

National Assessment Governing Board

Reporting and Dissemination Committee

August 3, 2012
10:00 a.m.-12:30 p.m.

AGENDA

10:00 – 11:00 am	<p>ACTION: NAEP Background Questions: Expert Panel Recommendations and Focused Reports <i>Larry Feinberg, NAGB Staff</i> <i>Mary Crovo, NAGB Staff</i></p> <p>[Joint meeting with Assessment Development Committee]</p>	Attachment A
11:00 – 11:10 am	<p>Review of NAEP Release: Science in Action Report <i>Stephaan Harris, NAGB Staff</i> <i>Valerie Marrapodi, Reingold Communications</i></p>	Attachment B
11:10 – 11:20 am	<p>Projected Schedule for Future NAEP Reports <i>Angela Glymph and Arnold Goldstein, NCEs</i></p>	Attachment C
11:20 – 11:30 am	<p>ACTION: Release Plan for Reading Vocabulary Report <i>Stephaan Harris</i></p>	Attachment D
11:30 am – 12:05 pm	<p>Governing Board Report on 12th Grade Preparedness</p> <ul style="list-style-type: none"> a. Content and Approach <i>Larry Feinberg</i> b. ACTION: Release Plan <i>Stephaan Harris</i> 	Attachment E
12:05 – 12:30 pm	<p>Implementation of Policy on Students with Disabilities and English-Language Learners <i>Keith Rust, Westat</i></p>	Attachment F
<i>Information Item</i>	<p>Plans for Puerto Rico Assessment and Reporting in 2013</p>	Attachment G

NOTE TO Reporting and Dissemination Committee on NAEP Background Questions and Contextual Data

At this meeting the Committee will be joined by the Assessment Development Committee to continue the Board's review of the expert panel report on improving NAEP background questions and making better use of the contextual data they provide. The two committees will discuss what recommendations to make for action by the full Governing Board.

The panel report, entitled, *NAEP Background Questions: An Underused National Resource*, was presented to the Board in March 2012 by the panel chair Marshall (Mike) Smith, former U.S. Under Secretary of Education and former dean of the Stanford University School of Education. The report made several dozen recommendations, some quite technical. Its main points were:

- Make greater use of background questions in NAEP reporting, both in regular report cards and special focused or analytical reports.
- Cluster questions around high-priority areas of interest in understanding student achievement and educational practice and policy.
- Obtain richer data by rotating questions in different years, dividing assessment samples, and increasing questionnaire length, but drop unproductive and redundant questions.
- Provide more context in presenting assessment results but avoid causal interpretations.

In response to a request for public comment, the Board received written statements from 24 persons and organizations. About ten persons offered brief comments at webinars conducted on the Internet.

Most comments were positive, including those of the two chief state school officers who responded, Commissioner Stephen Bowen, of Maine, and Lillian Lowery, Delaware Secretary of Education who now is state school superintendent in Maryland. There was very strong support for the panel recommendation for more special NAEP reports on topics of public interest. However, there was also some strongly-worded criticism. Former Governing Board Chairman Chester E. Finn, Jr., and former Commissioner of Education Statistics Mark Schneider both warned that NAEP would jeopardize its reputation if its reports could be construed as giving policy advice. Both were strongly critical of PISA—the Program for International Student Assessment—which the expert panel cited as a possible model for NAEP.

The American Educational Research Association (AERA) sent a very detailed positive response, but also cautioned against using NAEP to monitor implementation of the Common Core State Standards, as the panel suggested. In its response, the National Center for Education Statistics (NCES) supported many of the changes proposed to make questions more valid and reliable and data more accessible to the public. However, NCES expressed concern about additional costs and respondent burden and cautioned against fostering perceptions of intrusiveness.

Over the past year there has been somewhat greater use of background data in NAEP report cards after a virtual absence for almost a decade. NCES also announced plans for three additional focused reports on gender gaps, black male students, and private schools.

This tab includes the expert panel report, responses by NCES and AERA, a press article about the report, and presentations using NAEP contextual information at the National Conference on Student Assessment held in Minneapolis in June 2012.

NAEP BACKGROUND QUESTIONS: AN UNDERUSED NATIONAL RESOURCE

**A Report to the National Assessment
Governing Board by the Expert Panel
on Strengthening the NAEP
Background Questions**

February 22, 2012

Chair: Marshall S. Smith

*Members: Naomi Chudowsky, Alan Ginsburg,
Robert Hauser, Jennifer Jennings, and Sharon Lewis*

