12 percent of 8th graders in reading, compared with a 68 percent exclusion rate in 2009. But reading scores were almost unchanged compared with 2009 and better than prior years.

"By removing barriers, D.C. was bringing in some capable students who could show their skills and abilities," said Mr. Goldstein of the NCES.

Vol. 31, Issue 12, Pages 1,14-15
Although there has been a push to make "the nation's report card"—the National Assessment of Educational Progress—better reflect the academic performance of all children in America's schools, the effort hasn't gone far enough, Florida Education Commissioner Gerard Robinson wrote recently.

As colleague Lesli A. Maxwell and I noted late last year, overall, the numbers of 4th and 8th grade students who took NAEP and were identified as having disabilities or being English-language learners rose in 2011, continuing a long-term trend that began more than a decade ago when NAEP first allowed students to use accommodations, such as additional time, when taking the exams.

As the Tampa Bay Times wrote, last week's letter from Commissioner Robinson comes two months after NAEP results showed Florida's reading and math gains have stalled after years of steep increases. In addition, last week's Quality Counts report showed Florida tumbling from the sixth-ranked state to 11th place among states, with NAEP scores playing a role in that drop.

Maryland, which Robinson pointed out as having among the lowest inclusion rates—only 31 percent of their identified students with disabilities in 4th grade and 30 percent in 8th grade—was the top ranked state in Quality Counts.

On the most recent NAEP administration, 40 states, plus the District of Columbia, met the goal of including 95 percent of all students in the original testing sample for the reading assessment for grades 4 and 8. In math, Oklahoma was the only state to fall short of the 95 percent inclusion goal for both grades, while Maryland did so in grade 8.

The Times reported that Cornelia Orr, the executive director of NAGB and the former testing honcho in Florida, said that while the percentages of excluded students in some states may appear large, the raw numbers of students are small, so it doesn't make a large difference in the overall scores.

Robinson told NAGB that they should consider a policy of only reporting or using state-level results if the minimum standards of inclusion are met. "This would ensure the validity of the reported results for the nation and for the participating states. States not meeting the minimum standards should face funding sanctions."
Florida education commissioner raises concerns about state comparisons with NAEP test

When it comes to comparing academic success across states, there is no better tool than the National Assessment of Educational Progress, better known as NAEP and often called the nation’s report card. Fast-rising NAEP scores are one of the main reasons why Florida’s ed reforms generated buzz over the past decade.

But now Florida’s education commissioner is raising concerns about the validity of state comparisons with NAEP, given big difference from state to state in the percentages of potentially struggling students that are excluded from taking it.

This week - and the timing can’t be overlooked - Commissioner Gerard Robinson fired off a letter to David Driscoll, chair of the National Assessment Governing Board that oversees NAEP, and proposed that the board not report NAEP data for states that do not test high percentages of students with disabilities and students learning English. “Because of the importance of NAEP results, it is imperative that NAGB seek ways to ensure all states meet the minimum requirements,” Robinson wrote in the letter, dated Tuesday and attached below. “To continue reporting results, despite significant variations in state inclusion rates, calls into question the validity of any conclusions drawn from state-level comparisons to the nation or to other states.”

This is worrisome all get out. But interesting on several levels.

The letter comes two months after the most recent NAEP results showed Florida’s reading and math gainst have largely stalled after a decade of steep increases. It was also sent on the eve of the latest Education Week Quality Counts report, which showed Florida falling from No. 6 to No. 11 among states. Those stagnant NAEP scores are one of two reasons why Florida tumbled.

So, is Florida’s concern about NAEP “inclusion rates” sour grapes? Or a valid attempt to find out if the highly regarded national test is really giving us apples-to-apples comparisons? Well ...

The rates, again, do vary wildly. And some of the states ahead of Florida in the rankings do exclude significantly higher numbers of SD and ELL students.

We put together a quick spreadsheet, attached below, that shows the inclusion rates for Ed Week’s Top 10 states, for both groups of kids, for all four NAEP tests at issue.

The inclusion goal set by the NAGB is for 85 percent of SD and ELL students selected in the NAEP sample to take the test. Florida is one of the few top-ranked states that clearly meets the goal in every case or nearly every case. Maryland, by contrast - the No. 1 state, according to Ed Week - misses it in nearly every case and in most cases isn’t even close. Other states fall short a majority of the time, too.

How much does it matter? Frankly, we don’t have the statistical chops to know and we haven’t had time to talk to people who do.

After Kathleen Shanahan, the state Board of Education chair, raised the issue with The Gradebook yesterday (in response to the the latest Ed Week rankings), we spoke briefly with Cornelia Orr, the executive director of NAGB and the former top testing official in Florida.

Orr said while the percentages of excluded students in some states may appear large, the raw numbers of students are small, so it doesn’t make a large difference in the overall scores. But could it make a small difference? Enough to swing a state’s average score a point? Enough to affect a state’s rank a notch?

The NAGB has been dealing with this issue for some time. It required the reporting of inclusion rates in the NAEP reports for the first time last year. (You can find the inclusion rates for all states in the 2011 reading report here and the 2011 math report here.)

In his letter, Robinson proposed that the NAGB only report NAEP results for states that meet the minimum standards. He also suggested funding sanctions for states that don’t.

(Image from palmbeachpost.com)
<table>
<thead>
<tr>
<th>States</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4th grade</td>
<td>8th grade</td>
<td>4th grade</td>
<td>8th grade</td>
<td>4th grade</td>
<td>8th grade</td>
<td>4th grade</td>
<td>8th grade</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>ELL</td>
<td>SD</td>
<td>ELL</td>
<td>SD</td>
<td>ELL</td>
<td>SD</td>
<td>ELL</td>
</tr>
<tr>
<td>Maryland</td>
<td>31</td>
<td>52</td>
<td>30</td>
<td>45</td>
<td>57</td>
<td>86</td>
<td>43</td>
<td>74</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>71</td>
<td>82</td>
<td>69</td>
<td>70</td>
<td>84</td>
<td>89</td>
<td>80</td>
<td>78</td>
</tr>
<tr>
<td>New York</td>
<td>90</td>
<td>86</td>
<td>85</td>
<td>79</td>
<td>94</td>
<td>94</td>
<td>93</td>
<td>94</td>
</tr>
<tr>
<td>Virginia</td>
<td>81</td>
<td>92</td>
<td>77</td>
<td>78</td>
<td>84</td>
<td>95</td>
<td>81</td>
<td>87</td>
</tr>
<tr>
<td>Arkansas</td>
<td>89</td>
<td>98</td>
<td>87</td>
<td>97</td>
<td>92</td>
<td>98</td>
<td>88</td>
<td>96</td>
</tr>
<tr>
<td>New Jersey</td>
<td>50</td>
<td>55</td>
<td>64</td>
<td>*</td>
<td>81</td>
<td>89</td>
<td>75</td>
<td>96</td>
</tr>
<tr>
<td>Georgia</td>
<td>54</td>
<td>69</td>
<td>62</td>
<td>60</td>
<td>87</td>
<td>95</td>
<td>74</td>
<td>92</td>
</tr>
<tr>
<td>Vermont</td>
<td>86</td>
<td>92</td>
<td>85</td>
<td>*</td>
<td>90</td>
<td>*</td>
<td>93</td>
<td>*</td>
</tr>
<tr>
<td>West Virginia</td>
<td>90</td>
<td>*</td>
<td>89</td>
<td>*</td>
<td>91</td>
<td>*</td>
<td>89</td>
<td>*</td>
</tr>
<tr>
<td>Ohio</td>
<td>59</td>
<td>83</td>
<td>62</td>
<td>73</td>
<td>84</td>
<td>94</td>
<td>65</td>
<td>96</td>
</tr>
<tr>
<td>FLORIDA</td>
<td>89</td>
<td>92</td>
<td>87</td>
<td>83</td>
<td>91</td>
<td>96</td>
<td>88</td>
<td>95</td>
</tr>
<tr>
<td>Nation</td>
<td>77</td>
<td>89</td>
<td>76</td>
<td>86</td>
<td>84</td>
<td>96</td>
<td>80</td>
<td>93</td>
</tr>
</tbody>
</table>

Source: 2011 NAEP reports
* sample size insufficient to permit reliable estimate
PISA – Programme for International Student Assessment

Introduction

In this session, Mr. Andreas Schleicher representing the OECD Programme for International Student Assessment (PISA) will make a presentation to the Board about PISA. During his presentation, Mr. Schleicher will explain the distinctions of PISA that make it different from or similar to the National Assessment of Educational Progress. Of particular interest to the Board is the use of ancillary and background information on students, teachers, and schools in preparing assessment reports which resonate with the public and policy makers. Also of interest to the Board are the processes used by OECD to determine the content included in the assessment and the proficiency levels for reporting scores.

This tab includes an overview of PISA (follows immediately), Mr. Schleicher’s bio as Attachment A and the PISA 2009 Results: Executive Summary as Attachment B.

Overview of PISA

The OECD Programme for International Student Assessment (PISA) is a survey of 15-year-olds in the principal industrialised countries. Every three years, PISA assesses how far students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society. PISA focuses on young people’s ability to use their knowledge and skills to meet real-life challenges. This orientation reflects a change in the goals and objectives of curricula themselves, which are increasingly concerned with what students can do with what they learn at school and not merely with whether they have mastered specific curricular content. PISA’s unique features include its:

- Policy orientation, which highlights differences in performance patterns and identifies features common to high performing students, schools and education systems by linking data on learning outcomes with data on student characteristics and other key factors that shape learning in and outside of school.
- Innovative concept of “literacy”, which refers both to students’ capacity to apply knowledge and skills in key subject areas and to their ability to analyse, reason and communicate effectively as they pose, interpret and solve problems in a variety of situations.
- Relevance to lifelong learning, which goes beyond assessing students’ competencies in school subjects by asking them to report on their motivation to learn, their beliefs about themselves and their learning strategies.
- Regularity, which enables countries to monitor their progress in meeting key learning objectives.
- Breadth of geographical coverage and collaborative nature, which, in PISA 2009, encompasses the 34 OECD member countries and 41 partner countries and economies.

To learn more about PISA and download publications and data, please visit: www.pisa.oecd.org
To learn more about the OECD, please visit www.oecd.org
Andreas Schleicher provides strategic oversight over OECD’s work on the development and utilisation of skills and their social and economic outcomes. This includes the Programme for International Student Assessment (PISA), the OECD Survey of Adult Skills (PIAAC), the OECD Teaching and Learning International Survey (TALIS) and the development and analysis of benchmarks on the performance of education systems (INES). Before joining the OECD, he was Director for Analysis at the International Association for Educational Achievement (IEA). He studied Physics in Germany and received a degree in Mathematics and Statistics in Australia. He is the recipient of numerous honours and awards, including the “Theodor Heuss” prize, awarded in the name of the first president of the Federal Republic of Germany for “exemplary democratic engagement”. He holds an honorary Professorship at the University of Heidelberg.
PISA 2009 Results: Executive Summary
This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

Please cite this publication as:
OECD (2010), PISA 2009 Results: Executive Summary

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Photo credits:
Getty Images © Ariel Skelley
Getty Images © Geostock
Getty Images © Jack Hollingsworth
Stocklib Image Bank © Yuri Arcurs

Corrigenda to OECD publications may be found on line at: www.oecd.org/publishing/corrigenda.
PISA™, OECD/PISA™ and the PISA logo are trademarks of the Organisation for Economic Co-operation and Development (OECD). All use of OECD trademarks is prohibited without written permission from the OECD.
© OECD 2010

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d’exploitation du droit de copie (CFC) at contact@cfcopies.com.
THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA)

PISA focuses on young people's ability to use their knowledge and skills to meet real-life challenges. This orientation reflects a change in the goals and objectives of curricula themselves, which are increasingly concerned with what students can do with what they learn at school and not merely with whether they have mastered specific curricular content. PISA's unique features include its:

- Policy orientation, which highlights differences in performance patterns and identifies features common to high-performing students, schools and education systems by linking data on learning outcomes with data on student characteristics and other key factors that shape learning in and outside of school.
- Innovative concept of “literacy”, which refers both to students’ capacity to apply knowledge and skills in key subject areas and to their ability to analyse, reason and communicate effectively as they pose, interpret and solve problems in a variety of situations.
- Relevance to lifelong learning, which goes beyond assessing students’ competencies in school subjects by asking them to report on their motivation to learn, their beliefs about themselves and their learning strategies.
- Regularity, which enables countries to monitor their progress in meeting key learning objectives.
- Breadth of geographical coverage and collaborative nature, which, in PISA 2009, encompasses the 34 OECD member countries and 41 partner countries and economies.

To learn more about PISA and to download our publications and data, please visit our website: www.pisa.oecd.org

To learn more about the OECD, please visit www.oecd.org
Executive Summary

WHAT STUDENTS KNOW AND CAN DO: STUDENT PERFORMANCE IN READING, MATHEMATICS AND SCIENCE ................................................................. 6

OVERCOMING SOCIAL BACKGROUND: EQUITY IN LEARNING OPPORTUNITIES AND OUTCOMES .......... 9

LEARNING TO LEARN: STUDENT ENGAGEMENT, STRATEGIES AND PRACTICES ...................................................... 12

WHAT MAKES A SCHOOL SUCCESSFUL? RESOURCES, POLICIES AND PRACTICES .................................................. 15

LEARNING TRENDS: CHANGES IN STUDENT PERFORMANCE SINCE 2000 ................................................................. 19
WHAT STUDENTS KNOW AND CAN DO: STUDENT PERFORMANCE IN READING, MATHEMATICS AND SCIENCE

PISA's conception of reading literacy encompasses the range of situations in which people read, the different ways written texts are presented, and the variety of ways that readers approach and use texts, from the functional and finite, such as finding a particular piece of practical information, to the deep and far-reaching, such as understanding other ways of doing, thinking and being. Research shows that these kinds of reading literacy skills are more reliable predictors of economic and social well-being than the number of years spent in school or in post-formal education.

Korea and Finland are the highest performing OECD countries, with mean scores of 539 and 536 points, respectively. However, the partner economy Shanghai-China outperforms them by a significant margin, with a mean score of 556. Top-performing countries or economies in reading literacy include Hong Kong-China (with a mean score of 533), Singapore (526), Canada (524), New Zealand (521), Japan (520) and Australia (515). The Netherlands (508), Belgium (506), Norway (503), Estonia (501), Switzerland (501), Poland (500), Iceland (500) and Liechtenstein (499) also perform above the OECD mean score of 494, while the United States, Sweden, Germany, Ireland, France, Denmark, the United Kingdom, Hungary, Portugal, and partner economy Chinese Taipei have scores close to the OECD mean.

The lowest performing OECD country, Mexico, has an average score of 425. This means that the gap between the highest and lowest performing OECD countries is 114 points – the equivalent of more than two school years. And the gap between the highest and lowest performing partner country or economy is even larger, with 242 score points – or more than six years of formal schooling – separating the mean performance of Shanghai-China and Kyrgyzstan (314).

Differences between countries represent, however, only a fraction of overall variation in student performance. Addressing the educational needs of such diverse populations and narrowing the gaps in student performance that have been observed remains a formidable challenge for all countries.

In 18 participating countries, including Mexico, Chile and Turkey, the highest reading proficiency level achieved by most students was the baseline Level 2. Level 2 is considered a baseline level of proficiency, at which students begin to demonstrate the reading skills that will enable them to participate effectively and productively in life. Students who do not reach Level 2 have difficulties locating basic information that meets several conditions, making comparisons or contrasts around a single feature, working out what a well-defined part of a text means when the information is not prominent, or making connections between the text and outside knowledge by drawing on personal experience and attitudes. The proportion of 15-year-olds in this situation varies widely across countries, from fewer than one in 10 in four countries and economies to the majority of students in 10 countries. Even in the average OECD country, where nearly one student in five does not reach Level 2, tackling such low performance remains a major challenge.

At the other end of the proficiency spectrum, an average of 7.6% of students attain Level 5, and in Singapore, New Zealand and Shanghai-China the percentage is more than twice the OECD average. However, for some countries, developing even a small corps of high-performing students remains an aspiration: in 16 countries, fewer than 1% of students reach Level 5. Students at this level are able to retrieve information requiring the reader to locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. They can critically evaluate information and build hypotheses drawing on specialised knowledge, develop a full and detailed understanding of a text whose content or form is unfamiliar, and understand concepts that are contrary to expectations.

Results from the PISA 2009 assessment show that nurturing high performance and tackling low performance need not be mutually exclusive. The countries with the very highest overall reading performance in PISA 2009, Finland and Korea, as well as the partner economies Hong Kong-China and Shanghai-China, also have among the lowest variation in student scores. Equally importantly, Korea has been able to raise its already-high reading performance even further, by more than doubling the percentage of students reaching Level 5 or higher since 2000.

Korea, with a country mean of 546 score points, performed highest among OECD countries in the PISA 2009 mathematics assessment. The partner countries and economies Shanghai-China, Singapore and Hong Kong-China rank first, second and third, respectively.

In the PISA 2009 mathematics assessment, the OECD countries Finland, Switzerland, Japan, Canada, the Netherlands, New Zealand, Belgium, Australia, Germany, Estonia, Iceland, Denmark, Slovenia and the partner countries and economies Chinese Taipei, Liechtenstein and Macao-China also perform significantly above the OECD average in mathematics.
Shanghai-China, Finland, Hong Kong-China and Singapore are the four highest performers in the PISA 2009 science assessment.

In science, New Zealand, Canada, Estonia, Australia, the Netherlands, Germany, Switzerland, the United Kingdom, Slovenia, Poland, Ireland and Belgium as well as the partner country and economies Chinese Taipei, Liechtenstein and Macao-China also perform significantly above the OECD average.

Some 14.6% of students in Shanghai-China and 12.3% of students in Singapore attain the highest levels of proficiency in all three assessment subjects.

High-level skills are critical for innovation and, as such, are key to economic growth and social development. On average, across OECD countries, 16.3% of students are top performers in at least one of the subject areas of science, mathematics or reading. However, only 4.1% of 15-year-old students are top performers in all three assessment subject areas.

Girls outperform boys in reading skills in every participating country.

Throughout much of the 20th century, concern about gender differences in education focused on girls’ underachievement. More recently, however, the scrutiny has shifted to boys’ underachievement in reading. In the PISA 2009 reading assessment, girls outperform boys in every participating country by an average, among OECD countries, of 39 PISA score points – equivalent to more than half a proficiency level or one year of schooling.

On average across OECD countries, boys outperform girls in mathematics by 12 score points while gender differences in science performance tend to be small, both in absolute terms and when compared with the large gender gap in reading performance and the more moderate gender gap in mathematics. The ranks of top-performing students are filled nearly equally with girls and boys. On average across OECD countries, 4.4% of girls and 3.8% of boys are top performers in all three subjects, and 15.6% of girls and 17.0% of boys are top performers in at least one subject area. While the gender gap among top-performing students is small in science (1% of girls and 1.5% of boys), it is significant in reading (2.8% of girls and 0.5% of boys) and in mathematics (3.4% of girls and 6.6% of boys).

Countries of similar prosperity can produce very different educational results.

The balance of proficiency in some of the richer countries in PISA looks very different from that of some of the poorer countries. In reading, for example, the ten countries in which the majority of students are at Level 1 or below, all in poorer parts of the world, contrast starkly in profile with the 34 OECD countries, where on average a majority attains at least Level 3. However, the fact that the best-performing country or economy in the 2009 assessment is Shanghai-China, with a GDP per capita well below the OECD average, underlines that low national income is not incompatible with strong educational performance. Korea, which is the best-performing OECD country, also has a GDP per capita below the OECD average. Indeed, while there is a correlation between GDP per capita and educational performance, this predicts only 6% of the differences in average student performance across countries.

The other 94% of differences reflect the fact that two countries of similar prosperity can produce very different educational results. Results also vary when substituting spending per student, relative poverty or the share of students with an immigrant background for GDP per capita.

The following table summarises the key data of this volume. For each country, it shows the average score of 15-year-olds in reading, mathematics and science as well as on the subscales that were used to measure reading skills in greater detail. Cells shaded in light blue indicate values above the OECD average. Cells shaded in medium blue indicate values below the OECD average. Cells shaded in dark blue indicate values that are not statistically different from the OECD average.
## What Students Know and Can Do: Student Performance in Reading, Mathematics and Science

### Executive Summary

#### Attachment B: PISA 2009 Results

**Figure 1. Comparing Countries’ and Economies’ Performance**

<table>
<thead>
<tr>
<th>On the overall reading scale</th>
<th>On the reading subscales</th>
<th>On the mathematics scale</th>
<th>On the science scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azerbaijan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD average</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Reading Scale

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>524</td>
</tr>
<tr>
<td>Croatia</td>
<td>521</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>520</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>515</td>
</tr>
<tr>
<td>Slovenia</td>
<td>514</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>512</td>
</tr>
<tr>
<td>Italy</td>
<td>512</td>
</tr>
<tr>
<td>Latvia</td>
<td>503</td>
</tr>
<tr>
<td>Lithuania</td>
<td>499</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>498</td>
</tr>
<tr>
<td>Netherlands</td>
<td>497</td>
</tr>
<tr>
<td>Norway</td>
<td>497</td>
</tr>
<tr>
<td>Poland</td>
<td>495</td>
</tr>
<tr>
<td>Portugal</td>
<td>494</td>
</tr>
<tr>
<td>Russia</td>
<td>493</td>
</tr>
<tr>
<td>Slovenia</td>
<td>493</td>
</tr>
<tr>
<td>Spain</td>
<td>493</td>
</tr>
<tr>
<td>Hungary</td>
<td>493</td>
</tr>
<tr>
<td>Italy</td>
<td>492</td>
</tr>
<tr>
<td>Latvia</td>
<td>491</td>
</tr>
<tr>
<td>Lithuania</td>
<td>490</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>490</td>
</tr>
<tr>
<td>Switzerland</td>
<td>490</td>
</tr>
<tr>
<td>Greece</td>
<td>489</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>489</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>489</td>
</tr>
<tr>
<td>Croatia</td>
<td>488</td>
</tr>
<tr>
<td>Israel</td>
<td>487</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>487</td>
</tr>
<tr>
<td>Austria</td>
<td>487</td>
</tr>
<tr>
<td>Lithuania</td>
<td>487</td>
</tr>
<tr>
<td>Turkey</td>
<td>487</td>
</tr>
<tr>
<td>United States</td>
<td>485</td>
</tr>
<tr>
<td>Vietnam</td>
<td>485</td>
</tr>
<tr>
<td>Korea</td>
<td>485</td>
</tr>
<tr>
<td>Korea average</td>
<td>485</td>
</tr>
</tbody>
</table>

**Source:** OECD, PISA 2009 Database: http://dx.doi.org/10.1787/888932343342

© OECD 2009

PISA 2009 RESULTS: EXECUTIVE SUMMARY
OVERCOMING SOCIAL BACKGROUND: EQUITY IN LEARNING OPPORTUNITIES AND OUTCOMES

The best performing school systems manage to provide high-quality education to all students.

Canada, Finland, Japan, Korea and the partner economies Hong Kong-China and Shanghai-China all perform well above the OECD mean performance and students tend to perform well regardless of their own background or the school they attend. They not only have large proportions of students performing at the highest levels of reading proficiency, but also relatively few students at the lower proficiency levels.

Disadvantaged students may have access to more teachers, but not necessarily to the best teachers.

With the exception of Turkey, Slovenia, Israel and the United States, where socio-economically disadvantaged schools also tend to be deprived in terms of basic resources, such as larger student-staff ratios, OECD countries place at least an equal, if not a larger, number of teachers into socio-economically disadvantaged schools as they do in advantaged schools. But despite this fact, disadvantaged schools still report great difficulties in attracting qualified teachers. In other words, in disadvantaged schools, quantity of resources does not necessarily translate into quality of resources since, in general, more advantaged students attend schools that have a higher proportion of full-time teachers who have an advanced university degree. Findings from PISA suggest that, in terms of teacher resources, many students face the double liability of coming from a disadvantaged background and attending a school with lower quality resources. Many countries also show a strong relationship between the socio-economic background of students and their success at school and, in some of these countries, these disparities are magnified by large variations in the schools’ socio-economic backgrounds, that is, in the backgrounds of the students’ peers.

Home background influences educational success, and schooling often appears to reinforce its effects. Although poor performance in school does not automatically follow from a disadvantaged socio-economic background, the socio-economic background of students and schools does appear to have a powerful influence on performance.

Socio-economic disadvantage has many facets and cannot be ameliorated by education policy alone, much less in the short term. The educational attainment of parents can only gradually improve, and average family wealth depends on the long-term economic development of a country and on a culture that promotes individual savings. However, even if socio-economic background itself is hard to change, PISA shows that some countries succeed in reducing its impact on learning outcomes.

While most of the students who perform poorly in PISA are from socio-economically disadvantaged backgrounds, some peers from similar backgrounds excel in PISA, demonstrating that overcoming socio-economic barriers to achievement is possible. Resilient students come from the bottom quarter of the distribution of socio-economic background in their country and score in the top quarter among students from all countries with similar socio-economic background. In Finland, Japan, Turkey, Canada and Portugal and the partner country Singapore, between 39% and 48% of disadvantaged students are resilient. In Korea and the partner economy Macao-China, 50% and 56% of disadvantaged students can be considered resilient, and this percentage is 72% and 76% in partner economies Hong Kong-China and Shanghai-China, respectively.

Across OECD countries, a student from a more socio-economically advantaged background (among the top one seventh) outperforms a student from an average background by 38 score points, or about one year’s worth of education, in reading. In New Zealand, France, the partner country Bulgaria and the partner economy Dubai (UAE), the gap between advantaged and disadvantaged students is more than 50 score points. On average across OECD countries, 14% of the differences in student reading performance within each country is associated with differences in students’ socio-economic background. In Hungary and the partner countries Peru, Bulgaria and Uruguay, more than 20% of the differences in student performance is associated with differences in background.

Regardless of their own socio-economic background, students attending schools with a socio-economically advantaged intake tend to perform better than those attending schools with more disadvantaged peers.

In the majority of OECD countries, the effect of the school’s economic, social and cultural status on students’ performance far outweighs the effects of the individual student’s socio-economic background. And the magnitude of the differences is striking. In Japan, the Czech Republic, Germany, Belgium and Israel and the partner countries Trinidad and Tobago and Liechtenstein, the performance gap between two students with similar socio-economic backgrounds, one of whom attends a school with an average socio-economic background and the other attends a school with an advantaged socio-economic background (among the top 16% in the country), is equivalent to more than 30 score points, on average, or more than a year’s worth of education.
Across OECD countries, first-generation students – those who were born outside the country of assessment and who also have foreign-born parents – score, on average, 52 score points below students without an immigrant background

In New Zealand, Canada and Switzerland, 20% to 25% of students are from an immigrant background while the proportions are even higher in Liechtenstein (30%), Hong Kong-China (39%), Luxembourg (40%) and Qatar (46%). In Macao-China and Dubai (UAE), that percentage is at least 70%. There is no positive association between the size of the immigrant student population and average performance at the country or economy level, and there is also no relationship between the proportion of students with an immigrant background and the performance gaps between native and immigrant students. These findings contradict the assumption that high levels of immigration will inevitably lower the mean performance of school systems.