Table of Contents

Executive Summary	3
Introduction	9
Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Areas	13
Questionnaire Overview	13
Recommendation 1a. Continually review and refine a core and second-tier set of background topics and questions that are common across NAEP surveys.	15
Recommendation 1b. Extend NAEP background questions to inform topics of current policy interest in the subjects assessed.	16
Recommendation 1c. Select clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics.	18
Recommendation 1d. Use consistency over time as a criterion to consider for question selection and wording.	20
Recommendation 1e. Delete duplicative or low-priority questions to make time for the Panel’s higher priority items.	21
Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for the Background Questions.	22
Recommendation 2a. Improve the validity and reliability of the current measures NAEP uses for its mandated student reporting categories.	23
Recommendation 2b. Enhance the validity of student responses at different grade levels.	25
Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-learn skills including student learning behaviors, motivation and expectations.	26
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”	27
Recommendation 2e. Coordinate NAEP background questions with those asked on international or domestic surveys.	28
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.	29
Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy	30
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.	31
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.	32
Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of the NAEP Background Questions	32

Recommendation 4a. Prepare special reports highlighting the background question findings.....	33
Recommendation 4b. Prepare an online compendium of key background indicators for States and participating urban districts.	34
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.....	36
Recommendation 4d: NAEP should encourage others to conduct exploratory studies of the NAEP background variables.	36
Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.....	37
5. Implementing the Panel Recommendations	40
Recommendation 5a. Exploit existing background data through special reports focused on issues and topics informed by background questions.	41
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.	41
Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.....	42
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.....	43
REFERENCES.....	44
EXPERT PANEL MEMBERSHIP AND AFFILIATIONS.....	46

Executive Summary

For more than four decades the National Assessment of Educational Progress (NAEP) has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions. Less known, but complementing the NAEP assessments, is a rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questionnaires have been cut back over the past decade. They now cover only a small fraction of important student, teacher, and school issues and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education including:

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.
- Tracking progress in implementing key instructional, curricular, and technological changes and educational policy initiatives, such as the Common Core standards.
- Monitoring student motivation and out-of-school learning as research-based factors affecting student achievement.
- Benchmarking high-performing states and urban districts and those with high achievement growth to identify factors that differentiate high-performers from lower-performers on NAEP. This domestic effort would parallel the extensive reporting of background variables in PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) that have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing and low-performing education systems.

The panel proposes building a strategy to make the NAEP background questions an important national resource for educators, policymakers, and the public. The panel sees the need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions through 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			
1. Ask Important Questions	2. Improve the Accuracy of Measures	3. Strengthen Sampling Efficiency	4. Reinstitute Meaningful Analyses & Reporting
<ul style="list-style-type: none"> •Core questions •Rotated questions •Policy questions •Theoretical frameworks •Consistent questions overtime •Delete duplicative or low-priority questions 	<ul style="list-style-type: none"> •Valid •Reliable •Coordinated (with domestic and international surveys) •Cognitive labs 	<ul style="list-style-type: none"> •Spiral sampling •Extended questionnaire time •Alternate surveys •Pooling item responses across surveys 	<ul style="list-style-type: none"> •Special background question reports •Online compendium of responses •Report descriptive not causal findings •Externally conducted research •Improve online tools
<ul style="list-style-type: none"> • Establish a single NAGB committee overseeing background questions • Review budget including need for staff to implement recommendations 			

Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important *Background Questions* around High-Priority Areas.

To the extent that you don't ask and analyze important questions, you can't expect to get back important answers. The panel recommends identifying topics falling into three question groups.

- A *common core* set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.
- A *second tier* of priority background question clusters would be rotated across assessment cycles. Important topics that might be explored include school-parent cooperation, school climate and discipline, school administration including support for learning, and out-of-school learning time.
- A *third tier* would be a set of *policy issues* that would be examined for six years and then rotated out with new ones added. For example, the initial set might start

with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations could be added, and two years after that a module on project-based or online learning.

Once question topics are identified, the panel urges the *selection of clusters of questions that collectively best portray different important aspects of research-based theoretical frameworks for the major educational topics. Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement.*

The Panel recommends two additional considerations to maximize the information worth of the questions chosen. The first is to pay greater attention to the *consistency of question selection and wording* to produce reliable time-series that measure change over time. A review of 400 questions asked about teachers found that about 300 are no longer used, with many replaced by just slightly different wording. A second recommendation is to balance the number of questions asked about a topic with the information value gained. Eight questions are asked about technology use in mathematics but there are no questions about student expectations despite the strong research connection with achievement.

Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for the Background Questions.

The panel urges attention to strengthening the validity, reliability and coordination of NAEP background questions. An important first step in this overall effort would be to improve the *validity, reliability and coordination of the current measures NAEP uses for its mandated student reporting categories.* The panel strongly supports the current review of the SES variables as it is critical to respond to the known limitations of the school-lunch proxy. These problems will worsen with expansion of the Department of Agriculture state pilots, which allow whole-school eligibility for schools serving concentrations of low-income students. The panel also believes that an expanded *cognitive interview capability*, such as a small standing panel of respondents to test out questions, would improve question validity and reliability. We recognize that this may increase costs but it would help make NAEP a better source of information.

The panel recommends improving question wording by replacing imprecise terms such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week.” Furthermore, major information benefits would accrue from coordinating the NAEP background questions with those asked on other international and domestic surveys. To illustrate, the PISA international survey covers number of hours of math instruction in-school and out- of-school; NAEP only asks about days taught math in-school and only about participation in math instruction outside of school and nothing about frequency.