Students in urban schools perform better than students in other schools, even after accounting for differences in socio-economic background.

In Turkey, the Slovak Republic, Chile, Mexico and Italy, as well as the partner countries Peru, Tunisia, Albania, Argentina and Romania, the performance gap between students in urban schools and those in rural schools is more than 45 score points after accounting for differences in socio-economic background. This is more than one year of education across OECD countries. That gap is 80 score points or more – or two years of schooling – in Hungary and in the partner countries Bulgaria, Kyrgyzstan and Panama. However, this pattern is not observed in Belgium, Finland, Germany, Greece, Iceland, Ireland, Israel, the Netherlands, Poland, Sweden, the United Kingdom and the United States.

On average across the OECD, 17% of students come from single-parent families and they score five score points lower than students from other types of families after accounting for socio-economic background.

Among OECD countries, the gap is particularly large in the United States where, after accounting for socio-economic background, the performance difference between students from single-parent families and those from other types of families stands at 23 score points. In Ireland, Poland and Mexico, the gap is 13 score points and in Belgium, Japan and Luxembourg it is 10 score points, double the average among OECD countries. Among partner countries and economies, students from single-parent families score 10 points lower than peers from other types of families after accounting for socio-economic background.

Parents’ engagement with their children’s reading life has a positive impact on their children’s reading performance. Students whose parents reported that they had read a book with their child “every day or almost every day” or “once or twice a week” during the first year of primary school performed higher in PISA 2009 than students whose parents reported that they had done this “never or almost never” or “once or twice a month”. On average across the 14 countries that had collected information on this question, the difference is 25 score points, but it ranges from 4 score points in the partner country Lithuania to 63 score points in New Zealand. Also, 15-year-olds whose parents discuss political or social issues once a week or more score 28 score points higher than those whose parents do not, or who talk about these issues less often. The performance advantage was largest in Italy, at 42 score points, and smallest in the partner economy Macao-China, and it is observed across all countries.

The following table summarises key data. For each country, it shows the average score of 15-year-olds in reading and seven equity measures from PISA: i) and ii) two measures focusing on those who achieve the baseline level of proficiency in PISA; the proportion of boys and girls who score below Level 2; iii) a measure of those who overcome socio-economic disadvantaged and do best given their weak prospects, the proportion of resilient students; iv) and v) two measures of the relationship between student background and performance: the percentage of variation in student performance explained by the student’s socio-economic background and the slope of the socio-economic gradient, the average gap in performance between students from different socio-economic backgrounds; and vi) and vii) two measures of equality in the distribution of educational resources, namely the quality and quantity of teachers. For the first five measures, cells shaded in light blue indicate values of quality or equity below the OECD average. Cells shaded in medium blue indicate values of equity below the OECD average. Cells shaded in dark blue indicate values that are not statistically different from the OECD average. In the last two columns, cells shaded in light blue indicate that disadvantaged schools are more likely to have more or better resources. Cells shaded in medium blue that advantaged schools are more likely to have more or better resources. Cells shaded in dark blue indicate values where disadvantaged and advantaged schools are equally likely to have more or better resources. In these two last columns, estimates in bold indicate that they are statistically different from the OECD average.
## SUMMARY OF PISA MEASURES OF EDUCATIONAL EQUITY

The table below shows the percentage of students in Level 2 or below, the percentage of girls in the top 10% of the student/teacher ratio, and the correlation between socio-economic background and reading scores.

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage of boys in Level 2</th>
<th>Percentage of girls in Level 2</th>
<th>Percentage of students in Level 2</th>
<th>Percentage of students in the top 10% of the student/teacher ratio</th>
<th>Correlation between socio-economic background and reading scores</th>
<th>Correlation between socio-economic background and reading scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD average</td>
<td>25</td>
<td>13</td>
<td>8</td>
<td>4.0</td>
<td>0.15</td>
<td>-0.15</td>
</tr>
<tr>
<td>Korea</td>
<td>539</td>
<td>9</td>
<td>2</td>
<td>14</td>
<td>-0.03</td>
<td>0.30</td>
</tr>
<tr>
<td>Finland</td>
<td>536</td>
<td>13</td>
<td>3</td>
<td>11</td>
<td>-0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Canada</td>
<td>524</td>
<td>14</td>
<td>6</td>
<td>10</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>New Zealand</td>
<td>436</td>
<td>23</td>
<td>3</td>
<td>14</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Japan</td>
<td>520</td>
<td>19</td>
<td>8</td>
<td>11</td>
<td>0.20</td>
<td>0.38</td>
</tr>
<tr>
<td>Australia</td>
<td>513</td>
<td>20</td>
<td>9</td>
<td>8</td>
<td>0.02</td>
<td>-0.09</td>
</tr>
<tr>
<td>Netherlands</td>
<td>508</td>
<td>21</td>
<td>9</td>
<td>8</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>Belgium</td>
<td>506</td>
<td>21</td>
<td>14</td>
<td>9</td>
<td>0.58</td>
<td>0.66</td>
</tr>
<tr>
<td>Norway</td>
<td>503</td>
<td>21</td>
<td>8</td>
<td>6</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>Estonia</td>
<td>505</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>0.00</td>
<td>0.43</td>
</tr>
<tr>
<td>Switzerland</td>
<td>503</td>
<td>22</td>
<td>11</td>
<td>8</td>
<td>0.24</td>
<td>0.06</td>
</tr>
<tr>
<td>Poland</td>
<td>500</td>
<td>23</td>
<td>7</td>
<td>9</td>
<td>-0.03</td>
<td>0.81</td>
</tr>
<tr>
<td>Iceland</td>
<td>500</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td>United States</td>
<td>500</td>
<td>21</td>
<td>14</td>
<td>7</td>
<td>0.10</td>
<td>-0.17</td>
</tr>
<tr>
<td>Sweden</td>
<td>497</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>-0.04</td>
<td>0.12</td>
</tr>
<tr>
<td>Germany</td>
<td>497</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>-0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>Ireland</td>
<td>496</td>
<td>23</td>
<td>11</td>
<td>7</td>
<td>-0.08</td>
<td>0.49</td>
</tr>
<tr>
<td>France</td>
<td>496</td>
<td>28</td>
<td>14</td>
<td>8</td>
<td>0.15</td>
<td>0.11</td>
</tr>
<tr>
<td>Denmark</td>
<td>495</td>
<td>19</td>
<td>11</td>
<td>6</td>
<td>0.16</td>
<td>0.27</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>494</td>
<td>23</td>
<td>14</td>
<td>6</td>
<td>-0.03</td>
<td>-0.10</td>
</tr>
<tr>
<td>Hungary</td>
<td>494</td>
<td>24</td>
<td>11</td>
<td>10</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Portugal</td>
<td>489</td>
<td>25</td>
<td>11</td>
<td>10</td>
<td>0.04</td>
<td>0.39</td>
</tr>
<tr>
<td>Italy</td>
<td>486</td>
<td>29</td>
<td>13</td>
<td>8</td>
<td>0.13</td>
<td>0.50</td>
</tr>
<tr>
<td>Slovenia</td>
<td>483</td>
<td>31</td>
<td>11</td>
<td>6</td>
<td>0.55</td>
<td>-0.22</td>
</tr>
<tr>
<td>Greece</td>
<td>483</td>
<td>30</td>
<td>13</td>
<td>7</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>Spain</td>
<td>483</td>
<td>24</td>
<td>15</td>
<td>9</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>478</td>
<td>33</td>
<td>14</td>
<td>5</td>
<td>0.37</td>
<td>-0.14</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>477</td>
<td>32</td>
<td>13</td>
<td>5</td>
<td>-0.21</td>
<td>-0.00</td>
</tr>
<tr>
<td>Israel</td>
<td>474</td>
<td>34</td>
<td>19</td>
<td>6</td>
<td>0.20</td>
<td>-0.20</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>474</td>
<td>33</td>
<td>19</td>
<td>6</td>
<td>0.39</td>
<td>0.20</td>
</tr>
<tr>
<td>Austria</td>
<td>470</td>
<td>35</td>
<td>20</td>
<td>5</td>
<td>0.64</td>
<td>-0.07</td>
</tr>
<tr>
<td>Turkey</td>
<td>464</td>
<td>33</td>
<td>15</td>
<td>10</td>
<td>0.04</td>
<td>-0.26</td>
</tr>
<tr>
<td>China</td>
<td>449</td>
<td>36</td>
<td>15</td>
<td>10</td>
<td>0.25</td>
<td>-0.03</td>
</tr>
<tr>
<td>Mexico</td>
<td>423</td>
<td>46</td>
<td>34</td>
<td>7</td>
<td>-0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Shanghai-China</td>
<td>356</td>
<td>7</td>
<td>2</td>
<td>19</td>
<td>0.32</td>
<td>-0.13</td>
</tr>
<tr>
<td>Hong Kong-China</td>
<td>533</td>
<td>11</td>
<td>5</td>
<td>18</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Singapore</td>
<td>526</td>
<td>12</td>
<td>9</td>
<td>14</td>
<td>0.22</td>
<td>-0.14</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>498</td>
<td>21</td>
<td>9</td>
<td>9</td>
<td>0.57</td>
<td>0.70</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>495</td>
<td>22</td>
<td>10</td>
<td>10</td>
<td>0.29</td>
<td>-0.07</td>
</tr>
<tr>
<td>Malaysia</td>
<td>487</td>
<td>12</td>
<td>9</td>
<td>2</td>
<td>-0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Latvia</td>
<td>484</td>
<td>27</td>
<td>9</td>
<td>8</td>
<td>0.19</td>
<td>0.38</td>
</tr>
<tr>
<td>Croatia</td>
<td>476</td>
<td>31</td>
<td>13</td>
<td>7</td>
<td>0.28</td>
<td>0.32</td>
</tr>
<tr>
<td>Lithuania</td>
<td>468</td>
<td>35</td>
<td>13</td>
<td>5</td>
<td>0.19</td>
<td>0.36</td>
</tr>
<tr>
<td>Dubai (UAE)</td>
<td>459</td>
<td>41</td>
<td>21</td>
<td>3</td>
<td>-0.01</td>
<td>-0.27</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>459</td>
<td>36</td>
<td>19</td>
<td>5</td>
<td>0.21</td>
<td>0.29</td>
</tr>
<tr>
<td>Serbia</td>
<td>442</td>
<td>43</td>
<td>23</td>
<td>4</td>
<td>0.06</td>
<td>-0.01</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>429</td>
<td>52</td>
<td>29</td>
<td>2</td>
<td>0.17</td>
<td>-0.21</td>
</tr>
<tr>
<td>Hungary</td>
<td>426</td>
<td>33</td>
<td>14</td>
<td>4</td>
<td>0.18</td>
<td>-0.21</td>
</tr>
<tr>
<td>Romania</td>
<td>324</td>
<td>51</td>
<td>30</td>
<td>2</td>
<td>0.11</td>
<td>-0.02</td>
</tr>
<tr>
<td>Thailand</td>
<td>423</td>
<td>55</td>
<td>33</td>
<td>7</td>
<td>0.16</td>
<td>-0.02</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>354</td>
<td>34</td>
<td>10</td>
<td>17</td>
<td>0.56</td>
<td>0.38</td>
</tr>
<tr>
<td>Colombia</td>
<td>413</td>
<td>50</td>
<td>45</td>
<td>6</td>
<td>-0.08</td>
<td>-0.14</td>
</tr>
<tr>
<td>Brazil</td>
<td>412</td>
<td>50</td>
<td>43</td>
<td>6</td>
<td>0.03</td>
<td>-0.20</td>
</tr>
<tr>
<td>Montenegro</td>
<td>408</td>
<td>61</td>
<td>37</td>
<td>3</td>
<td>0.38</td>
<td>-0.02</td>
</tr>
<tr>
<td>Jordan</td>
<td>405</td>
<td>62</td>
<td>34</td>
<td>3</td>
<td>-0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Singapore</td>
<td>406</td>
<td>64</td>
<td>34</td>
<td>2</td>
<td>0.20</td>
<td>0.36</td>
</tr>
<tr>
<td>Indonesia</td>
<td>402</td>
<td>65</td>
<td>42</td>
<td>6</td>
<td>0.16</td>
<td>-0.16</td>
</tr>
<tr>
<td>Argentina</td>
<td>398</td>
<td>59</td>
<td>45</td>
<td>3</td>
<td>0.22</td>
<td>-0.02</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>399</td>
<td>50</td>
<td>45</td>
<td>3</td>
<td>0.24</td>
<td>0.34</td>
</tr>
<tr>
<td>Albania</td>
<td>385</td>
<td>69</td>
<td>44</td>
<td>3</td>
<td>0.38</td>
<td>0.15</td>
</tr>
<tr>
<td>Qatar</td>
<td>372</td>
<td>72</td>
<td>54</td>
<td>1</td>
<td>0.38</td>
<td>0.15</td>
</tr>
<tr>
<td>Panama</td>
<td>372</td>
<td>72</td>
<td>54</td>
<td>1</td>
<td>0.38</td>
<td>0.15</td>
</tr>
<tr>
<td>Peru</td>
<td>370</td>
<td>70</td>
<td>60</td>
<td>1</td>
<td>0.48</td>
<td>-0.02</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>362</td>
<td>70</td>
<td>68</td>
<td>1</td>
<td>0.44</td>
<td>0.73</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>314</td>
<td>80</td>
<td>78</td>
<td>1</td>
<td>0.35</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Counties are ranked in descending order of the mean score in reading, separately for OECD and partner countries and economies.

Source: OECD, PISA 2009 Database, Tables 1.2.2, 1.2.3, 1.2.2.3, 1.3.1 and 1.3.3.
LEARNING TO LEARN: STUDENT ENGAGEMENT, STRATEGIES AND PRACTICES

PISA results show that mastering strategies that assist learning, such as methods to remember and understand or summarise texts and reading widely, are essential if students are to become proficient readers. Practicing reading by reading for enjoyment is most closely associated with better outcomes when it is accompanied by high levels of critical thinking and strategic learning. Across OECD countries, students who have low levels of awareness about which strategies are most effective for understanding, remembering and summarising information are less proficient readers than those who have high levels of awareness about these strategies, regardless of their reading habits.

In all countries, students who enjoy reading the most perform significantly better than students who enjoy reading the least. There has been considerable debate as to what type of reading may be most effective in fostering reading skills and improving reading performance. The results from PISA suggest that, although students who read fiction are more likely to achieve high scores, it is students who read a wide variety of material who perform particularly well in reading. Compared with not reading for enjoyment at all, reading fiction for enjoyment appears to be positively associated with higher scores in the PISA 2009 reading assessment, while reading comic books is associated with little improvement in reading proficiency in some countries, and with lower overall reading performance in other countries. Also, students who are extensively engaged in online reading activities, such as reading e-mails, chatting on line, reading news on line, using an online dictionary or encyclopaedia, participating in online group discussions and searching for information on line, are generally more proficient readers than students who do little online reading.

On average across OECD countries, 37% of students – and 45% or more in Austria, the Netherlands, and Luxembourg – report that they do not read for enjoyment at all. In all but a few countries, students who use appropriate strategies to understand and remember what they read, such as underlining important parts of the texts or discussing what they read with other people, perform at least 73 points higher in the PISA assessment – that is, one full proficiency level or nearly two full school years – than students who use these strategies the least. In Belgium, Switzerland and Austria, the quarter of students who use these strategies the most score an average of 110 points higher than the quarter of students who use them the least. That translates into a difference of roughly one-and-a-half proficiency levels or nearly three years of formal schooling.

In all countries, boys are not only less likely than girls to say that they read for enjoyment, they also have different reading habits when they do read for pleasure. Most boys and girls in the countries that took part in PISA 2009 sit side by side in the same classrooms and work with similar teachers. Yet, PISA reveals that in OECD countries, boys are on average 39 points behind girls in reading, the equivalent of one year of schooling. PISA suggests that differences in how boys and girls approach learning and how engaged they are in reading account for most of the gap in reading performance between boys and girls, so much so that this gap could be predicted to shrink by 14 points if boys approached learning as positively as girls, and by over 20 points if they were as engaged in reading as girls. This does not mean that if boys’ engagement and awareness of learning strategies rose by this amount the increase would automatically translate into respective performance gains, since PISA does not measure causation. But since most of the gender gap can be explained by boys being less engaged, and less engaged students show lower performance, then policy makers should look for more effective ways of increasing boys’ interest in reading at school or at home.

PISA reveals that, although girls have higher mean reading performance, enjoy reading more and are more aware of effective strategies to summarise information than boys, the differences within genders are far greater than those between the genders. Moreover, the size of the gender gap varies considerably across countries, suggesting that boys and girls do not have inherently different interests and academic strengths, but that these are mostly acquired and socially induced. The large gender gap in reading is not a mystery: it can be attributed to differences that have been identified in the attitudes and behaviours of boys and girls.

Girls are more likely than boys to be frequent readers of fiction, and are also more likely than boys to read magazines. However, over 65% of boys regularly read newspapers for enjoyment and only 59% of girls do so. Although relatively few students say that they read comic books regularly, on average across OECD countries, 27% of boys read comic books several times a month or several times a week, while only 18% of girls do so.

High-performing countries are also those whose students generally know how to summarise information. Across OECD countries, the difference in reading performance between those students who know the most about which strategies are best for summarising information and those who know the least is 107 score points. And students who say that they begin the learning process by figuring out what they need to learn, then ensure that
they understand what they read, figure out which concepts they have not fully grasped, try to remember the most important points in a text and look for additional clarifying information when they do not understand something they have read, tend to perform better on the PISA reading scale than those who do not.

**While factors such as predisposition, temperament, peer pressure and socialisation may contribute to boys having less interest in reading than girls, boys could be encouraged to enjoy reading more and to read more for enjoyment.**

PISA results suggest that boys would be predicted to catch up with girls in reading performance if they had higher levels of motivation to read and used effective learning strategies. In Finland, for example, if boys were equally aware as girls of the most effective ways of summarising complex information in their reading, their scores in the PISA assessment would be predicted to be 23 points higher. Similarly, in most of the countries that participated in PISA 2009, if the most socio-economically disadvantaged students had the same levels of awareness about these strategies as their most advantaged peers, their reading performance would be predicted to be at least 15 points higher.

Across OECD countries, if socio-economically disadvantaged students were as aware of effective strategies to summarise information as advantaged students, the performance gap between the two groups of students could be 20% narrower. The poor reading proficiency seen among socio-economically disadvantaged boys is of particular concern because, without the ability to read well enough to participate fully in society, these students and their future families will have fewer opportunities to escape a cycle of poverty and deprivation. On average in the OECD area, socio-economically disadvantaged boys would be predicted to perform 28 points higher in reading if they had the same level of awareness of effective summarising strategies as socio-economically advantaged girls and 35 points higher if they enjoyed reading as much as socio-economically disadvantaged girls.

In recent years, the gender gap in reading engagement has widened, as has the gender gap in reading performance. Changing students’ attitudes and behaviours may be inherently more difficult than providing equal access to high quality teachers and schools, two of the factors that explain the low performance of socio-economically disadvantaged students – an area where PISA shows that over the past decade, some countries have achieved significant progress.

The following table provides selected results.
- The first column shows students’ mean reading scores.
- The second column shows the percentage of students who reported high levels of awareness about effective learning strategies and who regularly read a wide range of materials, including fiction and non-fiction books or at least magazines and newspapers, for enjoyment (considered ‘wide and deep’ or ‘narrow and deep’ readers).
- The third column shows the score point differences in reading between boys and girls, with negative numbers indicating an advantage for boys and positive numbers indicating an advantage for girls.
- The fourth column shows gender differences in the percentage of ‘wide and deep’ or ‘narrow and deep’ readers.
- The fifth column shows the portion of the gender gap that would be predicted to be closed if boys had the same level of enjoyment of reading as girls.
- The sixth column shows the score point difference between the top and bottom quarters of the socio-economic distribution of students.
- The seventh column shows the differences in the share of students who are ‘wide and deep’ or ‘narrow and deep’ readers between the top and bottom quarters of the socio-economic distribution of students. Larger numbers indicate a higher share of ‘wide and deep’ or ‘narrow and deep’ readers among socio-economically advantaged students.
- The last column shows the portion of the socio-economic gap in reading performance that would be predicted to be closed if socio-economically disadvantaged students had the same level of awareness of effective reading strategies (here, summarising strategies) as socio-economically advantaged students.

Values that are larger than the OECD average are shown in light blue; while values that are smaller than the OECD average are shown in medium blue and values that are not statistically different from the OECD average are shown in dark blue.
### Table I. Comparing the contribution of students' engagement in reading and approaches to learning to reading performance and equity

| OECD average | Mean Score | Percentage of "wide and deep" and "narrow and deep readers" | Difference in reading performance (G - B) | Difference in the percentage of girls and boys who can be considered "wide and deep" and "narrow and deep" readers (G - B) | Proportion of the overall gender gap that could be closed if boys enjoyed reading as much as girls | Socio-economic differences in the percentage of students who are "wide and deep" and "narrow and deep" readers (top - bottom quarter of ESCS) | Socio-economic differences in the percentage of students who are "wide and deep" and "narrow and deep" readers (top - bottom quarter of ESCS) | Proportion of the socio-economic gap that could be closed if socio-economically disadvantaged students had values on the index of summarising as socio-economically advantaged students |
|--------------|------------|---------------------------------------------------------------|----------------------------------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| OECD average | 493        | 45                                                           | 39                                     | 11                                                                               | 61                                                                               | 89                                                                               | 17                                                                               | 20                                                                               |

#### Figures

**Figure III.**

Comparing the contribution of students' engagement in reading and approaches to learning to reading performance and equity.

**Note:**

- Statistically significantly above the OECD average
- Not statistically significantly different from the OECD average
- Statistically significantly below the OECD average

**Source:** OECD, PISA 2009 Database.
WHAT MAKES A SCHOOL SUCCESSFUL? RESOURCES, POLICIES AND PRACTICES

Since school is where most learning happens, what happens in school has a direct impact on learning. In turn, what happens in school is influenced by the resources, policies and practices approved at higher administrative levels in a country’s education system.

Successful school systems – those that perform above average and show below-average socio-economic inequalities – provide all students, regardless of their socio-economic backgrounds, with similar opportunities to learn.

Systems that show high performance and an equitable distribution of learning outcomes tend to be comprehensive, requiring teachers and schools to embrace diverse student populations through personalised educational pathways. In contrast, school systems that assume that students have different destinations with different expectations and differentiation in terms of how they are placed in schools, classes and grades often show less equitable outcomes without an overall performance advantage.

Earlier PISA assessments showed these expectations to be mirrored in how students perceived their own educational future. The results of these differences can also be seen in the distribution of student performance within countries and in the impact that socio-economic background has on learning outcomes:

- In countries, and in schools within countries, where more students repeat grades, overall results tend to be worse.
- In countries where more students repeat grades, socio-economic differences in performance tend to be wider, suggesting that people from lower socio-economic groups are more likely to be negatively affected by grade repetition.
- In countries where 15-year-olds are divided into more tracks based on their abilities, overall performance is not enhanced, and the younger the age at which selection for such tracks first occurs, the greater the differences in student performance, by socio-economic background, by age 15, without improved overall performance.
- In school systems where it is more common to transfer weak or disruptive students out of a school, performance and equity both tend to be lower. Individual schools that make more use of transfers also perform worse in some countries.

These associations account for a substantial amount of the differences in the outcomes of schooling systems. For example, the frequency with which students are transferred across schools is associated with a third of the variation in country performance. This does not necessarily mean that if transfer policies were changed, a third of country differences in reading performance would disappear, since PISA does not measure cause and effect. Transferring pupils who do badly may be partly a symptom, rather than a cause, of schools and school systems that are not producing satisfactory results, especially for lower-achieving students. It is worth noting that the schools with lower transfer rates tend to have greater autonomy and other means of addressing these challenges. The cluster of results listed above suggests that, in general, school systems that seek to cater to different students’ needs through a high level of differentiation in the institutions, grade levels and classes have not succeeded in producing superior overall results, and in some respects they have lower-than-average and more socially unequal performance.

Most successful school systems grant greater autonomy to individual schools to design curricula and establish assessment policies, but these school systems do not necessarily allow schools to compete for enrolment.

The incentive to deliver good results for all students is not just a matter of how a school’s student body is defined. It also depends on the ways in which schools are held accountable for their results and what forms of autonomy they are allowed to have – and how that could help influence their performance. PISA has looked at accountability both in terms of the information that is made available about performance and in terms of the use made of that information – whether by administrative authorities through rewards or control systems, or by parents, for example through their choice of school. Thus the issues of autonomy, evaluation, governance and choice interact in providing a framework in which schools are given the incentives and the capacity to improve. PISA 2009 finds that:

- In countries where schools have greater autonomy over what is taught and how students are assessed, students tend to perform better.
- Within countries where schools are held to account for their results through posting achievement data publicly, schools that enjoy greater autonomy in resource allocation tend to do better than those with less autonomy. However, in countries where there are no such accountability arrangements, the reverse is true.
- Countries that create a more competitive environment in which many schools compete for students do not systematically produce better results.
Within many countries, schools that compete more for students tend to have higher performance, but this is often accounted for by the higher socio-economic status of students in these schools. Parents with a higher socio-economic status are more likely to take academic performance into consideration when choosing schools.

In countries that use standards-based external examinations, students tend to do better overall, but there is no clear relationship between performance and the use of standardised tests or the public posting of results at the school level. However, performance differences between schools with students of different social backgrounds are, on average, lower in countries that use standardised tests.

After accounting for the socio-economic and demographic profiles of students and schools, students in OECD countries who attend private schools show performance that is similar to that of students enrolled in public schools. On average, socio-economically disadvantaged parents are over 13 percentage points more likely than socio-economically advantaged parents to report that they consider “low expenses” and “financial aid” as very important determinants in choosing a school. If children from socio-economically disadvantaged backgrounds cannot attend high-performing schools because of financial constraints, then school systems that offer parents more choice of schools for their children will necessarily be less effective in improving the performance of all students.

School systems considered successful tend to prioritise teachers’ pay over smaller classes.

School systems differ in the amount of time, human, material and financial resources they invest in education. Equally important, school systems also vary in how these resources are spent:

- At the level of the school system and net of the level of national income, PISA shows that higher teachers’ salaries, but not smaller class sizes, are associated with better student performance. Teachers’ salaries are related to class size in that if spending levels are similar, school systems often make trade-offs between smaller classes and higher salaries for teachers. The findings from PISA suggest that systems prioritising higher teachers’ salaries over smaller classes tend to perform better, which corresponds with research showing that raising teacher quality is a more effective route to improved student outcomes than creating smaller classes.

- Within countries, schools with better resources tend to do better only to the extent that they also tend to have more socio-economically advantaged students. Some countries show a strong relationship between schools’ resources and their socio-economic and demographic background, which indicates that resources are inequitably distributed according to schools’ socio-economic and demographic profiles.