Recommendation Area 3. Reform NAEP *Sampling* to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy.

The panel recommends that NAEP should consider expanding the depth of its background questions through a variety of strategies including spiral sampling (already under study), expanded questionnaire time and rotating background questions across samples. The panel notes that the depth of student information in particular is limited by the ten-minute questionnaire time limit compared with 30 minutes used for TIMSS and PISA. A combination of these strategies would allow NAEP to obtain far richer information while maintaining sampling accuracy and still keeping respondent burden to acceptable levels.

Recommendation Area 4. Reinstitute the *Analysis and Regular Reporting* of the NAEP Background Questions.

This set of recommendations would bolster the analysis and reporting of the background questions by means of separate publications, online tables, and improvements to the Data Explorer. The recommendations also include a reiteration of current policy to not use causal interpretations of point-in-time data.

The panel strongly recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. Exhibit B displays an illustrative overview table for in-school learning opportunities for math that suggests the rich potential information payoffs from background question analyses. A third benchmarking report should also be considered that explores the correlates of high-performing states and districts or those with high achievement growth. These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

Implementation of Recommendations

The panel urges the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES) to move quickly to begin implementing its recommendations to make the background questions a more useful resource, while also recognizing that implementation will take time.

Initial implementation should be undertaken through a three-part plan:

- Immediately produce *special reports on the background data* that analyze the considerable quantity of data already collected, but is largely unreported and unanalyzed.

Exhibit B . Illustrative Table of Background Question Indicators With a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment

	Grade 8 All Students	Eligible for National School Lunch	Grade 8 Students Absent 5 or more days last month	Grade 8 Students in Algebra	Grade 8 Students 5 or more Hours of Math Per Week	Grade 8 Students 1 Hour or More Math Homework	Grade 8 Does Math At An Afterschool or Tutoring Program	Grade 8 Entered Math Through Alternative Certification	Grade 8 Teacher Has Math Major/ Minor/ Special Emphasis	Grade 8 Full-time Math Specialist At School	Grade 8 Assigned To Math By Ability	Grade 8 26+ Students in Math Class	Grade 8 Computers Available to Teachers and Students
Jurisdictions	Scale Score	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages
National	284	44	7	42	37	17	21	17	38	17	76	45	84
Albuquerque	275	60	8	37	65	13	20	27	33	32	66	59	77
Atlanta	266	82	5	27	75	38	57	57	95	61	59	37	90
Austin	287	59	8	23	61	27	30	42	57	58	53	52	89
Baltimore City	261	85	9	46	93	41	38	38	79	53	85	37	71
Boston	282	76	9	66	76	39	30	13	69	12	61	47	56
Charlotte	285	52	8	35	87	18	29	44	47	33	86	76	70
Chicago	270	84	4	32	67	47	37	23	84	20	45	65	88
Cleveland	256	100	11	29	69	33	25	6	58	14	51	44	90
Dallas	274	85	7	32	46	27	39	61	66	13	45	24	57
Detroit	246	79	17	24	81	46	37	11	83	39	18	85	61
District of Columbia (DCPS)	255	70	12	53	65	29	39	57	68	40	53	20	86
Fresno	256	88	10	51	32	11	26	6	37	23	91	75	59
Hillsborough County (FL)	282	54	9	87	20	13	22	40	35	29	95	3	86
Houston	279	76	6	29	63	26	37	56	63	25	84	58	68
Jefferson County (KY)	274	60	7	40	68	14	20	21	34	36	77	80	80
Los Angeles	261	82	6	67	44	40	27	39	67	37	75	52	74
Miami-Dade	272	72	5	36	43	47	25	38	72	25	90	13	88
Milwaukee	254	81	13	30	78	43	31	37	74	82	28	86	78
New York City	272	87	10	28	83	26	39	35	65	36	60	83	79
Philadelphia	265	88	10	34	89	27	27	24	54	32	30	75	89
San Diego	278	60	8	69	48	13	27	11	40	17	78	72	80

Source: NAEP Data Explorer

- Move quickly to initiate a long-term effort to improve the relevance, quality, coherence, and usefulness of a *core and rotated set of background variables while implementing recommended improvements to improve measurement accuracy and sampling efficiency.*
- Further improve the *usability of the Data Explorer and other NCES online tools,* which are already valuable analytic supports.

The panel suggests that NAGB establish a separate standing committee to review all background questions and plans to improve their use. Currently, the Board’s responsibilities for background questions are divided between two of its standing committees. These subgroups do not coordinate their work and the background questionnaires are of secondary interest to both of them. A unified standing committee should regularly monitor and report on implementation of the panel’s recommendations by NCES and Governing Board staff.

In addition, the panel believes that the background questions and how they used in NAEP reporting warrant a periodic, rigorous, and independent evaluation similar to that conducted in the past on NAEP cognitive assessment items.

The panel recognizes that implementing its recommendations will involve resource considerations in terms of time, money, and personnel. One approach to this problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate

lower-priority activities, such as the duplicative collection of racial data and the disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

In conclusion, the NAEP background questions are a unique national information resource. The Governing Board and NCES have a responsibility to develop this resource to better understand academic achievement and the contexts in which it occurs and, hopefully, to help spur educational improvement.

Introduction

The National Assessment of Educational Progress (NAEP) is a unique American education resource. For more than four decades the assessment has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions.¹

Representative samples of students regularly take NAEP assessments in reading, mathematics, science, and writing and the national, state, and urban district levels. Other subjects, including U.S. history, civics, and the arts, are tested at the national level only. Less known, but complementing the NAEP assessments, is a potentially rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questions have been cut back over the past decade. They now cover only a small fraction of important student, teacher and school issues, and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into important questions about the nature and quality of American primary and secondary education. What are the racial, ethnic and economic characteristics of schools at different achievement levels? What are the sources of curriculum content? What resources are available for students? What are the common instructional approaches teachers employ, and how do they adjust approaches to differing student needs? What preparation and training do teachers receive? How is teacher performance evaluated?

In turn, the answers to these survey questions can support important NAEP analyses. The analyses should focus on the unique advantages of NAEP for collecting data and trends over time on education-related background factors paired with achievement results that are representative of states and many urban districts. The following three examples

¹ Although this report focuses on the lack of reporting the background variables for the main NAEP, a similar weakness occurs in not reporting the background variables for the long-term trend NAEP. The report on the 2008 long-term trend assessments did include data on higher level course taking in math in 2008 in relation to that year's NAEP scores, but surprisingly did not report results for earlier years, although available.

illustrate potentially significant descriptive findings from the NAEP background questions for mathematics with respect to:

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.
 - In Arizona, a Hispanic grade-8 student is only 57 percent as likely to have a teacher of mathematics who has a major in mathematics as a white grade-8 student. In California, their chances are nearly equal.
- Tracking progress in implementing instructional, curricular, and technological changes and key education policy initiatives.
 - The proportion of students in schools with no eighth-graders enrolled in algebra is 15 percent nationally. Among urban districts, Miami-Dade and Houston have only 5 percent of their students in schools without a grade-8 algebra course, but Detroit and Milwaukee have over 80 percent of eighth-graders in such schools.
- Monitoring student motivation and out-of-school learning as factors affecting student achievement.
 - More than 45 percent of the grade 4 students in several Southern states (Louisiana, South Carolina and Texas) participated in after-school math instruction. But in several highly rural states (Maine, Oregon and Vermont) the participation rate in after-school math instruction was only about 25 percent.

Moreover, the *extensive reporting of the background variables in PISA and TIMSS* have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing education systems (Darling-Hammond, 2010). These data have been used to examine characteristics of high-performing systems, such as Singapore and Korea, and to study the nature of instruction in subjects such as math and science, where the U.S. performs poorly. In a similar fashion the NAEP data could be used to guide benchmarking of high-performing states and urban districts or jurisdictions experiencing substantial performance growth. This benchmarking activity would be a means to generate hypotheses for further verification through 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).¹²
- 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

¹² In 2011 NAEP began to use the background variables again in its main assessment reports, but with only a single background table related to instruction for each subject and grade. The 2010 Civics, Geography and U.S. History reports also contained a background table related to instruction for the different grades.

variables in participating countries.

The panel believes NAEP should return to its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, the questions could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education and the context for understanding achievement and its improvement. The panel believes there is a need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions through 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.

Exhibit 1. Expert Panel Recommendations to Strengthen NAEP Background Questions in Four Areas			
1. Ask Important Questions	2. Improve the Accuracy of Measures	3. Strengthen Sampling Efficiency	4. Reinstitute Meaningful Analyses & Reporting
<ul style="list-style-type: none"> •Core questions •Rotated questions •Policy questions •Theoretical frameworks •Consistent questions overtime •Delete duplicative or low-priority questions 	<ul style="list-style-type: none"> •Valid •Reliable •Coordinated (with domestic and international surveys) •Cognitive labs 	<ul style="list-style-type: none"> •Spiral sampling •Extended questionnaire time •Alternate surveys •Pooling item responses across surveys 	<ul style="list-style-type: none"> •Special reports •Online compendium of responses •Report descriptive not causal findings •Externally conducted research •Improve online tools
<ul style="list-style-type: none"> • Establish a single NAGB committee overseeing background questions • Review budget include need for staff to implement recommendations 			

The panel recognizes that these recommendations would require commitments of resources and that the Governing Board and the Commissioner of Education Statistics are in the best position to decide on any tradeoffs between existing and proposed features of NAEP that may be required within NAEP’s budget.

Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Information Areas

To the extent that you don't ask and analyze important questions you can't expect to get back important answers. This section recommends strategies for focusing clusters of questions on important information topics within the confines of NAEP questionnaire timelines and administration procedures. Consistent with the NAEP framework, important questions are ones that would primarily focus on the factors that research has shown are related to student achievement. Background questions would also address the implementation of major national policies where NAEP surveys can provide a view from the field state-by-state. In this way, NAEP can report on the distributions and trends of many of the factors and policies important for student achievement.

Questionnaire Overview

With each administration of the subject area assessment, NAEP includes separate student, teacher and school background questionnaires. Although a few questions about subgroups are specified in the NAEP legislation, the Governing Board has the discretion to determine most questions. Exhibit 2 displays the overall number of questions and general question content for each of the three respondent questionnaires on the most recently- reported reading and mathematics surveys.

Exhibit 2. Overview of the Most Current NAEP Mathematics and Reading Background Questionnaires for Students, Teachers and Schools									
	Students 10 Min			Teachers 30 Min			Schools 30 Min		
	Questions: - Student & family background and out-of-school learning - Subject specific: self-perception and school courses content			Questions (subject specific): - Teachers Background: education and training; - Classroom Organization and Instructional practices			Questions: - School Characteristics (including a special charter school survey) - Subject specific: course, student placement, staff composition, training, technology		
	Gr. 4 (2011)	Gr. 8 (2011)	Gr. 12 (2009)	Gr. 4 (2011)	Gr. 8 (2011)	Gr. 12 (2009)	Gr. 4 (2011)	Gr. 8* (2011)	Gr. 12* (2009)
Math: 2011	31	30	40	48	31		39	49	48
Reading: 2011	32	26	34		30				
*School questionnaire for grades 8 and 12 covers reading, math and science. Teacher questionnaire is not administered at grade 12. Source: NAEP Background Questionnaires. Available Feb 2012: http://nces.ed.gov/nationsreportcard/bgquest.asp									

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.”³

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

³ See NCES description of non-cognitive items and questions available December 2011 online: <http://nces.ed.gov/nationsreportcard/tdw/instruments/noncog.asp>.

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.

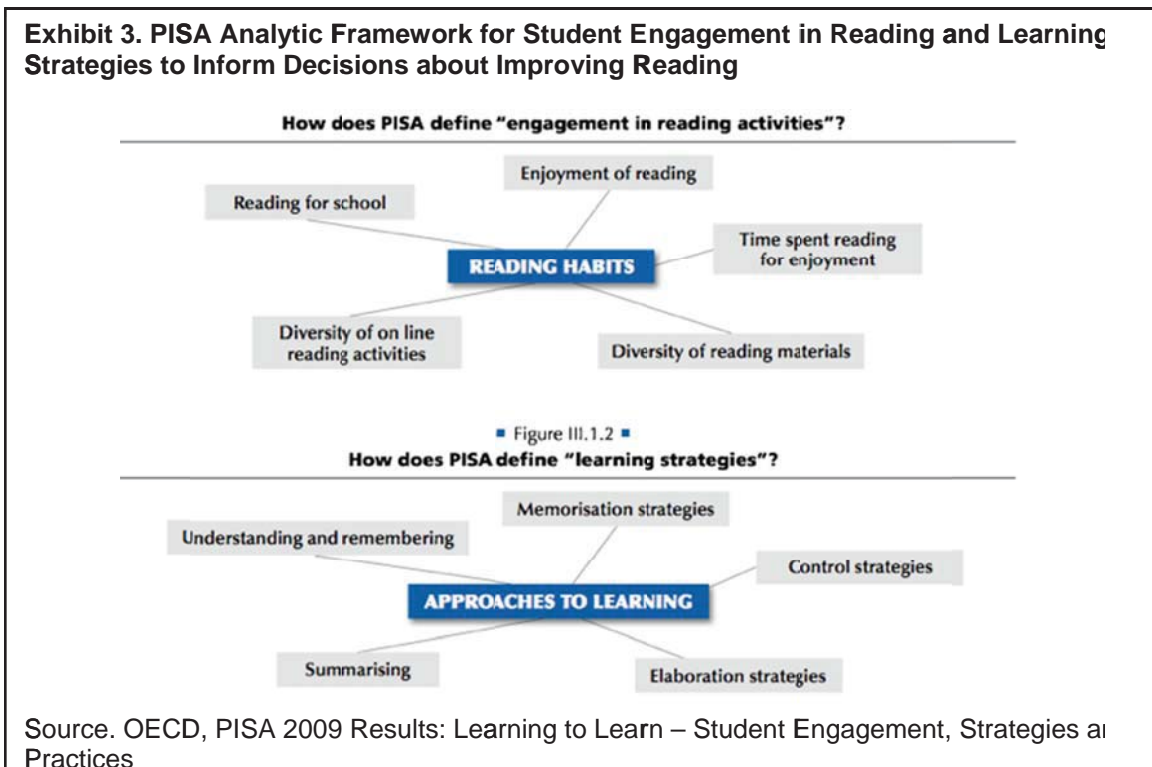
However, currently, NAEP does not formally publish an accompanying document with each assessment that lays out the theoretically-based frameworks that underlie the selection of the background questions and their connection with learning and achievement.