- In other respects, the overall lack of a relationship between resources and outcomes does not show that resources are not important, but that their level does not have a systematic impact within the prevailing range. If most or all schools have the minimum resource requirements to allow effective teaching, additional material resources may make little difference to outcomes.

In more than half of all OECD countries, over 94% of 15-year-old students reported that they had attended pre-primary school for at least some time.

Students who had attended pre-primary school tend to perform better than students who have not. This advantage is greater in school systems where pre-primary education lasts longer, where there are smaller pupil-to-teacher ratios at the pre-primary level and where there is higher public expenditure per pupil at that level of education. Across all participating countries, school systems with a higher proportion of students who had attended pre-primary education tend to perform better.

Schools with better disciplinary climates, more positive behaviour among teachers and better teacher-student relations tend to achieve higher scores in reading.

Across OECD countries, 81% of students report that they feel they can work well in class most of the time, 71% report that they never, or only in some classes, feel that other students don’t listen, and 72% say that their teacher never, or only in some lessons, has to wait a long time before students settle down to learn.

Meanwhile, 28% of students in OECD countries are enrolled in schools whose principals report that their teaching staff’s resistance to change negatively affects students or that students’ needs are not met; 23% attends schools whose principals report that students are not encouraged by teachers in the school; 22% attend schools whose principals believe that learning is hindered by low teacher expectations; and 17% of students attend schools whose principals say that teacher absenteeism hampers learning.
Comparing school systems...

...some characteristics are related to better reading performance and/or greater equity throughout the system. For example...

![Figure I.a](image)

### Executive Summary

#### Attachment B: PISA 2009 Results

**Comparing school systems...**

...some characteristics are related to better reading performance and/or greater equity throughout the system. For example...

| OECD   | Reading performance | Strength of the relationship between students’ socio-economic background and reading performance | Percentage of students in schools that transfer students to other schools due to low achievement, behavioural problems or special learning needs | Index of school responsibility for curriculum and assessment
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean score</strong></td>
<td>% variance explained</td>
<td>%</td>
<td>Mean index</td>
<td>Existence of standards-based external examinations</td>
</tr>
<tr>
<td>Australia</td>
<td>515</td>
<td>12.7</td>
<td>2.9</td>
<td>0.17</td>
</tr>
<tr>
<td>Austria</td>
<td>470</td>
<td>16.6</td>
<td>52.2</td>
<td>-0.11</td>
</tr>
<tr>
<td>Belgium</td>
<td>506</td>
<td>18.3</td>
<td>48.3</td>
<td>0.17</td>
</tr>
<tr>
<td>Canada</td>
<td>524</td>
<td>8.6</td>
<td>12.6</td>
<td>-0.66</td>
</tr>
<tr>
<td>Chile</td>
<td>449</td>
<td>18.7</td>
<td>24.0</td>
<td>-0.09</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>478</td>
<td>12.4</td>
<td>22.1</td>
<td>0.02</td>
</tr>
<tr>
<td>Denmark</td>
<td>495</td>
<td>14.5</td>
<td>6.1</td>
<td>0.05</td>
</tr>
<tr>
<td>Estonia</td>
<td>501</td>
<td>7.6</td>
<td>9.9</td>
<td>0.22</td>
</tr>
<tr>
<td>Finland</td>
<td>536</td>
<td>7.8</td>
<td>3.7</td>
<td>-0.15</td>
</tr>
<tr>
<td>France</td>
<td>496</td>
<td>16.7</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>Germany</td>
<td>497</td>
<td>17.9</td>
<td>24.0</td>
<td>-0.25</td>
</tr>
<tr>
<td>Greece</td>
<td>483</td>
<td>12.5</td>
<td>42.2</td>
<td>-1.25</td>
</tr>
<tr>
<td>Hungry</td>
<td>494</td>
<td>26.0</td>
<td>14.1</td>
<td>0.11</td>
</tr>
<tr>
<td>Iceland</td>
<td>500</td>
<td>6.2</td>
<td>0.0</td>
<td>0.23</td>
</tr>
<tr>
<td>Ireland</td>
<td>496</td>
<td>12.6</td>
<td>0.0</td>
<td>0.01</td>
</tr>
<tr>
<td>Israel</td>
<td>474</td>
<td>12.5</td>
<td>23.1</td>
<td>-0.01</td>
</tr>
<tr>
<td>Italy</td>
<td>486</td>
<td>11.8</td>
<td>20.1</td>
<td>0.20</td>
</tr>
<tr>
<td>Japan</td>
<td>520</td>
<td>8.6</td>
<td>8.4</td>
<td>1.06</td>
</tr>
<tr>
<td>Korea</td>
<td>539</td>
<td>11.0</td>
<td>6.3</td>
<td>0.79</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>472</td>
<td>18.0</td>
<td>68.0</td>
<td>-0.06</td>
</tr>
<tr>
<td>Mexico</td>
<td>425</td>
<td>18.5</td>
<td>32.8</td>
<td>-0.92</td>
</tr>
<tr>
<td>Netherlands</td>
<td>508</td>
<td>12.8</td>
<td>15.2</td>
<td>1.04</td>
</tr>
<tr>
<td>New Zealand</td>
<td>521</td>
<td>16.6</td>
<td>2.7</td>
<td>0.81</td>
</tr>
<tr>
<td>Norway</td>
<td>501</td>
<td>8.6</td>
<td>1.4</td>
<td>-0.57</td>
</tr>
<tr>
<td>Poland</td>
<td>500</td>
<td>14.8</td>
<td>8.2</td>
<td>0.31</td>
</tr>
<tr>
<td>Portugal</td>
<td>489</td>
<td>16.5</td>
<td>0.8</td>
<td>-0.93</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>477</td>
<td>14.6</td>
<td>30.2</td>
<td>0.08</td>
</tr>
<tr>
<td>Slovenia</td>
<td>483</td>
<td>14.3</td>
<td>21.8</td>
<td>-0.38</td>
</tr>
<tr>
<td>Spain</td>
<td>484</td>
<td>13.6</td>
<td>7.3</td>
<td>-0.38</td>
</tr>
<tr>
<td>Sweden</td>
<td>497</td>
<td>13.4</td>
<td>3.4</td>
<td>0.21</td>
</tr>
<tr>
<td>Switzerland</td>
<td>501</td>
<td>14.1</td>
<td>21.1</td>
<td>-0.62</td>
</tr>
<tr>
<td>Turkey</td>
<td>464</td>
<td>19.0</td>
<td>35.1</td>
<td>-1.04</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>494</td>
<td>13.7</td>
<td>2.2</td>
<td>0.38</td>
</tr>
<tr>
<td>United States</td>
<td>500</td>
<td>16.8</td>
<td>12.6</td>
<td>-0.20</td>
</tr>
<tr>
<td><strong>OECD average</strong></td>
<td>491</td>
<td>14.0</td>
<td>17.8</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

| Region | Reading performance | Strength of the relationship between students’ socio-economic background and reading performance | Percentage of students in schools that transfer students to other schools due to low achievement, behavioural problems or special learning needs | Index of school responsibility for curriculum and assessment
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>385</td>
<td>10.7</td>
<td>17.4</td>
<td>-0.42</td>
</tr>
<tr>
<td>Argentina</td>
<td>396</td>
<td>19.6</td>
<td>14.6</td>
<td>-0.57</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>362</td>
<td>7.4</td>
<td>15.4</td>
<td>-0.64</td>
</tr>
<tr>
<td>Brazil</td>
<td>412</td>
<td>13.0</td>
<td>13.8</td>
<td>-0.56</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>429</td>
<td>20.2</td>
<td>34.5</td>
<td>-0.91</td>
</tr>
<tr>
<td>Colombia</td>
<td>413</td>
<td>16.6</td>
<td>41.3</td>
<td>-0.21</td>
</tr>
<tr>
<td>Croatia</td>
<td>476</td>
<td>11.0</td>
<td>18.1</td>
<td>-0.93</td>
</tr>
<tr>
<td>Dubai (UAE)</td>
<td>459</td>
<td>14.2</td>
<td>20.3</td>
<td>0.15</td>
</tr>
<tr>
<td>Hong Kong-China</td>
<td>533</td>
<td>4.5</td>
<td>32.1</td>
<td>0.92</td>
</tr>
<tr>
<td>Indonesia</td>
<td>402</td>
<td>7.8</td>
<td>40.5</td>
<td>0.13</td>
</tr>
<tr>
<td>Jordan</td>
<td>405</td>
<td>7.9</td>
<td>46.0</td>
<td>-1.20</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>390</td>
<td>12.0</td>
<td>12.9</td>
<td>-0.98</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>314</td>
<td>14.6</td>
<td>38.2</td>
<td>-0.25</td>
</tr>
<tr>
<td>Latvia</td>
<td>484</td>
<td>10.3</td>
<td>14.7</td>
<td>-0.54</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>499</td>
<td>8.4</td>
<td>0.0</td>
<td>-0.05</td>
</tr>
<tr>
<td>Lithuania</td>
<td>465</td>
<td>13.6</td>
<td>6.6</td>
<td>0.13</td>
</tr>
<tr>
<td>Macau-China</td>
<td>487</td>
<td>1.8</td>
<td>47.7</td>
<td>0.86</td>
</tr>
<tr>
<td>Montenegro</td>
<td>408</td>
<td>10.0</td>
<td>7.1</td>
<td>-0.97</td>
</tr>
<tr>
<td>Panama</td>
<td>371</td>
<td>18.1</td>
<td>12.4</td>
<td>-0.60</td>
</tr>
<tr>
<td>Peru</td>
<td>370</td>
<td>27.4</td>
<td>26.8</td>
<td>-0.19</td>
</tr>
<tr>
<td>Qatar</td>
<td>372</td>
<td>4.0</td>
<td>44.9</td>
<td>-0.61</td>
</tr>
<tr>
<td>Romania</td>
<td>424</td>
<td>13.6</td>
<td>40.3</td>
<td>-0.16</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>459</td>
<td>11.3</td>
<td>13.6</td>
<td>-0.36</td>
</tr>
<tr>
<td>Serbia</td>
<td>442</td>
<td>9.8</td>
<td>29.8</td>
<td>-1.03</td>
</tr>
<tr>
<td>Shanghai-China</td>
<td>556</td>
<td>12.3</td>
<td>15.3</td>
<td>-0.09</td>
</tr>
<tr>
<td>Singapore</td>
<td>526</td>
<td>15.3</td>
<td>0.7</td>
<td>-0.09</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>495</td>
<td>11.8</td>
<td>17.2</td>
<td>0.38</td>
</tr>
<tr>
<td>Thailand</td>
<td>423</td>
<td>13.3</td>
<td>9.8</td>
<td>0.76</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>416</td>
<td>9.7</td>
<td>13.8</td>
<td>-0.61</td>
</tr>
<tr>
<td>Tunisia</td>
<td>404</td>
<td>8.1</td>
<td>26.0</td>
<td>-1.29</td>
</tr>
<tr>
<td>Uruguay</td>
<td>429</td>
<td>20.7</td>
<td>0.5</td>
<td>0.00</td>
</tr>
</tbody>
</table>

1. In this index, 0 is the OECD average. The higher the value, the greater the autonomy granted to schools in that country/economy.
2. Values between 0 and 1 indicate that standards-based external examinations exist in some parts of the school system, but not throughout.
### Figure IV.b

Comparing schools within each country/economy...

...the climate in class is also associated with reading performance. For example...

For example...

- Students in schools with better disciplinary climates tend to perform better.
- In most countries/economies, students in schools with better teacher-student relations tend to perform better.

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Index 1</th>
<th>Change in the reading score per unit of the change in the reading score per unit of the index of disciplinary climate³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Dubai (UAE)</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>-0.45</td>
<td></td>
</tr>
<tr>
<td>Eastera</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td>Montenegro</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Serbia</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>Macau-China</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Hong Kong-China</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Qatar</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>OECD average</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Albania</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>-0.21</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>-0.20</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Panama</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>Uruguay</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Russian Federation</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>-0.26</td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td>-0.19</td>
<td></td>
</tr>
</tbody>
</table>

Note: Darker bars show statistically significant associations.

1. In this index, zero is the OECD average and one unit is the standard deviation across OECD countries. The higher the value, the better the school climate.
2. In this index, zero is the OECD average and one unit is the standard deviation across OECD countries. The higher the value, the better the teacher-student relations.
LEARNING TRENDS: CHANGES IN STUDENT PERFORMANCE SINCE 2000

The design of PISA does not just allow for a comparison of the relative standing of countries in terms of their learning outcomes; it also enables each country to monitor changes in those outcomes over time. Such changes indicate how successful education systems have been in developing the knowledge and skills of 15-year-olds.

Indeed, some countries have seen impressive improvements in performance over the past decade, sometimes exceeding the equivalent of an average school year’s progress for the entire 15-year-old student population. Some of these countries have been catching up from comparatively low performance levels while others have been advancing further from already high levels. All countries seeking to improve their results can draw encouragement – and learn lessons – from those that have succeeded in doing so in a relatively short period of time.

Changes in student performance over time prove that a country’s performance in reading is not set in stone. In both absolute and relative terms, educational results can improve, and they cannot be regarded either as part of fixed “cultural” differences between countries or as inevitably linked to each country’s state of economic development.

Since both PISA 2000 and PISA 2009 focused on reading, it is possible to track how student performance in reading changed over that period. Among the 26 OECD countries with comparable results in both assessments, Chile, Israel, Poland, Portugal, Korea, Hungary and Germany as well as the partner countries Peru, Albania, Indonesia, Latvia, Liechtenstein and Brazil all improved their reading performance between 2000 and 2009, while performance declined in Ireland, Sweden, the Czech Republic and Australia.

**Between 2000 and 2009, the percentage of low performers in Chile dropped by more than 17 percentage points, while the share of top performers in Korea grew by more than 7 percentage points.**

In many countries, improvements in results were largely driven by improvements at the bottom end of the performance distribution, signalling progress towards greater equity in learning outcomes. Among OECD countries, variation in student performance fell by 3%. On average across the 26 OECD countries with comparable data for both assessments, 18% of students performed below the baseline reading proficiency Level 2 in 2009, while 19% did so in 2000. Among countries where between 40% and 60% of students performed below Level 2 in 2000, Chile reduced that proportion by the largest amount, and Mexico and the partner country Brazil also show important decreases in their share of low performers. Among countries where the proportion of students performing below Level 2 was smaller than 40% but still above the OECD average of 19%, the partner country Latvia reduced the proportion by 13 percentage points, while Portugal, Poland, Hungary, Germany, Switzerland and the partner country Liechtenstein reduced the share by smaller amounts. In Denmark, the percentage of students below Level 2 fell from an already below-average level.

The share of top performers – those students who attain reading proficiency Level 5 or 6 in reading – increased in Japan, Korea and the partner economy Hong Kong-China such that these countries now have the largest proportions of high-achieving students among the countries participating in the 2009 assessment. Several countries that had above-average proportions of top performers in 2000 saw those proportions decrease in 2009. Notable among them was Ireland, where the proportion of top performers fell from 14% to 7%, which is below the OECD average.

Between 2000 and 2009, Poland, Portugal, Germany, Switzerland and the partner countries Latvia and Liechtenstein raised the performance of their lowest-achieving students while maintaining the performance level among their highest-achieving students. Korea, Israel and the partner country Brazil raised the performance of their highest-achieving students while maintaining the performance level among their lowest-achieving students. Chile and the partner countries Indonesia, Albania and Peru showed improvements in reading performance among students at all proficiency levels.

On average, OECD countries narrowed the gap in scores between their highest- and lowest-performing students between 2000 and 2009; some also improved overall performance. In Chile, Germany, Hungary, Poland, Portugal, and the partner countries Indonesia, Latvia and Liechtenstein, overall performance improved while the variation in performance decreased. In many cases, this was the result of improvements among low-achieving students.

**The gender gap in reading performance did not narrow in any country between 2000 and 2009.**

The gender gap in reading performance widened in Israel, Korea, Portugal, France and Sweden, and in the partner countries and economies Romania, Hong Kong-China, Indonesia and Brazil between 2000 and 2009. The fact that girls outperform boys in reading is most evident in the proportion of girls and boys who perform below baseline
Executive Summary

proficiency Level 2. Across OECD countries, 24% of boys perform below Level 2 compared to only 12% of girls. The proportion of girls performing below this level decreased by two percentage points between 2000 and 2009, while the share of low-achieving boys did not change during the period.

Across the OECD area, the percentage of students with an immigrant background increased by an average of two percentage points between 2000 and 2009. The performance gap between students with and without an immigrant background remained broadly similar over the period. However, some countries noted large reductions in the performance advantage of students without an immigrant background. In Belgium, Switzerland and Germany, the gap narrowed by between 28 and 38 score points due to improvements in reading proficiency among students with an immigrant background. However, the gap is still relatively wide in these countries.

Across OECD countries, overall performance in mathematics remained unchanged between 2003 and 2009, as did performance in science between 2006 and 2009.

In mathematics, students in Mexico, Turkey, Greece, Portugal, Italy, Germany and the partner countries Brazil and Tunisia improved their mathematics scores considerably, while students in the Czech Republic, Ireland, Sweden, France, Belgium, the Netherlands, Denmark, Australia and Iceland saw declines in their performance. On average across the 28 OECD countries with comparable results in the 2003 and 2009 assessments, the share of students below mathematics proficiency Level 2 remained broadly similar over the period, with a minor decrease from 21.6% to 20.8%. Among the OECD countries in which more than half of students performed below mathematics proficiency Level 2 in 2003, Mexico shrunk this proportion by 15 percentage points, from 66% to 51%, by 2009 while Turkey reduced it from 52% to 42% during the same period. Meanwhile, the percentage of top performers in mathematics in those 28 OECD countries decreased slightly, from 14.7% in 2003 to 13.4% in 2009. Portugal showed the largest increase – four percentage points – in top performers in mathematics.

In science, 11 of the 56 countries that participated in both the 2006 and 2009 assessments show improvements in student performance. Turkey, for example, saw a 30 score point increase, nearly half a proficiency level, in just three years. Turkey also reduced the percentage of students below science proficiency Level 2 by almost 17 percentage points, from 47% to 30%. Portugal, Chile, the United States, Norway, Korea and Italy all reduced the share of lowest performers in science by around five percentage points or more, as did the partner countries Qatar, Tunisia, Brazil and Colombia. Performance in science declined considerably in five countries.

On average across OECD countries, the percentage of students who report reading for enjoyment daily dropped by five percentage points.

Enjoyment of reading tends to have deteriorated, especially among boys, signalling the challenge for schools to engage students in reading activities that 15-year-olds find relevant and interesting. On average across OECD countries, the percentage of students who said they read for enjoyment every day fell from 69% in 2000 to 64% in 2009. On the other hand, changes in student-teacher relations and classroom climate have generally been favourable or, at least, they have not deteriorated as many would have expected. Generally, students have become more confident that they can get help from their teachers. Across the 26 OECD countries that participated in both assessments, 74% of students in 2000 agreed or strongly agreed with the statements, “If I need extra help, I will receive it from my teachers” or “Most of my teachers treat me fairly”, while in 2009, 79% of students agreed or strongly agreed with those statements. Overall, aspects of classroom discipline have also improved. Thus there is no evidence to justify the notion that students are becoming progressively more disengaged from school.
## Figure V.

**Comparing a Summary of Changes in Reading Performance**

<table>
<thead>
<tr>
<th>Country</th>
<th>Change in reading performance between 2000 to 2009</th>
<th>Mean score in reading 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peru</td>
<td>All students: -3.4</td>
<td>Girls: 35</td>
</tr>
<tr>
<td>Chile</td>
<td>All students: -2.5</td>
<td>Girls: 35</td>
</tr>
<tr>
<td>Albania</td>
<td>All students: -2.9</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Indonesia</td>
<td>All students: -3.5</td>
<td>Girls: 39</td>
</tr>
<tr>
<td>Latvia</td>
<td>All students: -3.5</td>
<td>Boys: 39</td>
</tr>
<tr>
<td>Israel</td>
<td>All students: -1.5</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Poland</td>
<td>All students: -2.4</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Portugal</td>
<td>All students: -2.7</td>
<td>Boys: 39</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>All students: -2.7</td>
<td>Boys: 39</td>
</tr>
<tr>
<td>Brazil</td>
<td>All students: -3.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Hungary</td>
<td>All students: -4.0</td>
<td>Boys: 39</td>
</tr>
<tr>
<td>Germany</td>
<td>All students: -4.0</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Greece</td>
<td>All students: -4.5</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Hong Kong-China</td>
<td>All students: -4.5</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Switzerland</td>
<td>All students: -4.8</td>
<td>Boys: 39</td>
</tr>
<tr>
<td>Mexico</td>
<td>All students: -4.0</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Belgium</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Italy</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Denmark</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Norway</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Japan</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Korea</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>United States</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Ireland</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>New Zealand</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>France</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Switzerland</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Thailand</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Canada</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Ireland</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>New Zealand</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Spain</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Australia</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Sweden</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Argentina</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
<tr>
<td>Iceland</td>
<td>All students: -4.2</td>
<td>Boys: 35</td>
</tr>
</tbody>
</table>

Note: Countries are ranked in descending order of the change in reading performance between 2000 and 2009 for all students.

Source: OECD, PISA 2009 Database, Tables V.1, V.2.2, V.2.4 and V.4.3

http://dx.doi.org/10.1787/888932359948
ORGANISATION FOR ECONOMIC CO-OPERATION
AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation’s statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.
EXECUTIVE SUMMARY

THE NAEP BACKGROUND QUESTIONS:
AN UNDERUSED NATIONAL RESOURCE

A REPORT TO THE NATIONAL ASSESSMENT GOVERNING BOARD
by the EXPERT PANEL ON STRENGTHENING
NAEP BACKGROUND QUESTIONS

For more than four decades the National Assessment of Educational Progress (NAEP) has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions. Less known, but complimenting the NAEP assessments, is a rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questionnaires have been cut back over the past decade. They now cover only a small fraction of important student, teacher, and school issues and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education including:

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.

- Tracking progress in implementing key instructional, curricular, and technological changes and educational policy initiatives, such as the Common Core standards.

- Monitoring student motivation and out-of-school learning as research-based factors affecting student achievement.

- Benchmarking high-performing states and urban districts and those with high achievement growth to identify factors that differentiate high-performers from lower-performers on NAEP. This domestic effort would parallel the extensive reporting of background variables in PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) that have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing and low-performing education systems.
The panel proposes building a strategy to make the NAEP background questions an important national resource for educators, policymakers, and the public. The panel sees the need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions though regular publications, and improve the capacity for analysis by users around the world. We offer recommendations in four areas (see Exhibit A):

1. Ask Important Questions.
2. Improve the Accuracy of Measures.
3. Strengthen Sampling Efficiency.
4. Reinstitute Meaningful Analysis and Reporting.

Exhibit A. Expert Panel Recommendations to Strengthen NAEP Background Questions in Four Areas

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Core questions</td>
<td>• Valid</td>
<td>• Spiral sampling</td>
<td>• Special background question reports</td>
</tr>
<tr>
<td>• Rotated questions</td>
<td>• Reliable</td>
<td>• Extended questionnaire time</td>
<td>• Online compendium of responses</td>
</tr>
<tr>
<td>• Policy questions</td>
<td>• Coordinated (with domestic and international surveys)</td>
<td>• Alternate surveys</td>
<td>• Report descriptive not causal findings</td>
</tr>
<tr>
<td>• Theoretical frameworks</td>
<td>• Cognitive labs</td>
<td>• Pooling item responses across surveys</td>
<td>• Externally conducted research</td>
</tr>
<tr>
<td>• Consistent questions</td>
<td></td>
<td></td>
<td>• Improve online tools</td>
</tr>
<tr>
<td>overtime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Delete duplicative or low-priority questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- Establish a single NCB committee overseeing background questions
- Review budget including need for staff to implement recommendations

Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Areas.

To the extent that you don’t ask and analyze important questions, you can’t expect to get back important answers. The panel recommends identifying topics falling into three question groups.

- A common core set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.

- A second tier of priority background question clusters would be rotated across assessment cycles. Important topics that might be explored include school-parent
cooperation, school climate and discipline, school administration including support for learning, and out-of-school learning time.

- A third tier would be a set of policy issues that would be examined for six years and then rotated out with new ones added. For example, the initial set might start with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations could be added, and two years after that a module on project-based or online learning.

Once question topics are identified, the panel urges the selection of clusters of questions that collectively best portray different important aspects of research-based theoretical frameworks for the major educational topics. Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement.

The Panel recommends two additional considerations to maximize the information worth of the questions chosen. The first is to pay greater attention to the consistency of question selection and wording to produce reliable time-series that measure change over time. A review of 400 questions asked about teachers found that about 300 are no longer used, with many replaced by just slightly different wording. A second recommendation is to balance the number of questions asked about a topic with the information value gained. Eight questions are asked about technology use in mathematics but there are no questions about student expectations despite the strong research connection with achievement.

**Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for the Background Questions.**

The panel urges attention to strengthening the validity, reliability and coordination of NAEP background questions. An important first step in this overall effort would be to improve the validity, reliability and coordination of the current measures NAEP uses for its mandated student reporting categories. The panel strongly supports the current review of the SES variables as it is critical to respond to the known limitations of the school-lunch proxy. These problems will worsen with expansion of the Department of Agriculture state pilots, which allow whole-school eligibility for schools serving concentrations of low-income students. The panel also believes that an expanded cognitive interview capability, such as a small standing panel of respondents to test out questions, would improve question validity and reliability. We recognize that this may increase costs but it would help make NAEP a better source of information.

The panel recommends improving question wording by replacing imprecise terms such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week.” Furthermore, major information benefits would accrue from coordinating the NAEP background questions with those asked on other international and domestic surveys. To illustrate, the PISA international survey covers number of hours of math instruction in-school and out-of-school; NAEP only asks about days taught math in-
school and only about participation in math instruction outside of school and nothing about frequency.

**Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy.**

The panel recommends that NAEP should consider expanding the depth of its background questions through a variety of strategies including spiral sampling (already under study), expanded questionnaire time and rotating background questions across samples. The panel notes that the depth of student information in particular is limited by the ten-minute questionnaire time limit compared with 30 minutes used for TIMSS and PISA. A combination of these strategies would allow NAEP to obtain far richer information while maintaining sampling accuracy and still keeping respondent burden to acceptable levels.

**Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of the NAEP Background Questions.**

This set of recommendations would bolster the analysis and reporting of the background questions by means of separate publications, online tables, and improvements to the Data Explorer. The recommendations also include a reiteration of current policy to not use causal interpretations of point-in-time data.

The panel strongly recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. Exhibit B displays an illustrative overview table for in-school learning opportunities for math that suggests the rich potential information payoffs from background question analyses. A third benchmarking report should also be considered that explores the correlates of high-performing states and districts or those with high achievement growth. These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

**Implementation of Recommendations**

The panel urges the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES) to move quickly to begin implementing its recommendations to make the background questions a more useful resource, while also recognizing that implementation will take time.

Initial implementation should be undertaken through a three-part plan:

- Immediately produce special reports on the background data that analyze the considerable quantity of data already collected, but is largely unreported and unanalyzed.
### Exhibit B: Illustrative Table of Background Question Indicators With a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment

<table>
<thead>
<tr>
<th>Jurisdictions</th>
<th>Grade 8 All Students</th>
<th>Eligible for National School Lunch</th>
<th>Grade 8 Students Absent 5 or more days last month</th>
<th>Grade 8 Students 5 or more Hours of Math Per Week</th>
<th>Grade 8 Students 1 Hour or More Homework</th>
<th>Grade 8 Does Math At An After-school or Tutoring Program</th>
<th>Grade 8 Does Math Through Alternative Certification</th>
<th>Grade 8 Teacher Has Math Major/Minor/ Special Emphasis</th>
<th>Grade 8 Full-time Math Specialist At School</th>
<th>Grade 8 Assigned To Math By Ability</th>
<th>Grade 8 26+ Students in Math Class</th>
<th>Grade 8 Computers Available to Teachers and Students</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>284</td>
<td>44</td>
<td>7</td>
<td>42</td>
<td>37</td>
<td>17</td>
<td>21</td>
<td>17</td>
<td>38</td>
<td>37</td>
<td>76</td>
<td>45</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albuquerque</td>
<td>275</td>
<td>60</td>
<td>8</td>
<td>37</td>
<td>65</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>33</td>
<td>32</td>
<td>66</td>
<td>59</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlanta</td>
<td>266</td>
<td>82</td>
<td>5</td>
<td>27</td>
<td>75</td>
<td>38</td>
<td>57</td>
<td>57</td>
<td>95</td>
<td>61</td>
<td>59</td>
<td>37</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austin</td>
<td>287</td>
<td>59</td>
<td>8</td>
<td>23</td>
<td>61</td>
<td>27</td>
<td>30</td>
<td>42</td>
<td>57</td>
<td>58</td>
<td>53</td>
<td>52</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baltimore City</td>
<td>261</td>
<td>85</td>
<td>9</td>
<td>46</td>
<td>93</td>
<td>41</td>
<td>28</td>
<td>28</td>
<td>79</td>
<td>53</td>
<td>86</td>
<td>37</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>282</td>
<td>76</td>
<td>9</td>
<td>66</td>
<td>76</td>
<td>39</td>
<td>30</td>
<td>13</td>
<td>69</td>
<td>12</td>
<td>61</td>
<td>47</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlotte</td>
<td>285</td>
<td>52</td>
<td>8</td>
<td>35</td>
<td>87</td>
<td>18</td>
<td>29</td>
<td>44</td>
<td>47</td>
<td>33</td>
<td>86</td>
<td>76</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago</td>
<td>270</td>
<td>84</td>
<td>4</td>
<td>32</td>
<td>67</td>
<td>47</td>
<td>37</td>
<td>23</td>
<td>84</td>
<td>20</td>
<td>45</td>
<td>65</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleveland</td>
<td>256</td>
<td>100</td>
<td>11</td>
<td>29</td>
<td>69</td>
<td>33</td>
<td>25</td>
<td>6</td>
<td>58</td>
<td>14</td>
<td>51</td>
<td>44</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dallas</td>
<td>274</td>
<td>85</td>
<td>7</td>
<td>32</td>
<td>46</td>
<td>27</td>
<td>39</td>
<td>61</td>
<td>66</td>
<td>13</td>
<td>45</td>
<td>24</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detroit</td>
<td>246</td>
<td>79</td>
<td>17</td>
<td>24</td>
<td>81</td>
<td>46</td>
<td>37</td>
<td>11</td>
<td>83</td>
<td>39</td>
<td>18</td>
<td>85</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia (DCPS)</td>
<td>255</td>
<td>70</td>
<td>12</td>
<td>53</td>
<td>65</td>
<td>29</td>
<td>39</td>
<td>57</td>
<td>68</td>
<td>40</td>
<td>53</td>
<td>80</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresno</td>
<td>256</td>
<td>88</td>
<td>10</td>
<td>51</td>
<td>32</td>
<td>11</td>
<td>26</td>
<td>6</td>
<td>37</td>
<td>23</td>
<td>91</td>
<td>75</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillsborough County (FL)</td>
<td>282</td>
<td>54</td>
<td>9</td>
<td>87</td>
<td>20</td>
<td>13</td>
<td>22</td>
<td>40</td>
<td>35</td>
<td>29</td>
<td>95</td>
<td>3</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houston</td>
<td>279</td>
<td>76</td>
<td>6</td>
<td>29</td>
<td>63</td>
<td>26</td>
<td>37</td>
<td>56</td>
<td>63</td>
<td>25</td>
<td>84</td>
<td>58</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jefferson County (KY)</td>
<td>274</td>
<td>60</td>
<td>7</td>
<td>40</td>
<td>68</td>
<td>14</td>
<td>20</td>
<td>21</td>
<td>34</td>
<td>36</td>
<td>77</td>
<td>80</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Los Angeles</td>
<td>261</td>
<td>82</td>
<td>6</td>
<td>67</td>
<td>44</td>
<td>40</td>
<td>27</td>
<td>39</td>
<td>67</td>
<td>37</td>
<td>75</td>
<td>52</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miami-Dade</td>
<td>272</td>
<td>72</td>
<td>5</td>
<td>36</td>
<td>43</td>
<td>47</td>
<td>25</td>
<td>38</td>
<td>72</td>
<td>25</td>
<td>90</td>
<td>13</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milwaukee</td>
<td>254</td>
<td>81</td>
<td>13</td>
<td>30</td>
<td>78</td>
<td>43</td>
<td>31</td>
<td>37</td>
<td>74</td>
<td>82</td>
<td>28</td>
<td>86</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York City</td>
<td>272</td>
<td>87</td>
<td>10</td>
<td>28</td>
<td>83</td>
<td>26</td>
<td>39</td>
<td>35</td>
<td>65</td>
<td>36</td>
<td>60</td>
<td>83</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia</td>
<td>265</td>
<td>88</td>
<td>10</td>
<td>34</td>
<td>89</td>
<td>27</td>
<td>27</td>
<td>24</td>
<td>54</td>
<td>32</td>
<td>30</td>
<td>75</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego</td>
<td>278</td>
<td>60</td>
<td>8</td>
<td>69</td>
<td>48</td>
<td>13</td>
<td>27</td>
<td>11</td>
<td>40</td>
<td>17</td>
<td>78</td>
<td>72</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer

- Move quickly to initiate a long-term effort to improve the relevance, quality, coherence, and usefulness of a core and rotated set of background variables while implementing recommended improvements to improve measurement accuracy and sampling efficiency.

- Further improve the usability of the Data Explorer and other NCES online tools, which are already valuable analytic supports.

The panel suggests that NAGB establish a separate standing committee to review all background questions and plans to improve their use. Currently, the Board’s responsibilities for background questions are divided between two of its standing committees. These subgroups do not coordinate their work and the background questionnaires are of secondary interest to both of them. A unified standing committee should regularly monitor and report on implementation of the panel’s recommendations by NCES and Governing Board staff.

In addition, the panel believes that the background questions and how they used in NAEP reporting warrant a periodic, rigorous, and independent evaluation similar to that conducted in the past on NAEP cognitive assessment items.

The panel recognizes that implementing its recommendations will involve resource considerations in terms of time, money, and personnel. One approach to this problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate lower-priority activities, such as the duplicative collection of racial data and the...
disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

In conclusion, the NAEP background questions are a unique national information resource. The Governing Board and NCES have a responsibility to develop this resource to better understand academic achievement and the contexts in which it occurs and, hopefully, to help spur educational improvement.
NATIONAL ASSESSMENT GOVERNING BOARD

Expert Panel on Strengthening
NAEP Background Questions

MEMBERSHIP AND AFFILIATIONS

• Marshall S. Smith (Chair)
  Visiting Scholar, Carnegie Foundation for the Advancement of Teaching
  Former U.S. Under Secretary of Education
  Palo Alto, CA

• Naomi Chudowsky
  Co-Director
  Caldera Research, LLC
  Bend, OR

• Alan Ginsburg (Secretary)
  Education Consultant
  Former Director of Policy and Program Evaluation
  U.S. Department of Education
  Washington, DC

• Robert Hauser
  Executive Director
  Division of Behavioral and Social Sciences and Education
  National Research Council
  Washington, DC

• Jennifer Jennings
  Assistant Professor of Sociology
  New York University
  New York, NY

• Sharon Lewis
  Research Director
  Council of the Great City Schools
  Washington, DC
NAEP BACKGROUND QUESTIONS: AN UNDERUSED NATIONAL RESOURCE

A Report to the National Assessment Governing Board by the Expert Panel on Strengthening the NAEP Background Questions

February 22, 2012

Chair: Marshall S. Smith

Members: Naomi Chudowsky, Alan Ginsburg, Robert Hauser, Jennifer Jennings, and Sharon Lewis
Table of Contents

Introduction ..................................................................................................................................................3

Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent
Groups of Important Background Questions around High-Priority Areas ........7
   Questionnaire Overview ..........................................................................................................................7
   Recommendation 1a. Continually review and refine a core and second-tier set of
   background topics and questions that are common across NAEP surveys.........................9
   Recommendation 1b. Extend NAEP background questions to inform topics of current
   policy interest in the subjects assessed ..............................................................................................10
   Recommendation 1c. Select clusters of questions that collectively best measure
   different aspects of research-based theoretical frameworks for major educational
   topics ..................................................................................................................................................12
   Recommendation 1d. Use consistency over time as a criterion to consider for question
   selection and wording ..........................................................................................................................14
   Recommendation 1e. Delete duplicative or low-priority questions to make time for the
   Panel’s higher priority items ..................................................................................................................15

Recommendation Area 2. Strengthen the Validity, Reliability and Coordination
of the Measures and Clusters of Measures for the Background Questions ....16
   Recommendation 2a. Improve the validity and reliability of the current measures
   NAEP uses for its mandated student reporting categories ...............................................................17
   Recommendation 2b. Enhance the validity of student responses at different grade
   levels ..................................................................................................................................................19
   Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-
   learn skills including student learning behaviors, motivation and expectations..............20
   Recommendation 2d. Improve question reliability by replacing imprecise phrases such
   as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or
   more a week”. ..................................................................................................................................21
   Recommendation 2e. Coordinate NAEP background questions with those asked on
   international or domestic surveys ........................................................................................................22
   Recommendation 2f. Build on current NCES cognitive interview techniques by using
   cognitive laboratories, such as small standing panels, to field test questions to
   establish their validity and reliability ...............................................................................................23

Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the
Background Questions While Maintaining Sampling Accuracy .................24
   Recommendation 3a. Support NCES’s exploration of a spiral sample methodology to
   expand the scope of background questions, while assessing the possible loss in the
   representativeness of disaggregated data ..........................................................................................25
   Recommendation 3b. Consider other item-sampling reforms to obtain the needed
   questionnaire time including lengthening the student survey; establishing a 4-year
   interval between administration of some background questions; and pooling item
   responses across survey administrations .......................................................................................25

Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of
the NAEP Background Questions .........................................................................................26
   Recommendation 4a. Prepare special reports highlighting the background question
   findings ................................................................................................................................................27
Recommendation 4b. Prepare an online compendium of key background indicators for States and participating urban districts.

Recommendation 4c. NAEP’s reports should not indicate causal interpretations using the background questions. However, the NAEP data offer some unique advantages for generating relationships and hypotheses about factors that may be associated with performance and these findings should guide more rigorous in-depth follow-on analyses.

Recommendation 4d. NAEP should encourage others to conduct exploratory studies of the NAEP background variables.

Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.

5. Implementing the Panel Recommendations

Recommendation 5a. Exploit existing background data through special reports focused on issues and topics informed by background questions.

Recommendation 5b. Initiate a set of activities to build clusters of core and second-tier questions around high-priority topics for the 2015 NAEP administration.

Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.

Recommendation 5d. Champion implementation by creating a single NAGB subcommittee responsible for the background questions; provide adequate resource support, while ensuring efficient resource use; and publicize background question products and findings.

REFERENCES
Introduction

The National Assessment of Educational Progress (NAEP) is a unique American education resource. For more than four decades the assessment has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions.¹

Representative samples of students regularly take NAEP assessments in reading, mathematics, science, and writing and the national, state, and urban district levels. Other subjects, including U.S. history, civics, and the arts, are tested at the national level only. Less known, but complimenting the NAEP assessments, is a potentially rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questions have been cut back over the past decade. They now cover only a small fraction of important student, teacher and school issues, and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into important questions about the nature and quality of American primary and secondary education. What are the racial, ethnic and economic characteristics of schools at different achievement levels? What are the sources of curriculum content? What resources are available for students? What are the common instructional approaches teachers employ, and how do they adjust approaches to differing student needs? What preparation and training do teachers receive? How is teacher performance evaluated?

In turn, the answers to these survey questions can support important NAEP analyses. The analyses should focus on the unique advantages of NAEP for collecting data and trends over time on education-related background factors paired with achievement results that are representative of states and many urban districts. The following three examples illustrate potentially significant descriptive findings from the NAEP background questions for mathematics with respect to:

¹ Although this report focuses on the lack of reporting the background variables for the main NAEP, a similar weakness occurs in not reporting the background variables for the long-term trend NAEP. The report on the 2008 long-term trend assessments did include data on higher level course taking in math in 2008 in relation to that year’s NAEP scores, but surprisingly did not report results for earlier years, although available.
• Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.
  - In Arizona, a Hispanic grade-8 student is only 57 percent as likely to have a teacher of mathematics who has a major in mathematics as a white grade-8 student. In California, their chances are nearly equal.

• Tracking progress in implementing instructional, curricular, and technological changes and key education policy initiatives.
  - The proportion of students in schools with no eighth-graders enrolled in algebra is 15 percent nationally. Among urban districts, Miami-Dade and Houston have only 5 percent of their students in schools without a grade-8 algebra course, but Detroit and Milwaukee have over 80 percent of eighth-graders in such schools.

• Monitoring student motivation and out-of-school learning as factors affecting student achievement.
  - More than 45 percent of the grade 4 students in several Southern states (Louisiana, South Carolina and Texas) participated in after-school math instruction. But in several highly rural states (Maine, Oregon and Vermont) the participation rate in after-school math instruction was only about 25 percent.

Moreover, the extensive reporting of the background variables in PISA and TIMSS have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing education systems (Darling-Hammond, 2010). These data have been used to examine characteristics of high-performing systems, such as Singapore and Korea, and to study the nature of instruction in subjects such as math and science, where the U.S. performs poorly. In a similar fashion the NAEP data could be used to guide benchmarking of high-performing states and urban districts or jurisdictions experiencing substantial performance growth. This benchmarking activity would be a means to generate hypotheses for further verification though in-depth study. Specific examples of the use of NAEP background questions for domestic benchmarking might include examining:

- A high overall-performing state such as Massachusetts or a state like Texas that has a relatively small white-Hispanic performance gap compared with other states.
- A high-performing district such as New York City that has low-income students achieving above the national average for all low-income students in both reading and math at grades 4 and 8.
- The nearly one standard deviation growth in grade 4 math since 1990 and the instructional, curriculum and teacher changes that occurred over this period.

The panel recognizes the justifiable concern over misuse of the NAEP background variables in making causal interpretations. NAEP is not able to reduce countervailing explanations for causation like a well-designed experiment. Also, successive NAEP assessments will sample different students in the same grade, so the data are not a measure of change over time for the same students as in a true longitudinal design.
However, the panel believes that a valid concern over causal interpretations has led to a serious and unjustified overreaction. NAEP’s national and state representative data uniquely address many important descriptive questions. These data can track progress on variables shown by research to be important for achievement. The NAEP background questions can inform national policies by providing descriptive data about the quality of implementation. Also, because NAEP is already in the schools to administer its assessments, data can be collected at relatively low cost compared with other survey vehicles.

Yet for the past decade NAEP has stopped publishing all but the most minimal background information.

- NAEP no longer systematically reports on the responses to the background questions when publishing its assessment results, except for the congressionally required student reporting categories (e.g., race/ethnicity, low-income).\(^2\)
- In-depth special reports using the background questions are rare (e.g., the 2010 report on American Indian Educational Experiences was an exception).
- Data are made available almost entirely through an online database called the NAEP Data Explorer. This is a useful tool, but it is not a substitute for carefully prepared summary data tables and analyses. Most educators, policy makers and members of the public do not have the time or inclination to master use of the Data Explorer, but many would pay attention to focused reports and make use of summary tabular information.

Reporting the background questions would be a great service to the nation in identifying and tracking important national and state trends in education. Here, the panel finds that the NAEP background questionnaires severely limit their potential usefulness by not explicitly asking questions about the progress and challenges of implementing key national policies in different states and urban districts. Yet the *NAEP Background Information Framework* (2003), which sets out principles to guide background question selection and reporting, explicitly recognizes that the background questions should “focus on the most important variables related to public policy.”

NAEP’s de-emphasis of the background questions is in marked contrast to the significance that all the major international surveys – PISA (Program for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), and PIRLS (Progress in International Reading Literacy Study) – give to background variables in participating countries.

The panel believes NAEP should return to its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, the questions could provide rich insights into a wide range of important issues.

\(^2\) In 2011 NAEP began to use the background variables again in its main assessment reports, but with only a single background table related to instruction for each subject and grade. The 2010 Civics, Geography and U.S. History reports also contained a background table related to instruction for the different grades.
about the nature and quality of American primary and secondary education and the context for understanding achievement and its improvement. The panel believes there is a need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions though regular NAEP publications, and improve the capacity for analysis by data users.

To do so the panel has developed recommendations for improvements in four areas:

1. Ask Important Questions.
2. Improve the Accuracy of the Measures.
3. Strengthen Sampling Efficiency.
4. Reinstitute Meaningful Analysis and Reporting.

Within each area, Exhibit 1 identifies the specific individual recommendations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Core questions</td>
<td>Validate</td>
<td>Spiral sampling</td>
<td>Special reports</td>
<td></td>
</tr>
<tr>
<td>Rotated questions</td>
<td>Reliable</td>
<td>Coordinates</td>
<td>Online compendium of responses</td>
<td></td>
</tr>
<tr>
<td>Policy questions</td>
<td>Coordinated</td>
<td>Extended questionnaire</td>
<td>Report</td>
<td></td>
</tr>
<tr>
<td>Theoretical frameworks</td>
<td>with domestic and international surveys</td>
<td>Alternate surveys</td>
<td>Descriptive not causal findings</td>
<td></td>
</tr>
<tr>
<td>Consistent questions overtime</td>
<td>Cognitive labs</td>
<td>Pooling item responses across surveys</td>
<td>Externally conducted research</td>
<td></td>
</tr>
<tr>
<td>Delete duplicative or low-priority questions</td>
<td>Establish a single NAGB committee overseeing background questions</td>
<td>Reinstitute Meaningful Analyses &amp; Reporting</td>
<td>Improve online tools</td>
<td></td>
</tr>
<tr>
<td>• Establish a single NAGB committee overseeing background questions</td>
<td>• Review budget include need for staff to implement recommendations</td>
<td>• Special reports</td>
<td>• Online compendium of responses</td>
<td></td>
</tr>
</tbody>
</table>

The panel recognizes that these recommendations would require commitments of resources and that the Governing Board and the Commissioner of Education Statistics are in the best position to decide on any tradeoffs between existing and proposed features of NAEP that may be required within NAEP’s budget.
Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Information Areas

To the extent that you don’t ask and analyze important questions you can’t expect to get back important answers. This section recommends strategies for focusing clusters of questions on important information topics within the confines of NAEP questionnaire timelines and administration procedures. Consistent with the NAEP framework, important questions are ones that would primarily focus on the factors that research has shown are related to student achievement. Background questions would also address the implementation of major national policies where NAEP surveys can provide a view from the field state-by-state. In this way, NAEP can report on the distributions and trends of many of the factors and policies important for student achievement.

Questionnaire Overview

With each administration of the subject area assessment, NAEP includes separate student, teacher and school background questionnaires. Although a few questions about subgroups are specified in the NAEP legislation, the Governing Board has the discretion to determine most questions. Exhibit 2 displays the overall number of questions and general question content for each of the three respondent questionnaires on the most recently-reported reading and mathematics surveys.

<table>
<thead>
<tr>
<th></th>
<th>Students 10 Min</th>
<th>Teachers 30 Min</th>
<th>Schools 30 Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Student &amp; family</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>background and out-of-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>school learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Subject specific:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-perception and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>school courses content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questions (subject</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>specific):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Teachers Background:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>education and training;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Classroom Organization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and Instructional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>practices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- School Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(including a special</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>charter school survey)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Subject specific:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>course, student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>placement, staff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>composition, training,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>technology</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gr. 4 (2011) 31 30 40 48 31 39 49 48
Gr. 8 (2011) 32 26 34
Gr. 12 (2009) 39 49 48
Gr. 4 (2011) 31 39 48
Gr. 8* (2011) 49
Gr. 12* (2009) 48

*School questionnaire for grades 8 and 12 covers reading, math and science. Teacher questionnaire is not administered at grade 12.
A 10-minute student questionnaire consisting of approximately 30 questions asks about family background, school and home experiences, and out-of-school learning activities.

- Since NAEP does not administer a questionnaire to survey parents, the student questionnaire is the primary source of information on students’ home characteristics and out-of-school learning activities. (School records do provide an alternative source for race, ethnicity and school lunch eligibility data).
- With respect to socio-economic status, grade 4 students are only asked about household items (computers in the home, numbers of books). Students in grades 8 and 12 are also queried about their mother’s and father’s highest level of education.
- A few questions are asked about students’ out-of-school learning-related activities -- talk about things studied in school, read for fun on your own time, or studying and reading at an after-school program.
- A few items are included about student self-perception and enjoyment of a specific subject, for example whether reading and math are favorite subjects.
- Students are asked a number of questions about their classes in the subject assessed – for example, the frequency of reading aloud and discussing what they read in class, and in math many questions about using technology (calculators, graphing programs and spreadsheets).

A 30-minute teacher questionnaire of 30-40 questions is filled-out by the teacher in grade 4 or 8 in the subject assessed, usually the classroom teacher at grade 4 and the English or mathematics teacher at grade 8. This questionnaire covers:

- Teacher background information on race/ethnicity, education, certification and experience and professional development.
- Classroom organization items about class size, hours of instruction and ability grouping.
- Instructional items about topic emphasis, instructional approach, homework, evaluating student progress and access to resources and technology. The math questionnaire includes extensive questions about calculators of all types, computers, the Internet and CD-ROMs.

A 30-minute school questionnaire of about 40 questions covers:

- Overall school characteristics including grades, status as a charter, student composition and turnover, teacher absenteeism, volunteerism, and Title I federal program participation.
- Subject-specific items about specialist staff, structuring of content with standards and assessments, resource availability with emphasis on technology,
- Special charter school questionnaire about legal status and focus of charter.

Looking across the surveys, several issues of questionnaire coverage emerge:

- The student questionnaire includes items obtainable elsewhere and may be duplicative. For example, student-reported information on classroom instructional approaches overlaps with information on the teacher questionnaire.
• Although the three surveys collectively cover a broad range of important background topics, the surveys omit a few topics with a strong base supporting their relationship to achievement. Two examples are the degree to which schools reach out to parents, and school discipline and the climate for learning.
• The questionnaires largely ignore major national policy issues prominent over the last decade involving the response to federal mandates for state-based student testing and high-stakes accountability.

The panel believes there is a need to address these and other issues of questionnaire content through a systematic process for identifying topics and questions that best relate to understanding NAEP student achievement results without being excessively burdensome or invasive.

**Recommendation 1a. Continually review and refine a core and second-tier set of background topics and questions that are common across NAEP surveys.**

- NAEP should build on its current process for specifying a common core set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.
- NAEP should develop a second tier of priority background question clusters that could be rotated across assessment cycles. Important topics that might be explored include school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time.
- NAEP should prioritize core and second tier items in terms of information value and respondent time, select high-priority items, and eliminate current low-priority items.
- NAEP should regularly publish its background questionnaires and provide justifications for all questions asked in terms of research and policy. Core and second-tier background questions should be identified.

**Discussion**

This recommendation would expand NAEP’s current set of core background questions focused primarily on the congressionally required student subgroups. The panel recommends including as an additional part of the core, a second cluster for instructional and other school learning opportunities. This cluster would allow examination of student learning environments by describing the curriculum, instructional approaches, and teacher qualifications. Many of these types of questions are now included in the teacher questionnaire and would be folded into this category.

A third core cluster of core questions is recommended to cover the area of student motivation and control over the environment. Measures such as whether students believe that success depends more on ability than effort or students’ locus of control have been
documented over several decades as strongly related to academic performance (Coleman, 1966; Chen & Stevenson, 1995). Also, students’ educational expectations predict their educational achievement and occupational expectations predict occupational attainment (ETS, 2010). When good teachers and a positive school environment influence student motivation and expectations this in turn will lead to improved achievement.

A second tier set of question clusters is proposed to focus on items for which there is strong research backing of their relation to achievement, but for which rotated items across alternate assessments (e.g., every four years) would be acceptable. As noted above, these second tier clusters could describe school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time. Specific clusters should vary across time as achievement levels and educational practices and policies change.

Together these clusters of items would view gains in school achievement as driven by a simple theory that sees gains in learning as a function of the curriculum, learning time, quality of instruction and student motivation. These core and second-tier clusters meet the principle in the Board’s Background Information Framework that “The information obtained be of value in understanding academic performance and taking steps to improve it” (2003 Background Information Framework).

The Panel recognizes that in defining these clusters NAEP will have to establish tradeoffs in terms of meeting the constraints of questionnaire length and cost. These decisions should be based on the priority of a question or question cluster in terms of information value balanced against respondent burden and costs. To make room for new high-priority items, NAEP should consider eliminating or reducing low-value or duplicative questions as noted below. Time constraints may also be addressed by rotating questions on alternate survey administrations (i.e., four-year intervals). NAEP also constrains the student questionnaire length to ten minutes when TIMSS even at grade 4 is 30 minutes.

**Recommendation 1b. Extend NAEP background questions to inform topics of current policy interest.**

- Implementation of this recommendation could focus on three rotating sets of policy questions each extended over a six-year period. For example, the initial set might start with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations would be added, and two years after that a module on project-based or online learning. After six-years, questions on a new policy issue would be introduced to replace the first. Using this approach each of the question sets would have three observations over the six-year time.

- The panel concurs with the 2003 Background Report caution to include only policy-relevant questions that are answered on the basis of fact rather than opinion. That is, the responses to policy-relevant questions should be objective and not reflect personal beliefs. Questions should ask about policy responses, such as training received to understand new standards or the extent to which new
standards have changed instructional content or approaches. Questions should not elicit judgments about personal policy preferences.

- The policy information collected should not duplicate what can be obtained from other sources, such as description of the law or state implementation plans. Instead, NAEP is uniquely positioned to obtain ground-level information by surveying teachers and principals about policy implementation and challenges. This would not be designed nor suited to address legal compliance with federal policy, which is the role of program monitoring. Instead, it would provide information to improve the quality of policy and practice.

- Indeed, many national policies such as the Common Core are not federal at all. In this example, NAEP would track the implementation of standards in the Common-Core states, identifying changes in instructional content and emphasis compared with non-Common cores states. NAEP teacher surveys could further address the extent of staff training and understanding of the new standards and instructional challenges.