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).



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

Question Category	Total Questions 2011	Total Number Asked in 2005 or Earlier	% of 2011 questions Asked in 2005 or Earlier
Student			
Characteristics	10	8	80%
Curriculum	34	0	0%
Course Offerings	78	28	36%
School Demographics	18	13	72%
School Resources	43	0	0%

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 G

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 multi-dimensional. 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.”⁵

The Panel recommends assessing the potential implications of changes in multi-racial student populations for the valid measurement of the racial/ethnic student classification.

Starting in 2011 NAEP collected multi-racial data from school records and included it in the main subject-matter reports. In 2008, the U.S. Census (2011) reported the multiracial population at 7.0 million or 2.3% of the population. This number is for the full U.S. population and the percentage for the school age children would be expected to be higher to reflect the growing number of inter-racial families in the U.S. NAEP now collects these race / ethnicity data two ways – from school records and student reports. The student reports allow students to check more than one box within racial and ethnic categories. NAEP should compare the self-identified reports with the official school records.

Recommendation 2b. Enhance the validity of student responses at different grade levels.

- Assess whether the same construct (e.g., SES) is best measured by different and increasingly more valid items across grades 4, 8 and 12.

Discussion

A younger (grade 4) NAEP respondent is likely to have more difficulty accurately going

⁵ Quote from NCES Jan. 26, 2012 memo from Peggy Carr to Larry Feinberg.

through a typical question-answer process, which involves 4 steps: (1) understanding and interpreting the question being asked; (2) retrieving the relevant information from memory; (3) integrating this information into a summarized judgment; and (4) reporting this judgment by translating it to the format of the presented response scale (Borgers & Hox, 2000).

The Panel recognizes that NAEP questionnaire design already gives considerable attention to differences in the ability of students at different age groups to go through these four steps to respond accurately to background questions. Thus, NAEP dropped a question about parent's education for grade 4 students because of research suggesting that responses from grade 4 students were less reliable than from older students. However, balanced against possible student response error is the loss of potentially useful information from eliminating questions. The Panel recommends NAEP explore the inclusion in the grade 4 questionnaires of questions that ask about mother's and father's highest education. The exploration should compare the error rates in estimating SES with and without the grade 4 parent education item.

The Panel also recommends that NAEP consider how the same construct (e.g., SES) can be measured by increasingly more valid and multi-dimensional clusters of items for students in upper grades.

Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-learn skills including student learning behaviors, motivation and expectations.

- Learning-to-learn skills refer to a cluster of personal qualities, habits and attitudes and include learning strategies, motivations and expectations. These soft-skills have shown a strong predictive relationship with math and reading achievement and workforce performance over decades (Coleman report, ETS paper on ECLS, NAEP, TIMSS and PISA). The Panel also notes that motivation and expectation questions are a regular component in major NCES national longitudinal surveys and international surveys at the primary and secondary level. However, developing questions that accurately measure non-cognitive skills through subjective responses to survey questions is challenging and should build on the considerable existing body of measurement in this area.

Discussion

To accurately measure some of the hard-to-measure concepts the Panel has recommended (1c above) that NAEP develop clusters of questions that collectively provide a good measure of different aspects of theoretically-based frameworks. Currently, the NAEP background questionnaire, especially the student questionnaire, is highly restricted by time constraints and does not contain the rich set of items needed to validly measure many learning attributes associated with student achievement.

Exhibit 6 provides an example of how PISA’s in-depth questioning draws out students’ approaches to understanding a particular type of text. In essence, the questionnaire creates more authentic learning situations from which to document students’ behaviors.

Exhibit 6. PISA’s In-Depth Student Questions Of How They Would Approach Remembering Information in a Text Approximates An Authentic Assessment Item

Reading task: You have to understand and remember the information in a text.

How do you rate the usefulness of the following strategies for understanding and memorizing the text?

Possible strategy	Score					
	<i>Not useful at all</i>			<i>Very useful</i>		
	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>	<i>(5)</i>	<i>(6)</i>
a) I concentrate on the parts of the text that are easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) I quickly read through the text twice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) After reading the text, I discuss its content with other people.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) I underline important parts of the text.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) I summarize the text in my own words.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) I read the text aloud to another person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Source: OECD PISA 2009 Student Questionnaire

The Panel recommends that NAEP explore including these rich behavior questions for grades 8 and 12 even if it would require expanding the student questionnaire time for completion.

Recommendation 2d. Improve question reliability by replacing imprecise phrases such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week”.

Discussion

NAEP should ask questions involving frequency of behaviors or intensity of services in a form that elicits the most precise meaning to these terms. In this regard, some NAEP questions are not specific and the reliability of responses to these questions may be low.

The following illustrates two questions on the NAEP 2009 teacher questionnaire asking teachers about frequencies of time spent on science. Question a) asks about time spent on physical science in terms using categories such as “Little”, “Some” or “A lot” that could

mean quite different amounts of time depending on teacher norms. By contrast, question b) uses the preferred wording in which response times are expressed in clear distinct time intervals.