Discussion

The panel’s review of the current background questionnaires concluded that they insufficiently incorporate questions about school and teacher responses to policies that could strengthen policy implementation and promote student achievement. Examples of policy-relevant issues that NAEP could but currently does not report on include characteristics of instruction in schools that made adequate yearly progress, the degree to which teacher evaluations incorporate student outcomes, or the nature and extent of coordination between school and after-school instruction.

This recommendation would reinforce NAGB (2003) guidance that identifies “informing educational policy” as a reason for collecting non-cognitive information. It would also support NCES commitments to convening “a policy/contextual issues panel when needed to identify policy/contextual issues that NAEP might address in the future, and to outline the relevant constructs and identify data needed to address these issues.”

The panel recognizes that policy issues should be regularly refreshed as new policies emerge that build on or replace prior strategies. Our proposal aims for roughly a six-year issue cycle to give policies sufficient time to be implemented and effect improvements. The three policies suggested in the recommendations reflect the likely timeframe of implementation. The initial focus is on Common Core implementation, which is already underway in many states. Next a question set would be added on how schools evaluate their teachers. This would include questions on how evaluations of teachers take into consideration the outcomes of a teacher’s students, as this relatively new policy takes

---

hold. The third suggestion of project-based and online learning reflects expectations that the role of technology in providing instruction will substantially increase.

**Recommendation 1c. Select clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics.**

- Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement. NAEP unlike TIMSS or PISA currently fails to publish clearly defined, research-based theoretical frameworks that guide question selection. Accordingly, NAEP should make explicit and publically available the underlying theoretical frameworks for question selection. The Panel recognizes that the research basis for the theoretical justifications may be less than perfect and are sometimes subject to post-hoc rationalizations. Nonetheless, the objective syntheses of research across a variety of settings to form theoretical frameworks for clusters of variables significantly enhances the odds of collecting survey information that will accurately and usefully inform practice and policy.
- Background questions should fit together to portray different important aspects of a topic (e.g., the different dimensions of SES).

**Discussion**

The 2003 *Background Information Framework* for NAEP states the principle that “Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be “directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information.” NAEP to its credit employs panels involving contractors and multiple external groups in its question development.

However, currently, NAEP does not formally publish an accompanying document with each assessment that lays out the theoretically-based frameworks that underlie the selection of the background questions and their connection with learning and achievement.

NCES has a good start toward building the necessary research foundation for developing such frameworks in the papers prepared by the Education Testing Service (ETS). ETS (2010) has developed three in-depth literature reviews, one each to support the topics currently or potentially addressed in the student, teacher and school questionnaires. The student and school questionnaire reviews also compare the current NAEP content items with the content measured in other large-scale national and international assessments.

The panel’s proposal would build-on the current literature reviews by:
• Using the research to develop theoretical frameworks that identify for major topics the component variables around which to build clusters of questions. The current ETS literature reviews although useful, are largely a description of discrete findings. Exhibit 3 is an example of how PISA presents a research-based, theoretical framework to organize background questions around the components of student engagement in reading and reading strategies. In this example, PISA operationalizes engagement in reading in terms of five components: reading for school, enjoyment of reading, time spent reading for enjoyment, diversity of reading materials, and diversity of online reading activities. Multiple questions then ask students about their reading behaviors with respect to each component.

• Organizing literature reviews around topics, which is preferable to the current organization around three separate questionnaires. Some topics may cut across the student, teacher and school questionnaires. For example, the current ETS literature review considers family involvement only in terms of the student questionnaire and the items describing home learning activities and resources. A broader research-based theoretical framework around the issue of parental involvement would extend the construct to include how teachers and schools reach out and support families, not just what families do by themselves. Indeed, Title I longitudinal evaluations have shown that student achievement improves when schools reach out and support parental involvement. (USED, 2001).

Once developed, these research-based frameworks would form the basis for developing valid and reliable questions to measure the different aspects of a topic domestically and to coordinate measurement with major international surveys. (Section 2 below).

Exhibit 3. PISA Analytic Framework for Student Engagement in Reading and Learning Strategies to Inform Decisions about Improving Reading

<table>
<thead>
<tr>
<th>How does PISA define “engagement in reading activities”?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading for school</td>
</tr>
<tr>
<td>Enjoyment of reading</td>
</tr>
<tr>
<td>Time spent reading for enjoyment</td>
</tr>
<tr>
<td>Diversity of online reading activities</td>
</tr>
<tr>
<td>Diversity of reading materials</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How does PISA define “learning strategies”?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding and remembering</td>
</tr>
<tr>
<td>Memorisation strategies</td>
</tr>
<tr>
<td>Control strategies</td>
</tr>
<tr>
<td>Summarising</td>
</tr>
<tr>
<td>Elaboration strategies</td>
</tr>
</tbody>
</table>

Source. OECD, PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices
Recommendation 1d. Use consistency over time as a criterion to consider for question selection and wording.

- NAEP’s inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.
- Recognizing that NAEP needs to periodically refresh its question set, nonetheless NAEP question selection seems haphazard – important questions may not be asked for two or more assessments and then they may reappear with changed wording that disrupt the time series reporting.
- Rather than total eliminating some potentially important survey questions on a topic, NAEP should consider rotating questions so that a question may be asked only once every 4-6 years.
- When rewording is necessary, NAEP should do bridge studies to link the new question responses with prior ones to form an unbroken time series of responses.

Discussion

The opportunity to assess progress on a background indicator over time is lost when NAEP no longer asks a prior question or disrupts the time series by asking essentially the same question in a somewhat different way. Because NAEP is the only major regular state-by-state assessment, question disruption results in a loss of important information to understand changes in a state educational context.

The panel examined the extent to which time series are available on the background question items for a sample of five broad questionnaire categories (Exhibit 4). The examination computed the percentage of questions asked under each category on the 2011 questionnaire for which there was also information for the same question for 2005 or earlier (at least a six-year trend).

- Between 70%-80% of the 2011 items about student characteristics or school demographics could be traced back to 2005 or earlier years.
- The three remaining categories that dealt with more judgmental measurement had much weaker time series availability. Only one-third of the 2011 questions asking about course offerings yielded at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

Some question categories become confusing to the user because of the considerable number of questions no longer asked. A case in point under the group of teacher factor questions is the “Preparation, Credentials and Experiences” category that contains over 400 questions of which more than 300 are no longer used, with many replaced by just slightly different wording. Moreover, what appears to be the exact same question maybe listed a number of times and in different places. Each instance of this all too common occurrence requires the user to search through and find all similar items and try and identify the one, if any, that is available and relevant.
Recognizing that at times changes in question wording may be necessary, the Panel recommends conducting bridge studies that would compare responses in the same year for prior and newly revised questions on a topic. NAEP’s 2004 assessments in math and reading conducted a bridge study to compare results from students randomly assigned to the original and revised versions of the assessment (NCES, 2004). Bridge studies were also conducted for the new frameworks in reading and 12th grade math that were introduced in 2009. A similar process could be developed to bridge question changes in important areas of the background questionnaires.

Strategies for holding down the added expense of bridge studies should be carefully explored. Recognize that in conducting a bridge study on background questions, smaller representative samples of the kind used for polling may be adequate and preferable in minimizing error to having no bridge study at all. Also, it may be feasible to add background questions to other bridge studies such as those employed for the assessment.

### Exhibit 4. Percent of Background Questions Asked in 2011 Which Were Also Asked in 2005 or Earlier For a Sample of Question Categories

<table>
<thead>
<tr>
<th>Question Category</th>
<th>Total Questions 2011</th>
<th>Total Number Asked in 2005 or Earlier</th>
<th>% of 2011 questions Asked in 2005 or Earlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Characteristics</td>
<td>10</td>
<td>8</td>
<td>80%</td>
</tr>
<tr>
<td>Curriculum</td>
<td>34</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Course Offerings</td>
<td>78</td>
<td>28</td>
<td>36%</td>
</tr>
<tr>
<td>School Demographics</td>
<td>18</td>
<td>13</td>
<td>72%</td>
</tr>
<tr>
<td>School Resources</td>
<td>43</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer

**Recommendation 1e. Delete duplicative or low-priority questions to make time for the Panel's higher priority items.**

- Several question groups on the student questionnaire are duplicative of information asked on the school or teacher survey. With the 10-minute limited time constraints on the student survey, these duplicative items should be reviewed for elimination and replaced by higher-priority items in the areas recommended by the panel.
- There seem to be an excessive number of background variables collected around a particular topic in some subjects.

**Discussion**

With the student questionnaire currently only 10 minutes long, each question must bring information value or be eliminated and replaced by a high-value item. The Panel has identified two item clusters as duplicative and candidates for elimination

- Student’s race/ethnicity asked on the student questionnaire is also obtainable from
Exhibit 5. NAEP’s 2011 Grade 8 Student Questionnaire Asks 8 Questions About Technology Use

1. How often do you use these different types of calculators in your math class? a) Basic four-function (addition, subtraction, multiplication, division) b) Scientific (not graphing) c) Graphing
2. When you take a math test or quiz, how often do you use a calculator? a) Never b) Sometimes c) Always
3. For each of the following activities, how often do you use a calculator? a) To check your work on math homework assignments; b) To calculate the answers to math homework problems; and c) To work in class on math lessons led by your teacher.
4. What kind of calculator do you usually use when you are not in math class? a) None; b) Basic four-function (addition, subtraction, multiplication, division); c) Scientific (not graphing); d) Graphing
5. How often do you use a computer for math at school?
6. Do you use a computer for math homework at home?
7. On a typical day, how much time do you spend doing work for math class on a computer? Include work you do in class and for homework.
8. When you are doing math for school or homework, how often do you use these different types of computer programs?
   a) A spreadsheet program for math class assignments;
   b) A program to practice or drill on math facts (addition, subtraction, multiplication, division).
   c) A program that presents new math lessons with problems to solve
   d) The Internet to learn things for math class
   e) A calculator program on the computer to solve or check problems for math class
   f) A graphing program on the computer to make charts or graphs for math class
   g) A statistical program to calculate patterns such as correlations or cross tabulations
   h) A word processing program to write papers for math class.
   i) A program to work with geometric shapes for math class

school records that represent the official record and
- Student information on classroom instructional approaches overlaps with information on the teacher questionnaire.

In addition to direct item duplication, inefficiencies in question selection come about through an imbalance of questions in an area that is disproportionate to its information importance. Exhibit 5 lists the sixteen questions about technology on the 2011 student questionnaire for the eighth grade assessment in mathematics. This is over one-quarter of the items and, while easily measurable, the level of detail may be hard to justify in terms of information value.

**Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for Background Questions.**

The panel urges attention to strengthening the validity, reliability and coordination of NAEP background questions.

A validity study of the NAEP background questions would assess whether they capture the concept NAEP intends the questions to measure. Concepts such as student socioeconomic status, student expectations, teacher qualifications, instructional content are challenging to define and quantify.
Two common approaches to assessing validity are:

1. **Construct validity** assesses whether the question or set of questions accurately captures the underlying construct being measured, which is often multidimensional. Socio-economic status is a multidimensional concept about family and community position in society that is incompletely captured by a discrete measure of poverty status—eligibility for a free or reduced-price school lunch.

2. **Concurrent and predictive validity** assesses whether the questions measuring a concept relate well at the same time or in the future with another established measure of that concept. The different aspects of family involvement that relate to current or future achievement meet the concurrent or predictive validity test.

A **reliable measure** yields consistent results over repeated measures. Asking teachers a question about frequency of a behavior in terms such as how much emphasis do you place on a subject is imprecise and subject to the subjective opinion and local norms. A more reliable question would ask do you teach this subject once a week, twice a week or very day.

Coordination among a set of questions maximizes information content. A duplicative question yields no added information content. Matching a NAEP set of questions with comparable questions on international assessments is highly efficient as it potentially adds considerable information content at little or no extra respondent burden.

The following recommendations suggest improvements to the validity, reliability and coordination of the NAEP background questions.

**Recommendation 2a. Improve the validity and reliability of the current measures NAEP uses for its mandated student reporting categories.**

- Support the current NAGB and NCES reviews of the best way to measure student socioeconomic status (SES). The known limitations of the current school lunch proxy and the likelihood that even this proxy will no longer be available make this review critically important.
- Assess the implications of changes in multi-racial student populations for the racial/ethnic student classification.
- Examine the accuracy of state-by-state or urban school system performance differences because of variation in the percentages of special education students receiving accommodations.

**Discussion**

*The panel supports the current NAGB and NCES reviews to identify the best way to measure SES variables within the confines of the NAEP questionnaire structure.*
This review is critically important given the well-documented limitations of the current school lunch proxy and that the first three State systems are piloting free school lunches for all students in very high-poverty school systems.

Limitations of the current school lunch measure include:

- The current measure divides the population only into two groups of free and reduced price school-lunch eligibles and ineligibles and is therefore insensitive to income differences above and below the income eligibility thresholds. SES is more accurately reflected by continuous measures. For example, this is consistent with studies showing student achievement results are sensitive to income levels over a broad income range.\(^4\)

- School lunch eligibility is known to be underreported in secondary schools. Secondary students may not want the stigma of making known their families low-income and secondary students may not eat lunch at school. In fact, the grade 12 NAEP did not include school lunch for its 2009 report because of the problems of underreporting.

- The lengthy research literature on measuring SES consistently recommends multidimensional SES indices (Hauser & Warren, 1997) involving family resources, education and occupation. However, NAEP only reports the single student school lunch eligibility measure. NAEP’s SES Project Progress Report (Noel-Miller and Hauser, August 2011) shows that a simple weighted average of indicators of home possessions and parental educational attainment does quite as well as independently estimated regression estimates in predicting math and reading achievement across grade-levels and race-ethnic subgroups.

- The 2010 Healthy, Hunger-Free Kids Act includes a “community eligibility” option, which would permit schools in high-poverty areas to provide free breakfast and lunch to all students without sending home individual paper applications for parents to submit income data. Three states have been selected for 2011-12 pilot eligibility (Illinois, Kentucky and Tennessee) and more states are scheduled to participate in successive years. Moreover, one urban school system Cleveland already counts 100 percent of its students as eligible for school lunch.

Consistent with the research literature, PISA incorporates questions for age 15 respondents to support an international multidimensional, socio-economic index. PISA’s SES index elements consist of: occupational status of the father or mother, whichever is higher; the level of education of the father or mother, whichever is higher, converted into

\(^4\) In data from the Early Childhood Longitudinal Study (ECLS) measuring kindergarten students achievement on the ECLS reading achievement assessment, low-income students scored at about the 30th percentile, middle-income students scored at about the 45th percentile, and upper-income students scores at about the 70th percentile.” (Lacour & Tissington, 2011)
years of schooling; and the index of home possessions, obtained by asking students whether they had a desk at which they studied at home, a room of their own, a quiet place to study, educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (e.g. paintings), books to help them with their school work, a dictionary, a dishwasher, a DVD player or VCR, three other country-specific items and the number of cellular phones, televisions, computers, cars and books at home.

The panel recommends that NAEP also move toward a multidimensional index for SES using current background questions. The panel further supports a long-run direction along the lines NCES is exploring of a two-pronged approach: (1) Creating an enhanced student background questionnaire with items that probe resources in the home, parents’ education level, and parents’ employment status; and (2) Using geocoding software to link students’ home addresses to aggregate SES data available from the United States Bureau of the Census. The geocoding would reflect neighborhood and community factors that influence student performance.

In this context, the panel strongly supports the current NCES pilot to “generate SES information from the Census American Community Survey (ACS) data using school catchment zones, and which would make the collection of students’ home address unnecessary for any assigned (non-choice) school.”

The Panel recommends assessing the potential implications of changes in multi-racial student populations for the valid measurement of the racial/ethnic student classification.

Starting in 2011 NAEP collected multi-racial data from school records and included it in the main subject-matter reports. In 2008, the U.S. Census (2011) reported the multiracial population at 7.0 million or 2.3% of the population. This number is for the full U.S. population and the percentage for the school age children would be expected to be higher to reflect the growing number of inter-racial families in the U.S. NAEP now collects these race / ethnicity data two ways – from school records and student reports. The student reports allow students to check more than one box within racial and ethnic categories. NAEP should compare the self-identified reports with the official school records.

Recommendation 2b. Enhance the validity of student responses at different grade levels.

- Assess whether the same construct (e.g., SES) is best measured by different and increasingly more valid items across grades 4, 8 and 12.

Discussion

A younger (grade 4) NAEP respondent is likely to have more difficulty accurately going

---

5 Quote from NCES Jan. 26, 2012 memo from Peggy Carr to Larry Feinberg.
through a typical question-answer process, which involves 4 steps: (1) understanding and interpreting the question being asked; (2) retrieving the relevant information from memory; (3) integrating this information into a summarized judgment; and (4) reporting this judgment by translating it to the format of the presented response scale (Borgers & Hox, 2000).

The Panel recognizes that NAEP questionnaire design already gives considerable attention to differences in the ability of students at different age groups to go through these four steps to respond accurately to background questions. Thus, NAEP dropped a question about parent’s education for grade 4 students because of research suggesting that responses from grade 4 students were less reliable than from older students. However, balanced against possible student response error is the loss of potentially useful information from eliminating questions. The Panel recommends NAEP explore the inclusion in the grade 4 questionnaires of questions that ask about mother’s and father’s highest education. The exploration should compare the error rates in estimating SES with and without the grade 4 parent education item.

The Panel also recommends that NAEP consider how the same construct (e.g., SES) can be measured by increasingly more valid and multi-dimensional clusters of items for students in upper grades.

**Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-learn skills including student learning behaviors, motivation and expectations.**

- Learning-to-learn skills refer to a cluster of personal qualities, habits and attitudes and include learning strategies, motivations and expectations. These soft-skills have shown a strong predictive relationship with math and reading achievement and workforce performance over decades (Coleman report, ETS paper on ECLS, NAEP, TIMSS and PISA). The Panel also notes that motivation and expectation questions are a regular component in major NCES national longitudinal surveys and international surveys at the primary and secondary level. However, developing questions that accurately measure non-cognitive skills through subjective responses to survey questions is challenging and should build on the considerable existing body of measurement in this area.

**Discussion**

To accurately measure some of the hard-to-measure concepts the Panel has recommended (1c above) that NAEP develop clusters of questions that collectively provide a good measure of different aspects of theoretically-based frameworks. Currently, the NAEP background questionnaire, especially the student questionnaire, is highly restricted by time constraints and does not contain the rich set of items needed to validly measure many learning attributes associated with student achievement.
Exhibit 6 provides an example of how PISA’s in-depth questioning draws out students’ approaches to understanding a particular type of text. In essence, the questionnaire creates more authentic learning situations from which to document students’ behaviors.

Recommendation 2d. Improve question reliability by replacing imprecise phrases such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week”.

Discussion

NAEP should ask questions involving frequency of behaviors or intensity of services in a form that elicits the most precise meaning to these terms. In this regard, some NAEP questions are not specific and the reliability of responses to these questions may be low.

The following illustrates two questions on the NAEP 2009 teacher questionnaire asking teachers about frequencies of time spent on science. Question a) asks about time spent on physical science in terms using categories such as “Little”, “Some” or “A lot” that could
mean quite different amounts of time depending on teacher norms. By contrast, question b) uses the preferred wording in which response times are expressed in clear distinct time intervals.

Question a): In this class, about how much time do you spend on physical science?
   Answers: None = 4%, Little = 9%, Some = 27%, A lot = 60%

Question b): About how much time in total do you spend with this class on science instruction in a typical week
   Answers: Less than 1 hour = 1%, 1-2.9 hours = 4%, 3-4.9 hours = 60%, 5-6.9 hours = 25% , 7 hours or more = 9%

NAEP should specify responses to questions about frequency and intensity in a specific quantifiable format wherever feasible.

**Recommendation 2e. Coordinate NAEP background questions with those asked on international or domestic surveys.**

- NAEP should explore framing its questions with as identical wording as feasible to similar questions found on international assessments.
- NAEP should examine the feasibility of NAEP coordinating with the NCES household survey to administer the household survey to families of students who participate in the NAEP subject assessments. This coordination between the two surveys would link the results of adults in the household survey with students’ NAEP assessment scores.

**Discussion**

In recent years NAEP cognitive assessment results have been linked internationally to place NAEP national and state disaggregated performance on an international TIMSS or PISA scale. NCES now is linking the 2011 grade 8 mathematics and science assessments of NAEP and TIMSS so international benchmarks can be reported on NAEP. Potentially, many of the responses to the background questions can also be compared with similar questions asked on international assessments. Examples include time spent on homework, after-school learning, taking algebra in the eighth grade, or teacher preparation to teach math or science.

To make valid international comparisons, NAEP needs to word its questions so that they are very similar or identical to the wording of the comparable questions on international surveys. Comparability of wording will only be achieved through careful question linking.

Exhibit 7 illustrates the potential payoffs that could occur from linking NAEP responses to those on an international assessment measuring with respect student time learning in regular school lessons and out-of-school lessons compared with high-scoring Japan and Korea.
Almost 30% of U.S. age-15 students spend less than 2 hr. in a math class per week compared with less than 10% of Japanese students and 5 percent of Korean students. Moreover, those students with the lowest scores receive the least math instructional help in-school.

Eighty percent of U.S. age-15 students spend no time learning math in formal afterschool instruction compared with only a quarter of Japanese or Korean students.

It would be valuable for individual states to be able to compare their students’ math instructional time in-school and out-of-school with those of the Asian performers, but NAEP collects very little information about learning time. For example, it asks only about number of days a week in math instruction and not about number of hours and there is no information about time spent in math or other subjects after school. Had NAEP spelled out a basic theoretical framework identifying clusters of questions about time measurement (recommendation 1c) NAEP might have been more likely to align its questions to compare states with the interesting PISA national results.

Recommendation 2f. Build on current NCES cognitive interview techniques by using cognitive laboratories, such as small standing panels, to field test questions to establish their validity and reliability.

Discussion

NCES conducted cognitive laboratory investigations of the responses of students and teachers to questions from the 1996 and 1998 background questionnaires (Levine, Huberman, and Buckner, 2002). Cognitive interviews are an approach “to assess how respondents comprehend survey items and what strategies they use to devise answers.”

The 1990’s studies identified a number of general types of item problems:
Behavioral frequency discrepancies. These items ask about how frequently a student or teacher engages in specific activities or practices. The average level of agreement between fourth grade students and their teachers on items that used a four-point rating scale was only 38 percent; for eighth grade students and their teachers, the level of agreement was still only 51 percent. Guessing would yield agreement of 25 percent.

Time frame discrepancies. Differences between teachers and students in the period over which behavior is estimated were common. Teachers would generally think about the current year and students about a very immediate near-term period. Also, when teachers were asked about the frequency of a behavior such as how often a particular science topic was taught, teacher’s responses applied to only when science is taught. Thus the response option, “Almost every day,” was explicitly interpreted as “Almost every day that science is taught.”

Comprehension discrepancies. Different respondents may interpret items differently. When teachers responded to a question about frequency of a behavior with “students in your class,” some teachers would answer about the typical student and others would respond if any one student exhibited that behavior.

List format discrepancies: Loss of context. On a long list of items, students or teachers might forget the context in which the question was asked. A student might interpret a question about school behavior such as reading and respond with their general reading behavior in or out of school.

NAEP also conducted a cognitive laboratory analysis of the Responses of fourth and eighth graders to questionnaire items and parental assessment (Levine, et.al. 2001).

The Panel believes that cognitive lab interviews are able to detect and prevent many survey design problems. Hence, it recommends that NAEP use cognitive labs more extensively with an accompanying small panel of adult (teacher/principal) and child respondents to validate and improve background questions. In addition, small-scale pilot studies should be used to assess the feasibility, reliability, and external validity of survey items. We recognize that this may increase costs but it would help make the overall NAEP a better source of information.

Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy.

Limitations of time and concerns over data burden severely constrain the depth of the student background questions. As a result, NAEP often lacks the richness in its background questions that would enable it to replicate the constructs such as those PISA creates from lengthy multiple items around different aspects of research-based
frameworks. To further extend the richness of its data sets, PISA also enhances its basic student and principal questionnaires with optional supplemental questionnaires. NAEP should consider expanding the depth of its questions through a variety of strategies including spiral sampling (currently already under consideration by NAEP), expanded questionnaire time and rotating background questions across samples.

**Recommendation 3a. Support NCES’s exploration of a spiral sample methodology to expand the scope of background questions, while assessing the possible loss in the representativeness of disaggregated data.**

- Spiraling questions so that no student takes the full set of background questions would allow NAEP to expand the scope of its background items. The current 10-minute limit for the student questionnaire severely constrains the current scope and depth of the student questionnaires. By contrast PISA is able to support richer construct development with its 30-minute student questionnaire.
- In assessing questionnaire spiraling, it is important to consider how it would reduce NAEP’s ability to provide statistically-accurate state-by-state or urban district information, especially if broken out for different student sub-groups.

**Discussion**

The Panel supports exploring the proposed spiral sampling of questionnaire items in order to implement improvements in student questionnaire scope and depth. As noted, one such improvement would be to enable greater in-depth questioning through clusters of items that measure different aspects of research-based topic frameworks.

However, the Panel urges NCES to quantify how item spiraling will reduce NAEP’s ability to disaggregate state or urban district responses for specific population groups. For example, will background questions be available in sufficient sample size for all population groups for which cognitive student achievement data are reported?

Illustrating this point is an analysis of whether a state has changed its grade-8 access of students to a course in algebra during the two-year interval between successive NAEP assessments. It turns out that Alabama raised the percentage of its students in schools offering grade-8 algebra by 6 percentage points during the two years and Arizona decreased it by 5-percentage points. These changes are sizeable for two years, yet neither change was statistically significant. A spiral sampling approach would further reduce the odds of obtaining statistical significance.

**Recommendation 3b. Consider other item-sampling reforms to obtain the needed questionnaire time including lengthening the student survey; establishing a 4-year interval between administration of some background questions; and pooling item responses across survey administrations.**

- The ten-minute target length for responses to the student questionnaire does not seem grounded in empirical experience and NAEP would do well to consider the
merits and feasibility of a lengthier questionnaire. TIMSS grade 4 and 8 student questionnaires are targeted for 30 minutes at each grade and do not appear to suffer from high non-response rates.\(^6\)

- Some background questions with slow-moving trends may be adequately monitored through repeating survey questions at four-year intervals.
- Pooling item responses across successive surveys may also be a permissible strategy to expand the sample provided that response changes are sufficiently slow moving.

**Discussion**

These sample reforms could expand the number of background items surveyed over a multi-year period, while maintaining accurate State-by-state reporting of background questions. However, each involves its own tradeoffs in terms of questionnaire time and the availability of items on any one survey. The panel requests that NCES examine and report to NAGB the comparative strengths and weaknesses of different approaches to expanding questionnaire items.

**Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of NAEP Background Questions.**

Rich responses to relevant background questions would mean little if NAEP continues its present practice of including very few findings from the background questionnaires in its reports. The main exception is the reporting of achievement by the congressionally required student subgroups. For other background information, the only recourse for a potential user to these data is to conduct one’s own analyses using the NAEP Data Explorer. As a practical matter, this is an option that only professional researchers (and few others) will have the time and skills to undertake.

This set of recommendations would bolster the analysis and reporting of the background questions by means of separate publications, online tables, and improvements to the Data Explorer. The recommendations also include a caution to not repeat the mistakes of the past by excessive reporting of causal interpretations of point in-time data.

---

\(^6\) *TIMSS 2011 Assessment Design* (p126) describes expected student testing time at grade 4 of 72 minutes for the student achievement booklet and 30 minutes for the student questionnaire. The grade-8 times are 90 minutes for the student achievement booklet and 30 minutes for the student questionnaire.
Recommendation 4a. Prepare special reports highlighting the background question findings.

- The special reports would provide interested readers with key findings derived from the background questions. These special reports could be prepared and released either with the achievement report or during the interval between assessment administrations. The Panel recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. A third report that explores benchmarking to find correlates of high-performing states and districts should also be considered.
- These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

Discussion

Special reports would provide access to the background questions in manageable-size documents that don’t overwhelm the reader. An example of a NAEP special report is *The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*, which is Part II of the National Indian Education Study of 2009. Part II complements the Part I report on NAEP assessment results for American Indian students by providing information about students, their families and communities, and their school experiences.

More generally TIMSS and PISA illustrate two approaches to developing topics for the special reports. TIMSS includes individual chapters organized around different questionnaire topics:
- Students’ Backgrounds and Attitudes Towards Science
- The Science Curriculum
- Teachers of Science
- Classroom Characteristics and Instruction
- School Contexts for Science Learning and Instruction

The 2009 PISA has published a series of special reports, synthesizing lessons learned to improve academic achievement:

- *Overcoming Social Background: Equity in Learning Opportunities and Outcomes* looks at how successful education systems moderate the impact of social background and immigrant status on student and school performance.

The Panel recommends that NAEP give priority to preparing two initial special reports using current data.

- The first report would focus on learning opportunities and conditions in school including examining characteristics of teachers, curriculum and instruction and the distribution of these characteristics among schools with students of various racial and socioeconomic concentrations.
- The second report would explore the characteristics of learning opportunities after-school and in the home, again comparing students from different economic and social backgrounds.

These reports would help inform future background variable data collections by identifying data of the greatest value in what currently is collected.

Other future NAEP reports could take advantage of NAEP’s special data collections. One might examine the characteristics of high-performing states or jurisdictions. Another would explore the extensive NAEP question sets on technology use in instruction.

**Recommendation 4b. Prepare an online compendium of key background indicators for States and participating urban districts.**

**Discussion**

The state-by-state or urban district compendium would take advantage of NAEP’s unique capacity to report a consistent series of state and urban district background data over time. The Panel heard an example of such a report incorporating NAEP data in the STEM area that is being prepared by the nonprofit organization Change the Equation.

Exhibit 8 illustrates for the 22 districts participating in the 2011 Trial Urban Assessments a hypothetical mock-up of background question responses focused around grade 8 and mathematics. A few findings from the urban district data in Exhibit 8 illustrate the potential value of indicator comparisons:

- The systems with the highest percentage of students absent 5 or more days were Detroit, Milwaukee, DC and Cleveland, which were also places with lower student scores.
- For grade 8 students taking algebra, the highest scoring districts of Austin and Charlotte had relatively low rates of absenteeism.

---

7 From Change the Equation, a non-profit, non-partisan coalition of more than 100 CEOs who are committed to bringing high-quality Science, Technology, Engineering, and Mathematics (STEM) learning to every U.S. child.
• Although urban school systems have somewhat higher rates of students participating in math at an afterschool tutoring or school program, only Atlanta had at least half the students avail themselves of afterschool assistance.
• Urban districts for the most part have above national-average percentages of staff teaching math with a major, minor or special emphasis in mathematics.
• Access to the Internet at home is widespread among urban areas making school support for learning at home more feasible than might be generally believed.

| Exhibit 8. Illustrative Table of Background Question Indicators With a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment |

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Grade 8 All Students</th>
<th>Grade 8 Students Eligible for National School Lunch</th>
<th>Grade 8 Students Absent 6 or more days last month</th>
<th>Grade 8 Students 6 or more hours of Math Per Week</th>
<th>Grade 8 Students 6 or more hours at Home</th>
<th>Grade 8 Does Math At An Afterschool Program</th>
<th>Grade 8 Does Math Through Alternative Certification</th>
<th>Grade 8 Teacher Has Math Major/Minor or Special Emphasis</th>
<th>Grade 8 Does Math Specialist At School</th>
<th>Grade 8 Assigned To Math Ability</th>
<th>Grade 8 All Students In Math Class</th>
<th>Grade 8 Computers Available to Teachers and Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>284</td>
<td>44</td>
<td>7</td>
<td>42</td>
<td>37</td>
<td>17</td>
<td>21</td>
<td>17</td>
<td>38</td>
<td>17</td>
<td>75</td>
<td>45</td>
</tr>
<tr>
<td>Albuquerque</td>
<td>275</td>
<td>60</td>
<td>8</td>
<td>37</td>
<td>36</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>33</td>
<td>32</td>
<td>60</td>
<td>19</td>
</tr>
<tr>
<td>Atlanta</td>
<td>266</td>
<td>82</td>
<td>6</td>
<td>27</td>
<td>75</td>
<td>38</td>
<td>57</td>
<td>57</td>
<td>58</td>
<td>31</td>
<td>59</td>
<td>27</td>
</tr>
<tr>
<td>Austin</td>
<td>287</td>
<td>59</td>
<td>6</td>
<td>22</td>
<td>61</td>
<td>27</td>
<td>33</td>
<td>42</td>
<td>27</td>
<td>56</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>Baltimore City</td>
<td>261</td>
<td>85</td>
<td>0</td>
<td>46</td>
<td>95</td>
<td>41</td>
<td>28</td>
<td>24</td>
<td>76</td>
<td>53</td>
<td>85</td>
<td>27</td>
</tr>
<tr>
<td>Boston</td>
<td>282</td>
<td>76</td>
<td>6</td>
<td>36</td>
<td>76</td>
<td>38</td>
<td>30</td>
<td>13</td>
<td>18</td>
<td>12</td>
<td>61</td>
<td>47</td>
</tr>
<tr>
<td>Colorado Springs</td>
<td>266</td>
<td>52</td>
<td>6</td>
<td>35</td>
<td>87</td>
<td>18</td>
<td>29</td>
<td>44</td>
<td>47</td>
<td>33</td>
<td>85</td>
<td>70</td>
</tr>
<tr>
<td>Chicago</td>
<td>270</td>
<td>84</td>
<td>4</td>
<td>32</td>
<td>67</td>
<td>25</td>
<td>37</td>
<td>33</td>
<td>84</td>
<td>20</td>
<td>45</td>
<td>63</td>
</tr>
<tr>
<td>Cleveland</td>
<td>258</td>
<td>100</td>
<td>11</td>
<td>29</td>
<td>89</td>
<td>33</td>
<td>28</td>
<td>68</td>
<td>58</td>
<td>14</td>
<td>51</td>
<td>44</td>
</tr>
<tr>
<td>Dallas</td>
<td>274</td>
<td>65</td>
<td>7</td>
<td>32</td>
<td>46</td>
<td>27</td>
<td>36</td>
<td>61</td>
<td>60</td>
<td>23</td>
<td>45</td>
<td>54</td>
</tr>
<tr>
<td>Detroit</td>
<td>246</td>
<td>79</td>
<td>17</td>
<td>24</td>
<td>81</td>
<td>46</td>
<td>37</td>
<td>11</td>
<td>83</td>
<td>76</td>
<td>18</td>
<td>65</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>255</td>
<td>70</td>
<td>12</td>
<td>36</td>
<td>65</td>
<td>29</td>
<td>36</td>
<td>57</td>
<td>86</td>
<td>46</td>
<td>53</td>
<td>20</td>
</tr>
<tr>
<td>Fresno</td>
<td>256</td>
<td>88</td>
<td>10</td>
<td>51</td>
<td>52</td>
<td>11</td>
<td>28</td>
<td>6</td>
<td>57</td>
<td>23</td>
<td>91</td>
<td>75</td>
</tr>
<tr>
<td>Hillsborough County</td>
<td>282</td>
<td>54</td>
<td>9</td>
<td>34</td>
<td>57</td>
<td>13</td>
<td>22</td>
<td>40</td>
<td>50</td>
<td>40</td>
<td>95</td>
<td>3</td>
</tr>
<tr>
<td>Houston</td>
<td>279</td>
<td>75</td>
<td>6</td>
<td>29</td>
<td>83</td>
<td>26</td>
<td>37</td>
<td>56</td>
<td>83</td>
<td>25</td>
<td>84</td>
<td>58</td>
</tr>
<tr>
<td>Jefferson County</td>
<td>274</td>
<td>80</td>
<td>7</td>
<td>48</td>
<td>68</td>
<td>14</td>
<td>20</td>
<td>21</td>
<td>34</td>
<td>17</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>261</td>
<td>82</td>
<td>6</td>
<td>67</td>
<td>44</td>
<td>40</td>
<td>27</td>
<td>59</td>
<td>87</td>
<td>37</td>
<td>75</td>
<td>52</td>
</tr>
<tr>
<td>Miami-Dade</td>
<td>272</td>
<td>72</td>
<td>8</td>
<td>43</td>
<td>68</td>
<td>47</td>
<td>26</td>
<td>38</td>
<td>72</td>
<td>29</td>
<td>90</td>
<td>13</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>254</td>
<td>81</td>
<td>13</td>
<td>30</td>
<td>76</td>
<td>43</td>
<td>31</td>
<td>37</td>
<td>74</td>
<td>52</td>
<td>29</td>
<td>66</td>
</tr>
<tr>
<td>New York City</td>
<td>272</td>
<td>87</td>
<td>10</td>
<td>35</td>
<td>63</td>
<td>26</td>
<td>35</td>
<td>35</td>
<td>85</td>
<td>36</td>
<td>60</td>
<td>63</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>285</td>
<td>85</td>
<td>10</td>
<td>54</td>
<td>96</td>
<td>27</td>
<td>27</td>
<td>24</td>
<td>54</td>
<td>37</td>
<td>50</td>
<td>78</td>
</tr>
<tr>
<td>San Diego</td>
<td>276</td>
<td>60</td>
<td>8</td>
<td>48</td>
<td>48</td>
<td>13</td>
<td>27</td>
<td>11</td>
<td>40</td>
<td>17</td>
<td>77</td>
<td>72</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer

An actual set of NAEP urban or state indicators should be carefully developed to include the most informative research-based responses and would summarize other subjects and grades.

The Panel also recommends considering a larger online compendium of national, state or urban background question results be prepared and structured to easily find questions of interest around a topic. The typical educator or policymaker, who would benefit from the findings contained in the background questions, lacks the time to understand and delve into the questionnaires through the NAEP Data Explorer.

To facilitate online access to prepared tables of questions, the user might be given options to select: (a) questions based on a Google-type question search (b) questions as they appear on the student, teacher or school questionnaires; or (c) questions grouped by topic.
and grade. Once the questions are selected, tables at the different system levels would be automatically generated and viewed.

**Recommendation 4c. NAEP’s reports should not indicate causal interpretations using the background questions. However, the NAEP data offer some unique advantages for generating relationships and hypotheses about factors that may be associated with performance and these findings should guide more rigorous in-depth follow-on analyses.**

First, NAEP’s performance reporting by subject, population group or jurisdiction is often the primary source of objective national performance data overtime. These data naturally raise questions about the underlying factors that produce the high and low performance. However, the Panel concludes, as have other NAGB panels before it, that NAEP should not publish causal interpretations of the factors determining performance differences based on the NAEP data.

Second, it is important to differentiate NAEP’s use of rigorous external research to identify, measure and report on background variables that support or work against achievement (Barton, 2002). In such instances, NAEP is not generating the findings from its cross-sectional data, but instead drawing upon an external evidentiary research base for the questions selected. Examples would be the degree to which lower income or lower performing students have access to at least equal levels among opportunity-to-learn variables such as certified teachers or instructional time. Another example would be to compare high and low performers on such factors as alignment of instruction with standards that are systemically related to achievement.

**Recommendation 4d: NAEP should encourage others to conduct exploratory studies of the NAEP background variables.**

- This may be through initiating small-grant competitions for researchers to analyze NAEP background-question data or by partnering or supporting others to conduct their own analyses of the background variables.
- These grants would provide funds for researchers to explore interesting and potentially policy-relevant topics and methodologies.
- The independent reports supported through the external grants could use the background question data to inform national education policy debates without any direct NAEP organizational involvement and oversight over the findings. The external grantees might also explore issues and topics where analysts might employ NAEP data to explore correlations or associations.
- There is precedent for NAEP to support mini-grant competitions of this kind.

**Discussion**

Other statistical agencies routinely support in-depth analyses of their statistical data. For example, the Bureau of Labor Statistics (BLS) has its own employment research and
program development staff to conduct original research using BLS data. The ASA/NSF/Research Fellow program is jointly supported by American Statistical Association and The National Science Foundation with participation of the U.S. Census Bureau, and the Bureau of Economic Analysis. This program jointly supports a Federal Statistics Fellowship program bringing academic researchers to work with statisticians and social scientists in the three federal agencies for up to one year.

NAEP should consider launching a similar program through small grants ($10,000-$50,000) competitively given to independently conduct research using NAEP data including the background questions. The focus of this research would be primarily on measurement and other statistical issues to improve the election and quality of the background variables.

The Panel also suggests that NAEP consider various strategies for encouraging and supporting outside researchers to conduct analyses of the NAEP data. NCES may want to work cooperatively with other organizations and foundations in these efforts. For example, NCES partially supported with foundations the widely cited research by Grissmer (2000) to analyze the state-level NAEP repeated time series achievement and background questions to examine the impact of systemic reform on improved achievement.

Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.

- NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that may be analyzed through software such as Excel.
- Extend the Data Explorer to facilitate the manipulation and analyses of the background questions by themselves without the achievement results. Extending software to build-in multivariate analyses should be considered.

Discussion

NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that is analyzable though software such as Excel. Each NAEP table and chart contains useful breakouts of the overall assessment and background data, which have been extracted and organized to focus on particular topics. Analysts and researchers may want to build off these tables to add more data series, conduct descriptive statistical analyses or pull apart and regroup the data to emphasize different points. Currently, NAEP offers no direct means to work off of the tables and charts in the reports other than to reenter the data by hand or to try and recreate them using the NAEP Data Explorer.

The Panel urges NAEP reporting to follow the lead of PISA by attaching a “statlink” to a downloadable excel file of the data in the table so that the user is able to access directly the data content without burdensome data reentry. Exhibit 9 shows how statlink was used to highlight the U.S. score compared with Singapore. The published PISA chart was
modified to highlight the gap between the U.S. compared with top performing Singapore in the performance of the bottom quarter of the most disadvantaged students (low SES) within each country who achieve in the top quarter on PISA.

The Panel further recommends that NAEP strengthen the Data Explorer to facilitate the manipulation and analyses of the background questions by themselves without the achievement results. Extending software to build-in multivariate analyses should be considered.

While the NAEP data explorer is a typically excellent and easy to use tool when analyzing achievement results, analysis of the non-cognitive background variables can be quite challenging even for data experts. Several problems occur:

- Finding the question of interest in the Data Explorer is made more difficult by not having an alphabetic listing of question topics. A direct link from a question in the published student, school or teacher questionnaire to that question in the Data Explorer would also be helpful.
- The Data Explorer is designed to use the background questions as categories by which to classify student achievement scores (e.g., by whether a student participates in school-lunch) and not to independently analyze the background question responses themselves.

The following is a real-world example of the challenges that arose in using the Data Explorer to compare how much time teachers in each state spend on math instruction at the fourth grade.
• Step 1. Find whether this question is available on the NAEP Data Explorer.
  - Unfortunately, the Data Explorer does not contain a question search tool to determine if this question is available.
  - Look for “time spent on math instruction” under the curriculum section and find an item for class time spent on different science categories (e.g., earth science), but not for mathematics.
  - Look for “time spent on math instruction” under the “course offerings” section of the Data Explorer and find a question about “4th grade instruction in math” that covers time spent in class, but the latest data are for 1996.
  - Don’t give up, and go to the “classroom management” section of the Data Explorer and find “the 2011 question of interest: Amount of time required for math instruction.” This works but why is the question under classroom management and why is time spent in instruction listed in three different places?

• Step 2. Go to the Data Explorer to print a table displaying the distribution of time each state spends on math instruction at different grades. Instead obtain a table (Exhibit 10) that distributes State assessment scores by time intervals, but does not display the frequencies of the time intervals themselves.

Exhibit 10. Normal Data Explorer Display That Uses Background Variables (Time Spent Per Week on Math) As Classifiers To Distribute Achievement

The problem is that Data Explorer has a default that assumes interest in the distribution of assessment findings and not in the distribution of the background variables. The override selection to obtain a straightforward table of the time distribution of math scores is through a little known and not easily found path under the statistics option under edit reports. This permits the user to deselect assessment as the dependent variable and replace with the percentages distribution of the background question (Exhibit 11). This option should be highlighted in the NAEP general instructions and in the edit reports screen that everyone sees.

Finally the Panel understands that that the Data Explorer once had a capability to conduct multivariate analyses, but that is was removed by the NCES Chief Statistician because of concern about potentially disclosing personally identifiable information about sampled students. The Panel understands this concern, but
requests NCES to review the decision to determine whether disclosure safeguards can be built into an online multivariate capability.

Exhibit 11. Desired NAEP Data Explorer Display That Presents The Distribution of Time Spent On Math Per Week By State

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Less than 3 hours</th>
<th>3-4.9 hours</th>
<th>5-6.9 hours</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Alabama</td>
<td>4 (1.1)</td>
<td>3 (1.2)</td>
<td>62 (3.3)</td>
<td>31 (3.0)</td>
</tr>
<tr>
<td></td>
<td>Alaska</td>
<td>3 (0.5)</td>
<td>0 (0.0)</td>
<td>59 (2.2)</td>
<td>21 (2.1)</td>
</tr>
<tr>
<td></td>
<td>Arizona</td>
<td>3 (1.1)</td>
<td>5 (1.1)</td>
<td>37 (1.5)</td>
<td>35 (1.3)</td>
</tr>
</tbody>
</table>

NOTE: Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.


5. Implementing the Panel Recommendations

The panel report identifies four areas for improving the usefulness and use of the NAEP Background Questionnaires with respect to question selection, measurement, sampling, and analyses and reporting.

The panel recognizes that the benefits of the recommendations in each area should be balanced against their cost in relation to other expenditures in NAEP’s annual budget of over $130 million. A decision on the merits of each item involves potential tradeoffs that are outside the panel’s mandate and expertise. In considering resource priorities, however, the panel concludes that even though the background variables have been underused in recent years, they could, for a relatively modest expenditure, become the means for greatly increasing the usefulness and impact of NAEP. The panel therefore urges that its recommendations be implemented through:

- Producing *special reports* on the background data that analyze the considerable quantity of data already collected but largely unreported and unanalyzed.
- Moving quickly to initiate a long-term effort to improve the relevance, quality, coherence and usefulness of a *core and rotated set of background variables while implementing recommended improvements for measurement accuracy and sampling efficiency*.
- Further improving the *usability of the Data Explorer and other NCES online tools*, which are already of high quality.
Recommendation 5a. Exploit existing background data through special reports focused on issues and topics informed by background questions.

Discussion

The proposed special reports in 5a are designed to mine the unexploited investment in the largely unanalyzed background questions. These reports might be modeled on the special publication of background data from the National Indian Education Study of 2009, *Part II: The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*, cited in Recommendation 4a.

The special publications would describe:

- In-school learning opportunities and other educational experiences focusing on data already collected on curriculum, instruction, teachers and other school resources including technology.
- Out-of-school learning opportunities and other educational experiences including after-school and at home.
- The background characteristics of high performing states and school systems contrasted with low-performers. This benchmarking study would be purely descriptive, serving to guide follow-on research to improve understanding of the factors differentiating high and low performing states and districts.

These would be three synthesis reports, drawing on data from NAEP assessments across the curriculum and, where possible, trends over time.

Recommendation 5b. Initiate a set of activities to build clusters of core and second-tier questions around high-priority topics for the 2015 NAEP administration.

Discussion

Given the long lead times for questionnaire development, this effort needs to begin immediately in order to affect the 2015 NAEP reading and mathematics administration. The revised questionnaires would refocus the background questions to identify an expanded first-tier core and second-tier set of rotated question clusters, including a rotated set of policy issues (Strategies 1 and 2, Exhibit 12). As NAEP redefines its question sets, NAEP would improve measures through published evaluations of their validity, reliability and consistency with each major assessment (Strategy 3, Exhibit 12). To find the questionnaire time to develop in-depth question sets, Strategy 4 prepares a NAEP analysis and report on a combination of sampling reforms addressing spiraling questions and extra question time.
Exhibit 12. Longer-term Background Question Activities / Products

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Recom-</th>
<th>Activities/Products</th>
</tr>
</thead>
</table>
| Select core and rotated clusters of questions around research-based theoretical frameworks | 1a, 1c  | • Identify 1st tier core clusters (student sub-groups student learning opportunities, student motivation)  
• Identify 2nd tier rotated questions  
• Publish background questions with research-based justifications for question clusters |
| 2. Extend NAEP Background Questionnaires to monitor topics of current policy interest | 1b      | • Identify current and future policy issues that are suited for NAEP Background Question (Common Core, Teacher evaluation, online instruction.  
• Propose rotating cycle of 3 major policy areas beginning with 2013 assessment. |
| 3. Launch a process for the continual examination of the validity, reliability, efficiency, and consistency of measures | 2a, 2b, 2c 1d, 2f  | • Report on validity & reliability of SES & responses at different age levels  
• Implement quality review procedures for reliability and consistency of questions.  
• Launch a cognitive laboratory capability with possibly an available small standing supplementary panel. |
| 4. Report on item sampling reforms to incorporate extended question sets and topics including eliminating duplicative and low-priority items | 3a, 3b  | • Report on a strategy to add questions for cluster analyses and policy issues through questionnaire spiraling, alternating questions across assessment administrations, adding extra questionnaire time and eliminating low-priority items, |

Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.

Discussion

While the Data Explorer is an excellent tool for online access of NAEP achievement data, addressing weaknesses in the analyses and display of the background data in the Data Explorer and publications would extend the usefulness of NAEP’s current online tools.

- Simplify and clarify how to use the Data Explorer to analyze the distribution of responses on background questions.
- Explore the potential for conducting multivariate analyses through the Data Explorer
- Build links that allow the data in tables and charts in NAEP publications to transfer to excel spreadsheets for further analyses.
Recommendation 5d. Promote implementation by creating a single Governing Board committee responsible for all background questions; provide adequate resource support, while ensuring efficient resource use; and publicize background question products and findings.

Discussion

To promote implementation of the background question recommendations and make sure change occurs, the panel suggests that NAGB establish a separate standing committee to review all background questions and oversee a multi-year development plan to improve the questions and their use. Currently, the Board’s responsibilities for the background questions are divided between the Assessment Development and the Reporting and Dissemination Committees. A unified standing committee should regularly monitor and report on implementation of the panel’s recommendations by NCES and Governing Board staff.

The panel further recommends that a review be conducted of the resources needed in terms of time, money and personnel to implement the recommendations in this report. One approach to the problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate lower-priority activities, such as the duplicative collection of racial data and the disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

In conclusion, the NAEP background questions are a unique national information resource. The Governing Board and NCES have a responsibility to develop this resource to better understand academic achievement and the contexts in which it occurs and, hopefully, to help spur educational improvement.
REFERENCES


D.C. Available January 2012 online:


Noel-Miller and Hauser -- NAEP-SES Project Progress Report (August 2011)


TMSS 2011 assessment design. Available January 2012 online:


39
MEMBERSHIP AND AFFILIATIONS

- Marshall S. Smith (Chair)
  Visiting Scholar, Carnegie Foundation for the Advancement of Teaching
  Former U.S. Under Secretary of Education
  Palo Alto, CA

- Naomi Chudowsky
  Co-Director
  Caldera Research, LLC
  Bend, OR

- Alan Ginsburg (Secretary)
  Education Consultant
  Former Director of Policy and Program Evaluation
  U.S. Department of Education
  Washington, DC

- Robert Hauser
  Executive Director
  Division of Behavioral and Social Sciences and Education
  National Research Council
  Washington, DC

- Jennifer Jennings
  Assistant Professor of Sociology
  New York University
  New York, NY

- Sharon Lewis
  Research Director
  Council of the Great City Schools
  Washington, DC
NCES envisions the next decade as both vital to maintaining a quality assessment and a transition to the future. There are numerous environmental factors that will influence NAEP over the next decade, and beyond. Examples include:

- transition from a paper-and-pencil to technology-based assessment,
- the emerging interest in international comparisons for states and districts,
- advances in test development and new item types,
- advances in the cognitive sciences,
- development of measures of preparedness for post-secondary education and work,
- adoption of common content standards by most states that may or may not align with NAEP content frameworks,
- development of two separate multi-state-developed assessments based on the common core standards,
- inclusion of more contextual information in reports,
- growth of virtual schools and out-of-school and technology-based learning by students, and,
- budget reductions throughout the education and assessment enterprises.

In late 2010, NCES began a strategic planning process to provide guidance for making fundamental changes to the technical and operational components of NAEP. While the planning will be an integral part of the program over the next five years, the first phase is almost complete. Dr. Ed Haertel, former NAGB member, will summarize discussions from two culminating events wherein NCES received input from experts in a broad array of disciplines, and state and district policy, assessment, and curriculum areas.

Dr. Haertel is working with a panel of experts to prepare a draft report by March 31, 2012. A list of panel members is attached. Dr. Haertel’s short bio appears on page 3 of this tab.
Panel Members

Edward Haertel, chair - Stanford University
Russell Beauregard - Intel
Jere Confrey - North Carolina State University
Louis Gomez - University of California, Los Angeles
Brian Gong - National Center for the Improvement of Educational Assessment
Andrew Ho - Harvard University
Paul Horwitz - Concord Consortium
Brian Junker - Carnegie Mellon University
Roy Pea - Stanford University
Bob Rothman - Alliance for Excellent Education
Lorrie Shepard - University of Colorado at Boulder
Edward Haertel


Haertel has served on numerous state and national advisory committees related to educational testing, assessment, and evaluation. He chairs the Technical Advisory Committee concerned with California's school accountability system, chairs the National Research Council's Board on Testing and Assessment (BOTA), and from 2000 to 2003 chaired the Committee on Standards, Design, and Methodology of the National Assessment Governing Board.