Question a): In this class, about how much time do you spend on physical science?

Answers: None = 4%, Little = 9%, Some = 27%, A lot = 60%

Question b): About how much time in total do you spend with this class on science instruction in a typical week

Answers: Less than 1 hour = 1%, 1-2.9 hours = 4%, 3-4.9 hours = 60%, 5-6.9 hours = 25% , 7 hours or more = 9%

NAEP should specify responses to questions about frequency and intensity in a specific quantifiable format wherever feasible.

Recommendation 2e. Coordinate NAEP background questions with those asked on international or domestic surveys.

- NAEP should explore framing its questions with as identical wording as feasible to similar questions found on international assessments.
- NAEP should examine the feasibility of NAEP coordinating with the NCES household survey to administer the household survey to families of students who participate in the NAEP subject assessments. This coordination between the two surveys would link the results of adults in the household survey with students' NAEP assessment scores.

Discussion

In recent years NAEP cognitive assessment results have been linked internationally to place NAEP national and state disaggregated performance on an international TIMSS or PISA scale. NCES now is linking the 2011 grade 8 mathematics and science assessments of NAEP and TIMSS so international benchmarks can be reported on NAEP. Potentially, many of the responses to the background questions can also be compared with similar questions asked on international assessments. Examples include time spent on homework, after-school learning, taking algebra in the eighth grade, or teacher preparation to teach math or science.

To make valid international comparisons, NAEP needs to word its questions so that they are very similar or identical to the wording of the comparable questions on international surveys. Comparability of wording will only be achieved through careful question linking.

Exhibit 7 illustrates the potential payoffs that could occur from linking NAEP responses to those on an international assessment measuring with respect student time learning in regular school lessons and out-of-school lessons compared with high-scoring Japan and Korea.

Exhibit 7. Student Time Per Week Learning Math in Regular School Lessons and Out-of-School Lessons, PISA Age-15, 2006

Regular In-School Lessons: Mathematics (Age 15, 2006)										
	No Time		Less than 2 hr		2-4 hr		4-6 hr		6+ hr	
	Math Score	%	Math Score	%	Math Score	%	Math Score	%	Math Score	%
Japan		0	444	8	491	35	551	42	572	14
Korea	416	1	451	3	520	21	561	58	576	16
U.S.	429	5	430	23	465	20	511	38	490	15

Out-of-School: Mathematics (Age 15, 2006)										
	No Time		Less than 2 hr		2-4 hr		4-6 hr		6+ hr	
	Math Score	%	Math Score	%	Math Score	%	Math Score	%	Math Score	%
Japan	480	24	517	32	551	23	575	13	583	8
Korea	520	23	541	14	573	33	579	17	584	13
U.S.	512	79	478	11	454	5	456	3	433	2

Source: NAEP Data Explorer

- 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. Reinstigate 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.
- *Learning to Learn: Student Engagement, Strategies and Practices* examines 15-year-olds' motivation, their engagement with reading and their use of effective learning strategies.
- *What Makes a School Successful? Resources, Policies and Practices* examines how human, financial and material resources, and education policies and practices shape learning outcomes.

Students On Line: Digital Technologies and Performance, explores student use of information technologies for learning.

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 Equation⁷

Exhibit 8 illustrates for the 22 districts participating in the 2011 Trial Urban Assessments a hypothetical mock-up of background question responses focused around grade 8 and mathematics. A few findings from the urban district data in Exhibit 8 illustrate the potential value of indicator comparisons:

- The systems with the highest percentage of students absent 5 or more days were Detroit, Milwaukee, DC and Cleveland, which were also places with lower student scores.
- For grade 8 students taking algebra, the highest scoring districts of Austin and Charlotte had relatively low rates of absenteeism.

⁷ From *Change the Equation*, a non-profit, non-partisan coalition of more than 100 CEOs who are committed to bringing high-quality Science, Technology, Engineering, and Mathematics (STEM) learning to every U.S. child.

- Although urban school systems have somewhat higher rates of students participating in math at an afterschool tutoring or school program, only Atlanta had at least half the students avail themselves of afterschool assistance.
- Urban districts for the most part have above national-average percentages of staff teaching math with a major, minor or special emphasis in mathematics.
- Access to the Internet at home is widespread among urban areas making school support for learning at home more feasible than might be generally believed.