Haertel is a Fellow of the American Psychological Association and of the American Educational Research Association. He is a past president of the National Council on Measurement in Education (NCME) and is currently Vice President of the National Academy of Education. Haertel is a recipient of the California Educational Research Association's Lifetime Achievement Award and of the NCME Award for Career Contributions to Educational Measurement.
National Assessment Governing Board
Nominations Committee

March 3, 2012

7:30 – 8:15 am

AGENDA

Closed Session  7:30 – 8:15 am

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 – 7:35 am</td>
<td>Welcome, Introductions, and Agenda Overview</td>
</tr>
<tr>
<td></td>
<td><em>David Alukonis, Chair</em></td>
</tr>
<tr>
<td>7:35 – 8:15 am</td>
<td><strong>ACTION</strong></td>
</tr>
<tr>
<td></td>
<td>Discussion of Finalists for Terms Beginning on October 1, 2012</td>
</tr>
<tr>
<td></td>
<td><em>Committee Members</em></td>
</tr>
</tbody>
</table>
“Making a Difference” Discussion

Introduction

This tab includes materials in support of the full Board “Making a Difference” discussion on Friday, March 2, 2012 and includes:

1. A brief overview of the “Making a Difference” initiative,
2. A summary of the December 2011 meeting discussions, and
3. A list of the various Board discussions held on this topic.

Other March 2012 “Making a Difference” materials:

- The minutes of the December 2011 meeting contain a full summary of the presentations and discussions which are summarized in this tab.

- The tab for the Ad Hoc Committee on NAEP Parent Engagement includes their draft report and recommendations. Their recommendations identify strategies the Governing Board and NAEP can use with parent groups to foster a sense of urgency and action about the need to improve achievement and reduce achievement gaps.

- The tab for the Reporting and Dissemination Committee includes, as Attachment B, brief reports on the status of Parent Outreach and Engagement, the Speaker’s Toolkit, and Focused Reports. It also includes a list of previously published (1994-98) topical NAEP reports, identifies planned focused reports, and suggests topics to be considered for future focused reports.

Please refer to these materials for additional details.

Overview of the “Making a Difference” Initiative

At the May 2010 meeting, Governing Board Chair David Driscoll challenged the Board to think about what the Board and NAEP can do to foster improved student achievement and close achievement gaps. He said the NAEP data suggest a need for urgent action so we must consider what the Board and NAEP can do. At almost every meeting since that date, the Board has discussed ideas and strategies the Board can implement or has reviewed the ideas and strategies to identify priority areas in which the Board should work. Not only has the Board collectively generated ideas for action, but individual Board members have provided proposals and outside groups have made recommendations for actions the Board can take.

An issue discussed throughout the Board’s deliberations has been the scope of the actions that can be taken while remaining consistent with the laws governing NAEP. In particular, there have been concerns about activities that might be considered “prohibited” under Public Law 107-279 Sec. 303(b)(4) Prohibited Activities. While the Board has not attempted to develop a
list of prohibited activities, the discussions remain mindful of this concern. The law describes the following two prohibitions:

(A) In General – The use of assessment items and data shall not be used by an agent of the federal government ... to rank, or compare, or otherwise evaluate individual students or teachers, or to provide rewards or sanctions for individual students, teachers, schools or local educational agencies is prohibited.

(B) Special Rule – Any assessment authorized ... shall not be used by an agent of the federal government ... to establish, require, or influence the standards, assessments, curriculum, including lesson plans, text books, or classroom materials, or instructional practices of States or local educational agencies.

As a result of stimulating remarks by Secretary of Education Arne Duncan at the November 2010 Board meeting, Chair Driscoll and Board member Tonya Miles began exploring what the Board could do with a focus on parents. The official establishment of the Ad Hoc Committee on Parent Engagement, chaired by Tony Miles, occurred at the March 2011 Board meeting along with the establishment of the committee’s mission: “to increase parent awareness about the urgency to improve levels of student achievement ... and reduce the size of achievement gaps” using NAEP data and resources. The Ad Hoc Committee will present its draft report at this March 2012 meeting.

Other meetings, as described in the last part of this section, have provided multiple opportunities to consider what the Board and NAEP can do to foster improved student achievement and close achievement gaps.

**December 2011 “Making a Difference” Discussion Summary**

Several sessions at the December 2011 meeting served to stimulate Board member thinking about how the Board and NAEP can make a difference.

**Secretary of Education Arne Duncan:** After swearing in the new and reappointed Board members, Secretary Duncan remarked that he enjoys the current popular debate on education policy but he is of the viewpoint that two real challenges are complacency and accountability. He emphasized that stakeholders need to continue to develop solutions on how better to serve the nation’s students. He also emphasized that parents need to be challenged and they need to demand accountability in order to seek better outcomes for their children.

**Kati Haycock and Ricki Price-Baugh:** Kati Haycock, President, Education Trust and Ricki Price-Baugh, Director of Academic Achievement, Council of the Great City Schools, spoke to the issue of how NAEP can be used to foster improved student achievement and close achievement gaps.

**Ms. Haycock** made the following suggestions for the Board:

1. Work with state officials to use NAEP data in public conversations about what to expect with the results on the new Common Core assessments.

2. Highlight the importance of state level results as a reliable indicator of student performance until the Common Core is fully implemented and tested for a few years.
3. Use NAEP data to confirm trends in student learning at the state and national levels.

4. Continue to use NAEP results to highlight performance and comparisons to international students.

5. Increase attention on NAEP assessments in subjects other than math, science and reading.

6. Continue to analyze data to highlight areas across the country where gains in student performance are being made, and identify what can be learned from those results.

**Ms. Price-Baugh** discussed a research project “Pieces of the Puzzle: Factors in the Improvement of Urban School Districts on NAEP” completed by the Council of the Great City Schools (CGCS) using NAEP trend data from the large city schools participating in the Trial Urban District Assessment (TUDA). She noted that several study findings (below) are very encouraging because they suggest steps that CGCS could take to accelerate progress in urban schools.

1. Large city schools made statistically significant gains in reading and math at both fourth and eighth grade levels. These gains were significantly greater than those of the nation, thus narrowing the gap between NAEP and the nation by statistically significant margins in both content areas.

2. Large cities made significant gains with all student groups in all subjects and grades except for Asian-American and Pacific Islander students.

3. Fourth and eighth graders performed better in reading for literary experience than reading for information.

4. Eighth grade students performed better in life science versus chemistry and physics.

5. Eighth grade students performed better in geometry and algebra than in number sense data and measurement.

Ms. Price-Baugh stated that the omission rates in the NAEP data revealed that students were not being taught to deal with questions that required analysis or argument based on complex text. The data also indicated that while a number of states have standards that are as rigorous as the NAEP frameworks, students were not being taught at that level. The results of the CGCS study serve as a strong indicator of where the Council will need to focus its work in preparation for the Common Core State Standards assessments. In addition, the study findings point to more detailed studies of student performance which can help identify underlying practices and determine improvements that can be made to instructional programs.

**Board Discussion:** The focus of the “Making a Difference” discussions at the December 2011 Governing Board meeting was to prioritize next steps, in particular, what could be done immediately or during the next year. At the direction of the Board Chair, staff developed nine proposals which were discussed in detail during the Committee meetings.

Vice Chair, Mary Frances Taymans, began the full Board discussion by reminding the Board about the objectives of the initiative.
• To address the urgency of the need to improve achievement for all students and close achievement gaps.
• To use the means available and appropriate for Board members individually and the Governing Board as a whole.

She requested that the discussion of these priorities focus on three important criteria as well as the constraints for implementing the activity. Ms. Taymans asked Board members to consider if the idea is within Board’s purview, the most effective approach, and how the idea could be conveyed in a way so others will act.

**Stage I – Committee Discussions:** The results of the Committee discussions were shared with the full Board on Saturday and are summarized in the chart which follows the Committee priorities listed below.

**Lou Fabrizio,** chair, Committee on Standards, Design and Methodology (COSDAM), indicated the Committee ranked the proposals as follows:
1. Topic #6 – NAEP Presentations for Parents
2. Topic #4 – NAEP Speaker’s Tool Kit & Resources
3. Topic #9 – Focused Reports and Studies

Further, Mr. Fabrizio said that during the Committee’s discussion an additional initiative was considered (as suggested by Kati Haycock), to have NAEP provide support to the states in preparation for reporting results from the Common Core State Standards.

**Eileen Weiser,** chair, Reporting and Dissemination Committee (R&D) reported that the Committee ranked the proposals as follows:
1. Topic #4 – NAEP Speaker’s Tool Kit & Resources
2. Topic #6 – NAEP Presentations for Parents
3. Topic #7 – Tell about TEL
4. Topic #9 – Focused Reports and Studies

**Alan Friedman,** chair, Assessment Development Committee (ADC) stated that ADC placed a high priority on:
1. Topic #5 – NAEP Resources for Teachers
2. Topic #4 – NAEP Speakers Tool Kit
3. Topic #9 – Focused Reports and Studies

A table summarizing the Committee priorities appears on the next page.
Making a Difference – Committee Priorities

<table>
<thead>
<tr>
<th>Item</th>
<th>ADC</th>
<th>R &amp; D</th>
<th>COSDAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Item a Day</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Quiz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Jeopardy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Speaker's Kit</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5-Teacher Res.</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-Parent Info.</td>
<td>Underway</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7-TEL</td>
<td>X 2nd tier</td>
<td>X later</td>
<td></td>
</tr>
<tr>
<td>8-NAEP Apps</td>
<td>X 2nd tier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-Focus Reports</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>10-Common Core/NAEP Info</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stage II – Full Board Discussion:

Vice Chair Mary Frances Taymans reviewed each Committee’s response to the initiatives and noted that topic #4 – NAEP Speaker’s Tool Kit and Resources and topic #9 – Focused Reports and Studies received the strongest support among the three committees. Then, Ms. Taymans opened the discussion to all Board members. The full Board discussion is captured in the December 2011 meeting minutes. The discussion resulted in the following suggestions.

Next Steps

- Find creative ways to use existing NAEP data.
- Determine if there is a need for new data collection on topics such as suspensions and expulsions which are not currently tracked at the national level.
- Explore the collection of information on alternative learning initiatives such as digital learning.
- Consider the possibility of merging the Speakers Toolkit with the Teacher Resource Kit.
- Use the same type of printed collateral documents when Board members have speaking engagements with various audiences.
- Consider setting up a speaker’s bureau to schedule speaking engagements for members and others who are knowledgeable about NAEP.
- Explain the similarities and differences between NAEP and the Common Core State Standards.
- Explore linkages to other assessments to determine students’ ability to apply science.
• Provide advance notice to the states on how well students might perform on Common Core State Standards assessments based on NAEP results. If the Board compares NAEP data to Common Core State Standards, it could inform changes in the curriculum, tools, and training.

• Take advantage of the unique position and opportunity to connect NAEP and the Common Core State Standards, particularly with the international assessments such as PISA, PIRLS, and TIMSS.

• Provide opportunities to debate the topics to allow for more participation on different points of view and levels of expertise. A lengthy exchange could help clarify what is possible, and allow more time for discussion on the limitations of reporting.

• Form partnerships, like those developed through the work on the Ad Hoc Committee on Parent Engagement, to create new avenues to spread the Board’s message.

• Explore NAEP’s role in showing meaningful correlations without causation, in the context of policymaking decisions. The Focused Reports initiative would provide the opportunity to delve into the substantive issues such as areas where students show strengths and weaknesses.

• Consider NAEP’s role in mining the wealth of data that is currently available and making it accessible and useful to the states.

• Hold a brainstorming session on the initiatives that are of the highest priority to the Board, and determine if it is possible to develop some strong statements supported by NAEP evidence.

• Have NCES staff compile a list of reports and activities that are currently available to provide context of what is possible in the future and share that information with the Board. (This is in the Reporting and Dissemination Tab.)

• Take advantage of the current conversations between teachers and principals about the Common Core State Standards and use the opportunity to make a connection with NAEP for greater visibility.

• Provide a high level overview of the assessment results.

• Revisit the initial charge to make sure the discussions stay focused and the exchange of ideas will lead to fulfilling the Chairman’s “Making a Difference” charge.

• Continue to think of the “Making a Difference” initiative as a process, which involves peeling back layers that will reveal options and have the Board reevaluate next steps at each stage.
Previous Board “Making a Difference” Discussions
(in chronological order)

May 2010 Generating ideas for the future of the National Assessment Governing Board: Summary is in the August 2010 Governing Board briefing materials.

Aug. 2010 Reviewing and prioritizing ideas for the future of the National Assessment Governing Board: Summary is in the November 2010 Governing Board briefing materials.

March 2011 Approving the appointment of the Ad Hoc Committee on Parent Engagement, and reviewing previous ideas and implementation status report: Summary is in the March 2011 Governing Board briefing materials.


NAEP 12th Grade Preparedness Research:
Analyses Relating Florida Students’ Performance on NAEP to Preparedness Indicators and Postsecondary Performance

Rebecca Moran, David Freund, and Andreas Oranje, ETS

As part of the National Assessment Governing Board’s efforts to enable NAEP to report on the preparedness of U.S. twelfth graders for postsecondary education or entry into job training programs, various studies were conducted to statistically relate performance on NAEP with results from other assessments that serve as indicators of preparedness for college entry, course placement, and entry into the workforce (National Assessment Governing Board, 2009). Both nationally-representative data, such as those used to establish a statistical link between NAEP and SAT, and data representative of students in individual states were of interest for such statistical relationship studies. The 2009 12th grade NAEP reading and mathematics assessments included a first-time pilot state assessment for 11 states; Florida was one of the participants. The Florida Department of Education (DOE) maintains a longitudinal data base (K-20 Education Data Warehouse) that includes college entrance and placement test scores and first-year college performance data for those students who attended public colleges in the state of Florida during the 2009-2010 academic year. Therefore, this Florida longitudinal data could provide information about postsecondary outcomes for many of the students who also participated in 12th grade NAEP.

This document describes the data and procedures used to evaluate Florida students’ performance in high school and first year of college relative to scores on the NAEP 12th grade reading and mathematics assessments and other test scores. This will be followed by a description of the analyses that serve as a follow-up to the statistical relationships established between NAEP and SAT at the national level. Of particular note are the analyses of postsecondary data to provide validity evidence for the potential preparedness reference points on the NAEP scales identified in the national statistical relationship studies.

Data

This study used data from Florida public school students who participated in the 2009 NAEP 12th grade reading or mathematics assessments, approximately 3,200 in math and 3,400 in reading. Analyses were conducted with the use of NAEP sampling weights to appropriately represent 12th grade public school students in Florida in 2009.

Matching NAEP and Florida DOE Data

The Governing Board entered into an agreement with the Florida DOE to obtain longitudinal data for public school students selected to participate in the 2009 NAEP 12th grade assessment. The process of matching data between the Florida database and NAEP participants was carried out in coordination with NAEP contractors, Westat and ETS, and the Florida DOE.

A critical requirement of the matching of student records was to protect students’ identity and maintain confidentiality. This was assured through the assignment of a unique pseudo ID for students sampled to participate in NAEP. At the time of sample selection of students for operational NAEP, Florida DOE staff appended the pseudo ID to files within the Florida DOE and transmitted the pseudo ID to Westat with other administration data. On all subsequent data files containing Florida data (e.g., ACT scores), only the pseudo ID was included on the files. The pseudo ID was used by Westat to match files from Florida back to the NAEP data.
files. Westat in turn provided files to ETS with the additional Florida data appended to NAEP student records. Throughout the process, ETS had no access to any Personally Identifiable Information (PII), such as names, birthdates, or social security numbers. This process was essentially identical to the matching process employed for the NAEP-SAT national linking study.

**Data Elements evaluated for use in the Florida Preparedness Research**

Of the variables available in the extensive Florida longitudinal dataset for the 2009 12th grade cohort, those examined for use in this research are described briefly below. Some of these data elements lacked sufficient power (i.e., small sample sizes in the linked set) and, therefore, are of limited value for extensive use in the current research.

1. **Florida Comprehensive Assessment Test** (FCAT) is Florida’s K-12 state assessment. Scores on the reading and mathematics tests from 3rd through 10th grades were available in the longitudinal dataset. Match rates were very high, with 10th grade scores in reading and math matched to approximately 94% of the NAEP test takers. However, concerns about the relevance of relating students’ 12th grade NAEP performance to FCAT scores earned two years earlier, while in 10th grade, precluded further analyses being pursued with the FCAT data.

2. **WorkKeys** is a job skills assessment system that helps employers select, hire, train, develop, and retain a high-performance workforce. WorkKeys includes three relevant tests: Applied Mathematics, Locating Information, and Reading for Information. Matched samples contained about 300 students (about 10% of the NAEP sample in each subject) for each WorkKeys test and, therefore, were inadequate for further analysis.

3. **Advanced Placement (AP)** college-level exams enable students to earn college credit and advanced placement in college courses. Approximately 36% of students in the NAEP samples took one or more AP tests. However, only 16% of the NAEP reading sample took a relevant reading AP exam (English or English Literature) and only 8% of the NAEP mathematics sample took a relevant math AP exam (calculus). The small sample sizes limited the efficacy of these data for further analysis.

4. **High School Program:** One of the background questions asked of students on the 12th grade NAEP assessment was, “Which of the following best describes your high school program?” Response options included (1) General, (2) Academic or college preparatory, and (3) Vocational or technical school. For the NAEP Florida sample, approximately 47% of students indicate their program was “general,” 43% indicated “academic or college preparatory,” and 9% indicated “vocational or technical school.” These data were examined in much greater detail and will be discussed in this memorandum.

5. **SAT and ACT College Entry Exams and ACCUPLACER College Placement Exam.** Approximately 43% of Florida’s NAEP sample took the SAT test; 47% took the ACT test, and 18% took the ACCUPLACER test. The match rates for SAT and ACT are relatively close to Florida SAT- and ACT test-taking rates and, therefore, the match was successful. This data is informative for NAEP preparedness research to the extent that College Board and ACT have developed preparedness benchmarks for their respective assessments.

6. **College Enrollment Status, First-Year Course-taking and Grade-point Average:** Data were available for students attending public colleges and universities in Florida for the 2009-2010 academic year. Approximately 54% of the students in Florida’s 2009 NAEP 12th grade sample attended a public postsecondary institution in Florida, with 36% attending community colleges and 17% attending four-year colleges and universities.
Analyses Conducted

As mentioned above, the purpose of this research activity was to explore the relationships between Florida students’ performance on the 12th grade NAEP assessments and various indicators of postsecondary preparedness to provide validity evidence for the potential preparedness reference points on the NAEP scales. These points were earlier identified by the national NAEP-SAT statistical relationship study. Table 1 provides potential reference points:

Table 1: Potential preparedness reference points based on NAEP-SAT statistical relationship study

<table>
<thead>
<tr>
<th>Statistical Projection</th>
<th>Percentage at or above 500 on SAT</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
<td>164</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>67%</td>
<td>169</td>
<td>313</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>175</td>
<td>325</td>
</tr>
</tbody>
</table>

Concordance

<table>
<thead>
<tr>
<th>SAT Subscore = 500</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>165</td>
<td>303</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NAEP “Proficient”</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>174</td>
<td>302</td>
<td></td>
</tr>
</tbody>
</table>

Thus, NAEP results were examined in relation to:

1. SAT performance, defined in terms of whether students met the SAT college readiness benchmark in each subject area (determined by College Board to be a score of 500 on the SAT critical reading and mathematics measures)
2. ACT performance; the benchmarks for college readiness established by ACT are 22 for math and 21 for reading
3. ACCUPLACER performance, where the cut scores used in Florida for placement into credit-bearing courses are 72 for math and 83 for reading
4. Students’ self-reported program of study in high school, whether vocational/technical, general, or academic/college preparatory. This information was collected as part of the NAEP student background questionnaire.
5. College enrollment: community college, four-year college/university, or unknown
6. First-year college course-taking: remedial or credit-bearing
7. First-year college grade point average: above or below a B-
Summary of Results

Relationships of NAEP reading and math performance with each of the seven indicators are provided in graphical form. Overall, the patterns of results observed for this cohort of Florida students did not contradict the potential preparedness reference points on the NAEP reading and mathematics scales identified through the national NAEP-SAT linking study.

Figure 1, which depicts the relationship between students’ self-reported program of study in high school and performance on the NAEP 12th grade mathematics assessment, is used as an example to familiarize the reader with the figures. These results will be discussed in more detail later on.

Figure 1: NAEP mathematics performance by self-reported program of high school study

Interpretation of Figures

- The vertical or y-axis represents the NAEP scale score distribution and is about 3 standard deviations (SD) in range to cover all results in this study while maintaining scale consistency across graphs. In other words, the mathematics SD is about 34 so the range of the y-axis is 100 points (100-200). The reading SD is slightly higher (37) and the range in the figures is 110 points (230-340).
- The number in each box is the average scale score for the designated group of students; the endpoints of lines above and below the box indicate the 75th and 25th percentiles of the score distribution, spanning the interquartile range (IQR). For example, the rightmost category (the mathematics score distribution for students in academic/college prep high school programs of study) shows that the average score is 159, the 25th percentile is about 138, and the 75th percentile is about 182, bringing the IQR to 44.
• The dark red horizontal line represents the NAEP *Proficient* cut-point, which is 174 for 12th grade mathematics.
• The light red shaded area covers the potential “preparedness” reference points obtained from the national NAEP-SAT linking analysis. The range for math is 164-175, as also presented in Table 1.
• The horizontal or x-axis includes each of the categories for the variable/measure; the percentage in the box below the category description indicates the estimated percentage of students in that category.

**Specific Results**

Figure 2 shows the relationship of NAEP with SAT. College Board has identified a score of 500 on the SAT critical reading and mathematics measures as college readiness benchmark scores (Wyatt, Kobin, Wiley, Camara, & Proestler, 2011). Based on this SAT benchmark, 53% of Florida’s 12th graders were deemed college ready for mathematics and 54% were for critical reading. Average NAEP scores for students who met the SAT preparedness benchmarks were near the NAEP *Proficient* cut scores and fell in or above the range of possible NAEP preparedness reference points. The average scores for these students were roughly one standard deviation higher than average NAEP scores for their peers who did not meet the SAT preparedness benchmark. In addition, the IQRs for these two groups were non-overlapping for mathematics and only marginally so for reading.

**Figure 2: NAEP performance by attainment of SAT college readiness benchmarks for mathematics and critical reading**
Similar to College Board, ACT has also established benchmarks for college readiness, which are an ACT score of 22 for mathematics and 21 for reading (ACT, 2010). Based on these benchmarks, Figure 3 indicates that 34% of Florida’s 12th graders were college-ready for mathematics and 46% were college-ready in critical reading. Average NAEP scores for students who met the ACT readiness benchmarks were near the NAEP Proficient cut scores and fell in or above the range of possible NAEP preparedness reference points. The average scores for these students were roughly one standard deviation higher than average NAEP scores for their peers who did not meet the ACT readiness benchmark. Similar to SAT results, IQRs for these groups were non- or marginally overlapping.

Figure 4 provides the relationship of NAEP with ACCUPLACER performance. For ACCUPLACER cut scores used in Florida for placement into credit-bearing courses are 72 for mathematics and 83 for reading. As is shown in the figure, 23% of Florida 12th graders who took ACCUPLACER achieved the benchmark in mathematics and 37% met the reading benchmark. The figure also shows that the placement cut scores do not align as well with the potential NAEP preparedness cut points from Table 1. This could be due to a weak relation between ACCUPLACER and NAEP, different standards and purposes, or a combination of both. However, the results certainly do not disconfirm the reasonableness of those preparedness cutpoints. In addition, it is important to note that ACCUPLACER results were available for only 18% of the students in the matched NAEP-Florida sample. Students in Florida may use ACT or SAT scores to qualify for entry into credit-bearing courses as alternatives to taking ACCUPLACER. Appendix A provides a more detailed treatment of matched sample sizes across the various indicators.
Figures 5a and 5b show the relationship between self-reported high school program of study and NAEP mathematics and reading performance, respectively. Of the Florida 12th graders in the NAEP mathematics sample, 42% indicated they were in academic/college prep curricula during high school. The average NAEP scale score for those students is 159. The 25th percentile estimate is 138; the 75th percentile is 182. Of the 42% of students in academic/college prep programs, about 31% were at or above the NAEP Proficient cut-point. Taking the range of potential preparedness reference points into consideration, the percent “prepared” ranged from about 31% to 45%.

In the NAEP reading sample, 43% indicated they were in academic/college prep curricula during high school. Based on the NAEP Proficient cutpoint, about 46% of Florida’s 12th graders in academic/college prep curriculums would be considered prepared for college in critical reading. The percent prepared is lower for students in both general programs and vocational/technical programs, although the relationship with NAEP is not as strong as for some of the other measures discussed earlier.
Figure 5a: NAEP mathematics performance by self-reported program of high school study

Figure 5b: NAEP reading performance by self-reported program of high school study
Figure 6a: NAEP mathematics performance by college enrollment

<table>
<thead>
<tr>
<th>College Enrollment</th>
<th>NAEP Mathematics Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>175</td>
</tr>
<tr>
<td>Missing</td>
<td>148</td>
</tr>
<tr>
<td>Community College</td>
<td>143</td>
</tr>
<tr>
<td>State University</td>
<td>162</td>
</tr>
</tbody>
</table>

Proportions:
- Total: 46%
- Missing: 36%
- Community College: 17%
- State University: <1%

Figure 6b: NAEP reading performance by college enrollment

<table>
<thead>
<tr>
<th>College Enrollment</th>
<th>NAEP Reading Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>306</td>
</tr>
<tr>
<td>Missing</td>
<td>283</td>
</tr>
<tr>
<td>Community College</td>
<td>274</td>
</tr>
<tr>
<td>State University</td>
<td>282</td>
</tr>
<tr>
<td>Both</td>
<td>306</td>
</tr>
</tbody>
</table>

Proportions:
- Total: 45%
- Missing: 37%
- Community College: 18%
- State University: <1%
Figures 6a and 6b depict the relationship of NAEP with college enrollment. It is important to note that the missing category contains a diverse group of students, including students who did not attend college, students who attended private schools in Florida, and students who attended colleges out of state. Average NAEP scores for students attending four-year state universities were close to the potential NAEP preparedness reference points, whereas average scores for their peers attending community colleges were below the reference points. Performance on NAEP of students in the missing category appeared relatively similar to those attending community colleges although, as expected, spanning a wider range of performances, particularly towards the lower end. The pattern was similar for both mathematics and reading. The rightmost data (for the category labeled “Both”) represents students taking courses in both community college and a state university. Note however that group is very small with less than 1% of students.

Figures 7a and 7b provide results for first-year college course taking, particularly whether students were enrolled in remedial courses. Similar to figures 6a and 6b, the missing category contains a host of situations ranging from Ivy League attendees to those not enrolled in college at all. The figures show that a substantially greater percentage of students taking no remedial courses during the first year of college scored at or above the NAEP preparedness reference points compared to students who required one or more remedial courses. The pattern was consistent for mathematics and reading and no particular pattern was found regarding the subject matter of the remedial course.

**Figure 7a: NAEP mathematics performance by first-year college course-taking: remedial or credit-bearing**

![Graph showing NAEP mathematics performance by first-year college course-taking: remedial or credit-bearing](image_url)
Figure 7b: NAEP reading performance by first-year college course-taking: remedial or credit-bearing

Figure 8a: NAEP mathematics performance by first-year college grade point average: above or below B-
Finally, NAEP performance was examined in relation to first-year college grade point average (GPA), which plays an important part of various preparedness benchmarks, particularly the GPA of B-. Figures 8a and 8b show that a greater percentage of students achieving GPA of B- or better during their first year of college scored at or above the NAEP preparedness reference points compared to students whose GPA was less than a B- during their first year of college. The pattern was consistent for mathematics and reading.

**Conclusion**

A number of indicators of postsecondary preparedness from the Florida Data Warehouse, such as college enrollment status and first-year college grade point averages, were studied in terms of their relationship to NAEP and, in particular, various potential benchmark points that were established based on national NAEP-SAT linking data. The results show relatively weak relationships with NAEP results, which would make these indicators relatively poor predictors with which to determine a preparedness benchmark on the NAEP scale. However, the purpose of this study was not to establish benchmarks, but to provide validity evidence (or disconfirming evidence) to the already-established benchmark region. These results generally indicate that this benchmark region is reasonable and certainly no evidence was found that would indicate that this region is unreasonable.

The limitations of the Florida data, namely the availability of data only for students enrolled in Florida public postsecondary institutions, must be taken into consideration when interpreting these results. Appendix A provides more details about the number and percentages of matched students that were used to support the findings in this memorandum.
References

ACT. (2010). *What Are ACT’s College Readiness Benchmarks?* Iowa City, IA: Author.

Appendix A

Figure A.1 provides sample sizes and percentages for the 2009 NAEP 12th grade Florida sample disaggregated by high school program, test-taking, college attendance, and remedial course-taking.

**Figure A.1: Sample sizes and Percentages for 2009 NAEP/Florida Grade 12 Math**
Ad Hoc Committee on NAEP Parent Engagement

Reaching Parents with NAEP Resources

March 2, 2012

Committee Members
Tonya Miles, Chair
Louis M. Fabrizio
Shannon Garrison
Doris R. Hicks
Hector Ibarra
Henry Kranendonk
Warren T. Smith
Blair Taylor

Staff
Ray Fields
Ad Hoc Committee on NAEP Parent Engagement

Overview of Recommendations

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

Committee Activity Timeline

November 2010  Recognize Need to Address NAEP Parent Engagement
March 2011  Approve Mission Statement and Establish Ad Hoc Committee on NAEP Parent Engagement
April 2011  First Ad Hoc Committee Teleconference
May 2011  First Committee Meeting
August 2011  Second Committee Meeting
October 2011  Second Teleconference
December 2011  Third Committee Meeting
February 2012  Third Teleconference
March 2012  Final Committee Meeting; Present Recommendations to the Board
Foreword

The National Assessment Governing Board, in overseeing the National Assessment of Educational Progress (NAEP or the Nation’s Report Card), is carrying out an initiative to raise public awareness about the status of student achievement in the United States.

The Governing Board believes that the low levels of student achievement and the persistent, large achievement gaps between student demographic subgroups are cause for alarm—for individuals, for families, for communities, and for the nation’s future.

Although the release of NAEP reports brings periodic public attention to this problem, this attention is not sustained for very long.

Consequently, the Governing Board is implementing an initiative to convey the urgency of improving achievement for all students and of closing achievement gaps between student subgroups by race, ethnicity and income levels, using NAEP data and resources.¹

One part of this initiative is aimed at reaching parents. In March 2011, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement, composed of Board members. The Ad Hoc Committee’s assignment was to study ways to reach parents with NAEP data and resources and to present the Committee’s recommendations to the Governing Board by March 2012.

The members of the Ad Hoc Committee have worked diligently over the past year and are pleased to present our report and recommendations on the following pages.

We would like to express appreciation for the important contributions of the National Center for Education Statistics in supporting the Ad Hoc Committee’s work and in embracing the objective of reaching more parents with NAEP data and resources. We also thank the Governing Board’s CCSSO² Policy Task Force members for their valuable comments and suggestions.

Tonya Miles
Chair
Ad Hoc Committee on NAEP Parent Engagement

¹ The authority for this initiative is found under the Governing Board’s duties in the NAEP legislation, Public Law 107-279. Specifically, Section 302(e)(1) authorizes the Board to “take appropriate actions needed to improve the form, content, use, and reporting of results” and “plan and execute the initial public release of National Assessment of Educational Progress reports.”

² The acronym CCSSO stands for Council of Chief State School Officers.
Introduction
The National Assessment Governing Board, recognizing that NAEP report releases were not conveying a sense of urgency, began an initiative in May 2010 to see what the Board could do to “make a difference” in fostering concern and action about the need to improve achievement and reduce achievement gaps, using NAEP data and resources. Toward this goal, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement. The Committee’s task was to develop recommendations on ways to reach parents with NAEP information. The purpose of this report is to document the work of the Ad Hoc Committee and present its recommendations.

Background
U.S. Secretary of Education Arne Duncan addressed the Governing Board on November 19, 2010. He focused on the urgent need to improve student achievement and reduce achievement gaps among student subgroups. He has said publicly that “our nation will pay the price socially and economically” if we fail to act with determination and dispatch and stressed to the Board that “we have to continue to awaken our country to the huge consequences” of inaction.

Secretary Duncan emphasized the important role of parents in improving student achievement. He told the story of President Obama meeting with the President of South Korea, Lee Myung-bak. President Obama asked him about education issues in South Korea. President Lee said his biggest challenge is that parents in South Korea are very assertive in demanding a good education from their schools and great effort from their children. He emphasized that this includes parents of all income levels.

Implicit in this story is the fact that South Korean students, as well as others in the world, outperform U.S. students in mathematics and science on TIMSS. Today’s students are tomorrow’s workers and leaders. It follows that failing to improve U.S. student achievement could have disastrous effects on the nation’s future work force and global competitiveness, and that parents have an important role to play in promoting improved student achievement.

Secretary Duncan continued by saying “I wish my biggest problem, my biggest challenge, was parents knocking down my door saying, ‘Get better faster!’” He said that there are good examples in the U.S. of parent initiatives that impact student achievement. But Secretary Duncan wanted to “scale up” parent engagement programs that “are really showing the ability to drive student achievement.”

---

3 The acronym TIMSS stands for the Trends in International Mathematics and Science Study.
The Secretary’s remarks and the Board’s initiative to make a difference served as the backdrop to Board member Tonya Miles asking what can the Board do to make NAEP data available to parents and guardians about student achievement, especially about the urgency of addressing achievement gaps by race, ethnicity, and income levels.

The question—“What can the Board do?”—is pertinent and important. Parents have a significant stake in the quality of their local schools and, most immediately, in their own children’s achievement.

Governing Board Chair David Driscoll recognized the opportunity and value of reaching parents with NAEP data. Therefore, at the conclusion of the November 2010 Governing Board meeting, he asked Ms. Miles, and she agreed, to lead a Board initiative to increase parent awareness about and access to NAEP data. The goal was to bring attention to the unacceptably low levels of student achievement in the U.S. and the disgraceful size of the achievement gaps.

**The Mission**

At the March 2011 Board meeting, the Executive Committee approved the mission statement for and established the Ad Hoc Committee on NAEP Parent Engagement (Appendix A). The Committee would be composed of Board members and chaired by Ms. Miles. The Ad Hoc Committee’s task was to present recommendations to the Governing Board by March 2012. The recommendations would describe steps and strategies the Governing Board and representatives of the NAEP program can take directly, and/or support the efforts of others to increase parent awareness about the urgency to improve the levels of student achievement in the U.S. and the urgency to reduce the size of achievement gaps by race, ethnicity, and income levels, using NAEP data and resources.

The mission statement indicated that the recommendations were to be clear about the limits on the Board’s role under the law. This was to ensure the Committee considered all appropriate options without exceeding the Board’s authority.

The recommendations were to help reach parents in feasible, innovative, and meaningful ways, across all income levels, and whether residing in urban, rural, or suburban areas. Finally, the recommendations were to include strategies to make NAEP parent engagement an ongoing part of the work of the Board and the NAEP program.

---

4 The term “parents” as used throughout this report is intended to refer to parents and guardians of school children.
Committee Activities
Members of the Ad Hoc Committee have met four times during the May 2011, August 2011, December 2011, and March 2012 Board meetings. The agendas for these meetings are in Appendix B.

The Ad Hoc Committee also has met three times in between Board meetings via conference calls:

- April 15, 2011 - reviewed the Committee’s mission statement and a timeline for completing their work (Appendix C)
- October 12, 2011 - focused on formulating the Committee’s preliminary recommendations for discussion at the December 2011 Board meeting (Appendix D)
- February 8, 2012 - reviewed the Committee’s initial draft report

In addition, Chair Tonya Miles and Ray Fields conducted meetings with leaders of three nationally recognized parent-related organizations. The purpose was to brief them on the Board’s initiative to reach parents with NAEP data, to receive their input and feedback, and to determine their interest in supporting this initiative. The three organizations are the National PTA, the Public Education Network, and the Center on School, Family, and Community Partnerships at Johns Hopkins University.

In connection with the August 2011 meeting in Washington, D.C., the Board conducted an outreach event with parent leaders and national and local parent organizations. The discussion with meeting participants, led by Ms. Miles and Governing Board Chair Driscoll, resulted in valuable feedback and input on the Board’s parent initiative. A summary of the discussion at this parent outreach meeting is in Appendix E.

Concluding Comment
Parents are the primary advocates for the quality of their children’s education. Having solid information about education achievement improves their ability to advocate and ask the right questions. NAEP can be one potentially valuable source of such information. Therefore, it is appropriate to seek ways to reach parents with NAEP data and resources. Some progress already has been made, in a small way, as will be seen in the activities and relationships described below. The recommendations that follow are offered as a set of feasible next steps, all within the Governing Board’s authority. All have the potential to reach parents in meaningful ways. Recommendations that the Governing Board decides to adopt should be assigned to appropriate Board committees and staff for implementation.
Recommendations

1. Specify the Target Audience: National, State, and Local Parent Leaders and Parent Organizations

The target audience needs to be defined. Approximately 55 million students are enrolled in public and private K-12 schools in the U.S. It is not feasible to reach the parents of all these children with NAEP data, nor is it within NAEP or the Governing Board’s scope to do so.

Further, the achievement of their own children is the most pressing and immediate interest of parents. Because NAEP does not provide individual student results, this interest of parents is not served by NAEP.

A unique aspect of NAEP is its ability to report patterns of overall and subgroup student performance within and across education systems. These patterns may reflect education system strengths and weaknesses that can affect the achievement of individual students. The NAEP data for the states and 21 urban districts5 provide ample evidence of differences in achievement across comparable groups at points in time and differences in gains in achievement over time. The NAEP data also document persistent and unacceptable achievement gaps between groups. This NAEP information does have potential interest for parents.

Also of potential interest to parents is how their education systems compare internationally. The linking studies the Board has endorsed, beginning in 2011, between NAEP and the international assessments (TIMSS and PIRLS6) will provide a way to compare student achievement at the state level in the U.S. with achievement in other nations.

State and local education policymakers use NAEP data to ask fundamental questions about the levels of student achievement in schools under their authority. For example, Tennessee Commissioner of Education Kevin Huffman discussed how he uses NAEP at a November 2011 meeting in Nashville on NAEP 12th grade academic preparedness. Commissioner Huffman said that he analyzes student subgroup results in his state (e.g., students on free and reduced lunch) in comparison to other states. Raising questions about how subgroup performance compares across jurisdictions can help highlight where state or local policies may or may not be working. Asking thoughtful questions about the implications of NAEP results can be a positive way for parents to begin a productive conversation with state education leaders seeking to improve

5 The 21 participants in the NAEP Trial Urban District Assessment Program are: Albuquerque, Atlanta, Austin, Baltimore City, Boston, Charlotte, Chicago, Cleveland, Dallas, Detroit, Fresno, Hillsborough County, Houston, Los Angeles, Louisville, Ky. (Jefferson County), Miami (Dade County), Milwaukee, New York City, Philadelphia, San Diego, and Washington, DC.

6 The acronym PIRLS stands for the Progress in International Reading Literacy Study.
achievement and close achievement gaps. Of course, while NAEP can be used as a source of information to help parents identify important questions to ask about the status of student achievement locally, the answers about what to do must be made by state and local officials with authority for the schools.

The Ad Hoc Committee believes there are groups of active parents and parent organizations who see the connection between system performance and the potential for impact on individual students. These include local and state leaders, often members of recognized parent and community organizations, who regularly work with the leaders of education systems, examine data, and ask fundamental questions to support and foster improved achievement and the closing of achievement gaps. These parent leaders and parent organizations should be the initial target audience for NAEP data and resources.

More specifically, because NAEP provides data for each of the 50 states and 21 urban districts, the initial target audience should be state and local parent leaders and parent organizations associated with these jurisdictions.

2. Establish Relationships with Recognized Parent and Community-based Organizations

To reach the target audience with NAEP data, it is important to work collaboratively with existing parent and community-based organizations. Many of these organizations have state affiliates and/or affiliates associated with local school districts. These organizations have direct access to parent and community leaders through email networks, social media, newsletters, and websites. These mechanisms are potentially effective, viable avenues for the dissemination of NAEP data and resources. In addition, these organizations often conduct national and state conferences, which could afford opportunities for presenting NAEP data and resources.

The Ad Hoc Committee has initiated conversations with the National PTA (NPTA), with positive results (see Appendix F). For example, the NPTA has begun announcing NAEP release events through its email networks and social media. In addition, Tonya Miles has been invited to make a presentation on March 7, 2012 at the NPTA Legislative Conference and on June 21, 2012 at the NPTA Annual Conference. Further, the NPTA assisted in recruiting parents for a meeting on February 16, 2012 to help review the NAEP presentation and materials for parents described in Recommendation 3.

Likewise, collaborative activity has occurred with the Public Education Network (PEN). Cornelia Orr, Governing Board Executive Director, made a presentation on NAEP and 12th grade academic preparedness at the PEN annual conference in November 2011. PEN also helped recruit experts from among its member organizations for a one-day meeting held on February 14,
2012 to provide input and feedback on the NAEP presentation and materials for parents described in Recommendation 3. PEN already transmits information about NAEP data and NAEP releases to its members and newsletter subscribers.

The Governing Board should continue to develop the relationships with the NPTA and PEN, and develop similar collaborative relationships with other organizations.

3. Develop Presentations and Materials Targeted to Parents for Use by Board Members and Others

Recognizing that the scope and depth of NAEP data and resources can be overwhelming, the Governing Board is working to develop a model PowerPoint presentation and associated materials for parents. Consistent with the information needs of the target audience in Recommendation 1, the presentation and materials can be customized for particular states and urban districts. The materials will include easy-to-understand charts and graphs and avoid the use of technical terms and jargon. In addition to explaining what NAEP is, the presentation will highlight NAEP data regarding the levels of achievement and the gaps between subgroups in ways that convey urgency.

The presentation and materials should be designed to help the audience understand the role of NAEP in the context of state and local assessments. Sample test items can be used to illustrate what content NAEP measures and how it is measured; consideration can be given to how this information about NAEP may complement state assessments. As noted in Recommendation 2, conducting input and feedback meetings with parent leaders and representatives of parent organizations is important to ensure that the level of detail and amount of information is appropriately tailored for the target audience.

The intent is for these resources to be available for use by Governing Board members invited to make presentations to the public and by interested parent and community-based organizations in making presentations specific to their locale.

The National Center for Education Statistics (NCES) is currently developing a general publication for parents. This publication will inform parents about what NAEP is, how it fits into the education landscape, and options to learn or do more. This publication will be debuted at the NPTA conference in June and displayed at the NAEP booth at the conference.
4. **Develop Parent Pages on the Governing Board and NAEP Websites**

Currently, the Governing Board website has no pages aimed at parents as the target audience. The NAEP website, managed by the National Center for Education Statistics (NCES) does have pages for parents whose child has been selected to take NAEP, but not for parents in general.

The Ad Hoc Committee invited NCES to examine what it can do to make NAEP information on the website more accessible to parents. As a first important step, NCES made the “parent” navigation button more prominent on the NAEP website landing page. NCES is exploring additional changes to make the NAEP data more accessible to parents. As they develop the parent publication mentioned in Recommendation 3, NCES will update the NAEP web pages to ensure consistency. This will help expand the NAEP website audience from just parents of students selected for the NAEP sample to all interested parents.

The Ad Hoc Committee asked the Board’s communications and website contractors, Reingold, Inc. and Quotient, to develop page mockups for parent pages on the Governing Board website (Appendix G). These should be further developed and incorporated as components of the Governing Board’s website redesign, which is currently underway. The model PowerPoint presentation and materials in Recommendation 3 should be available for easy downloading from the Governing Board website.

In addition, the Governing Board should seek ways to leverage mass communications (e.g., TV, radio, public service announcements, and social media) to reach parents with NAEP data and resources.

5. **Conduct a Parent Education Summit in Late Summer/Early Fall 2012**

The Ad Hoc Committee proposes a one-day parent summit on education for the late summer or early fall of 2012. The summit would be conducted in Washington, D.C. and available across the nation via live-streaming internet video, with the potential for live TV and radio coverage via C-SPAN.

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

In addition to Governing Board members, the audience of 150-300 would consist primarily of parent and community leaders, parent organizations, and leaders in education, business, civil rights, the religious community, and legislative policy.
To help convey the non-partisan, universal interest in achieving the summit objective, as well as to focus on its importance for the nation’s future, First Lady Michelle Obama and former First Lady Laura Bush would be invited to share the podium in delivering the keynote address.

A distinguished journalist or media representative, acknowledged for intellect and freedom from bias, would be invited to moderate and provide a concluding summary.

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

Individual and panel presentations would be made to address the national imperative for achieving the summit objective, from a wide range of perspectives which, taken together, would provide a compelling, unassailable argument for the urgent need to take action.

For example (not listed in priority order):

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

Edward Haertel, Panel Chair
Stanford University

New Orleans, LA
March 2, 2012
Panel Membership

- Edward Haertel (*Chair, Stanford University*)
- Russell Beauregard (*Intel Corporation*)
- Jere Confrey (*North Carolina State University*)
- Louis Gomez (*University of California, Los Angeles*)
- Brian Gong (*National Center for the Improvement of Educational Assessment*)
- Andrew Ho (*Harvard University*)
- Paul Horwitz (*Concord Consortium*)
- Brian Junker (*Carnegie Mellon University*)
- Roy Pea (*Stanford University*)
- Bob Rothman (*Alliance for Excellent Education*)
- Lorrie Shepard (*University of Colorado at Boulder*)
Charge to the Panel

• **Charge:**
  Develop a high-level vision for the future of NAEP as well as a plan for moving NAEP toward that vision

• **Audiences:**
  NCES, NAGB, NAEP Contractors, Policy Makers, Researchers, Concerned Citizens
Timeline

- NAEP Summit
- Panel Convened
- Draft Outline
- NAEP Summit (SEAs)
- Final Outline
- Form writing groups
- Initial Draft
- Final Draft
- Deliverable

- August 18-19, 2011
- October 2011
- November 2011
- January 24-25, 2012
- January 2012
- January 2012
- February 2012
- March 2012
- March 31, 2012
A Work in Progress…

- We are still discussing more ideas than are likely to appear in our March 31 draft report. What I am presenting today reflects some of our current thinking, but we have not reached consensus on all of these ideas. There is no assurance that any particular idea will appear in the final version.

- PLEASE SHARE YOUR REACTIONS AS WELL AS YOUR OWN IDEAS!
Overview of today’s presentation

- Context for this initiative
- NAEP infrastructure
- NAEP content frameworks
- NAEP and technology
- Embedded assessments
- Reporting
- NAEP’s continuing importance
Context for This Initiative

- A Changing Environment for NAEP
- More Ambitious Expectations
- Rapid Change in Technology
A Changing Environment for NAEP

- Preparing students for a changing world
- Relation to the CCSS
- Relation to PARCC and SBAC assessments
- Globalization
  - State participation in TIMSS, PISA, …
- Evolution of “education” beyond “schooling”
More Ambitious Expectations

- Reasoning and problem solving in complex, dynamic environments
- Communication and collaboration
  - Group problem solving
- Expanded views of literacy
  - Identifying need for, locating, and evaluating information
  - Fluency with new technologies (e.g., TEL)
- College and career readiness
Rapid Change in Technology

- Increasing educational use of (e.g.):
  - e-textbooks
  - interactive media
  - web-based resources

- Increasing availability of:
  - massive data warehouses
  - data mining

- Increasing communication/cooperation as states move toward "shared learning infrastructure"
NAEP Infrastructure

- Background
- NAEP’s place in “Assessment Ecosystem”
- NAEP Innovations Laboratory?
  - Illustrative topics
Background

- **NAEP is a Complex System**
  - Involves multiple organizations, areas of expertise

- **R&D Component is critical**
  - NAEP not only tracks achievement, but also drives assessment innovation.
  - NAEP’s methodology is emulated worldwide.
  - NAEP R&D is guided and funded through multiple, complex institutional mechanisms.
  - Systematic review might identify possible improvements.

- **NAEP as backbone of “assessment ecosystem”?**
- **NAEP Innovations Laboratory?**
Evolving Assessment “Ecosystem”

- Potential role for NAEP as backbone of evolving assessment infrastructure
  - Design changes to facilitate linkages between NAEP and other assessments
    - Bridge between state-level assessments and TIMSS, PISA, PIRLS, … ?
  - Explicit attention to NAEP vis-à-vis the CCSS
- Defining the state-of-the-art in large-scale assessment
NAEP Innovations Laboratory?

- **Purposes**
  - Strengthen and systematize NAEP R&D
  - Strengthen linkages to other assessment programs and facilitate dissemination

- **Features**
  - Access point for vetting new ideas
  - Organizational structure not yet specified
  - Would support both in-house and 3rd party studies
NAEP Innovations Laboratory?

- Step 1: Review existing structures for NAEP R&D
  - Design and Analysis Committee
  - Validity Studies Panel
  - IES Program on Statistical and Research Methodology in Education
  - NAEP Data Analysis and Reporting contract
  - Education Statistics Support Institute Network (ESSIN)
  - NAEP Secondary Analysis Grants program
  - …
NAEP Innovations Laboratory?

- Step 2: More clearly frame purposes NAEP R&D should serve
  - investigate / assure validity of NAEP findings
  - improve NAEP processes to reduce testing time, reporting time, measurement error, cost
  - expand the range of constructs assessed
  - enable NAEP to serve new purposes
    - e.g., linking to other assessments
Illustrative Topics

- Assessing home-schooled students? lower grades or pre-K? college students?
- R&D on new item types
- Interpretability of NAEP reports
- dynamic (evolving, non-static) content frameworks
- adaptive testing
- technology-enhanced accommodations
- Linkage to state data systems
- Linkage to massive data warehouses
- …
NAEP Content Frameworks

- Relation of NAEP content frameworks to the Common Core State Standards
- Dynamic content frameworks?
Relation to the CCSS

- Considered and rejected:
  - “CCSS” as replacement for NAEP frameworks
  - “CCSS” scales within NAEP

- Distinct functions for NAEP vs. CCSS
  - NAEP covers more content areas
  - NAEP covers more content within each area
    - in part due to focus on populations, not individuals
  - Broader frameworks facilitate linking
  - Value in multiple content frameworks
    - Measuring content outside the focus of instruction can inform wise deliberation regarding evolution of C&I
Dynamic Content Frameworks?

- **Current approach:**
  Content frameworks are held constant for a period of time, then replaced

- **Alternative to consider:**
  - Framework development panels are replaced by standing committees of content experts
  - Achievement is defined relative to a mix of knowledge and skills that is updated incrementally, analogous to the CPI
    - *Affords local continuity, but broad constructs may evolve over time*
    - *Caution: Implies need for a principled way to place relative values on alternative learning outcomes*
NAEP and Technology

- Technology and teaching-and-learning
- Technology and assessment
- Deeper links to state data systems and massive data warehouses
Technology & Teaching-&-Learning

- How might NAEP monitor the full range of complex activities students pursue in modern learning environments?
  - Changing representational modalities and user interface modalities
    - Gesture and touch, sketching, voice, visual recognition and search, augmented reality
  - Interaction with dynamic knowledge representations
Technology and Assessment

- Assessing Old Constructs in New Ways
  - New platforms for item presentation
  - New modalities for student response
  - Adaptive testing to improve precision (especially in the tails of the distribution)

- Assessing New Constructs
  - Technology and Engineering Literacy
  - Problem solving in adaptive environments
  - Technology-mediated collaboration
Deeper Links to State Data Systems

- Expand on current use of state data to improve efficiency of within-state samples
- Expand on initial efforts linking NAEP scales to State assessment scales
  - E.g., mapping of state “Proficient” definitions to NAEP score scales
- Consider building and maintaining integrated longitudinal data structures
  - Interpreting student performance on new kinds of tasks may require knowledge of instructional history
Embedded Assessments

- Embedded Assessments?
- Why EAs in NAEP?
- Can NAEP bridge the gap?
Embedded Assessments?

- Unclear exactly what EAs are, but most accounts suggest:
  - Assessments are linked more closely to ongoing instruction
  - Students engage in complex tasks or create complex (scorable) work products
  - Problems may have multiple solutions
  - Data collection may be “transparent” or unobtrusive
  - Standardization is much weaker than for conventional assessments (task specifications, administration conditions, scoring rubrics)
Why EAs in NAEP?

- Fundamental challenge of standardized assessment where there is no standard curriculum
  - Each item must be a self-contained package, providing all relevant context and content
  - Interpretation of test performance may be unclear if instructional history is not known
- Test tasks must be brief
  - No writing revision cycles, for example
Can NAEP bridge the gap?

- “EA” reflects the perennial desire to link classroom and external assessments
- May not comport with structure and culture of US educational system
Reporting

- Revisit role of achievement levels
- Greater use of “active” reporting
- Reporting R&D
Revisit Role of Achievement Levels

- Achievement Levels are popular
  - ALs are the major vehicle for NAEP reporting
  - ALs respond to demand for normative information
  - ALs seem easy to understand
  - AL reporting is widely emulated and written into law

- Achievement Levels are problematical
  - ALs are poor choice of statistics to summarize distributions
  - ALs are prone to serious misinterpretation
  - ALs are invested with surplus meaning

- Revisit heavy reliance on ALs for NAEP reporting
  - No recommendation to abolish ALs, but instead to accompany them with better methods for meaningful reporting
Greater Use of “Active” Reporting

- Dynamic displays, interactive graphs
  - NAEP Data Explorer and related tools are great resources
  - Bring these methods into online reports with embedded animations and customizable displays
  - These applications might be web-based or stand-alone
Reporting R&D

- Periodic audit of media accounts to identify any systematic miscommunications
- Research on usability of active displays
- Research on alternatives to Achievement Level reporting
- Research on communication about supported versus unsupported interpretations
  - Complexity will increase with availability of cross-assessment linkages
NAEP’s Continuing Importance

- NAEP’s traditional functions are still critically important
  - Low-stakes “audit” function
  - Designed to estimate achievement distributions, not individual scores
  - Source of assessment innovations critically needed to fulfill promise of new consortia

- NAEP may evolve to serve additional purposes
  - Backbone of new assessment “ecosystem” anchoring linkages among multiple assessment programs
Thank you