Exhibit 8. Illustrative Table of Background Question Indicators With a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment

Jurisdiction	Grade 8 All Students	Eligible for National School Lunch	Grade 8 Students Absent 5 or more days last month	Grade 8 Students in Algebra	Grade 8 Students 5 or more Hours of Math Per Week	Grade 8 Students 1 Hour or More Math Homework	Grade 8 Does Math At An Afterschool or Tutoring Program	Grade 8 Entered Math Through Alternative Certification	Grade 8 Teacher Has Math Major/ Minor/ Special Emphasis	Grade 8 Full-time Math Specialist At School	Grade 8 Assigned To Math By Ability	Grade 8 26+ Students in Math Class	Grade 8 Computers Available to Teachers and Students
	Scale Score	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages	Percentages
National	284	44	7	42	37	17	21	17	38	17	76	45	84
Albuquerque	275	60	8	37	65	13	20	27	33	32	66	59	77
Atlanta	266	82	5	27	75	38	57	57	95	61	59	37	90
Austin	287	59	8	29	61	27	30	42	57	58	53	52	89
Baltimore City	261	85	9	46	93	41	38	38	79	53	85	37	71
Boeton	282	76	9	66	76	39	30	13	69	12	61	47	56
Charlotte	285	52	8	35	87	18	29	44	47	33	86	76	70
Chicago	270	84	4	32	67	47	37	23	84	20	45	65	88
Cleveland	256	100	11	29	69	33	25	6	58	14	51	44	90
Dallas	274	85	7	32	46	27	39	61	66	13	45	24	57
Detroit	246	79	17	24	81	46	37	11	83	39	18	85	61
District of Columbia (DCPS)	255	70	12	59	65	29	39	57	66	40	53	20	86
Fresno	256	88	10	51	32	11	26	6	37	23	91	75	59
Hillsborough County (FL)	282	54	9	67	20	13	22	40	35	29	95	3	86
Houston	279	76	6	29	63	26	37	56	63	25	84	58	68
Jefferson County (KY)	274	60	7	40	66	14	20	21	34	36	77	80	80
Los Angeles	261	82	6	67	44	40	27	39	67	37	75	52	74
Miami-Dade	272	72	5	35	43	47	25	38	72	25	90	13	88
Milwaukee	254	81	13	30	78	43	31	37	74	82	28	86	78
New York City	272	87	10	28	83	26	39	35	65	36	60	83	79
Philadelphia	265	88	10	34	89	27	27	24	54	37	30	75	89
San Diego	278	60	8	66	46	13	27	11	40	17	78	72	80

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.

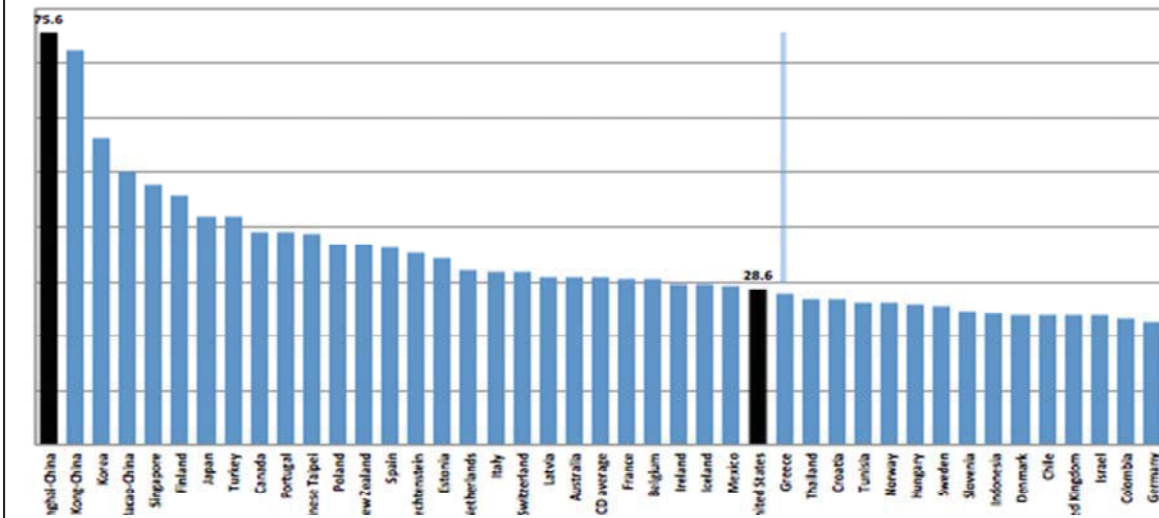
Discussion

NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that is analyzable through 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

Exhibit 9 The PISA Statlink To Excel Simplified Preparing This Graphic That Was Modified From the PISA Original To Highlight U.S. Performance Relative To Singapore

Chart: The Percentage of Disadvantaged Students (Low SES) Who Attain the Top Quarter On PISA Reading Performance Across All Countries



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

Average scale scores for mathematics, grade 4 by year, jurisdiction and time per week on math

Year	Jurisdiction	Less than 3 hours		3-4.9 hours		5-6.9 hours		7 hours or more	
		Average scale score	Standard Error	Average scale score	Standard Error	Average scale score	Standard Error	Average scale score	Standard Error
2011	Alabama	222	(3.5)	216	(7.4)	232	(1.3)	232	(1.4)
	Alaska	232	(5.9)	233	(3.5)	238	(1.2)	237	(1.9)
	Arizona	226	(5.1)	223	(4.3)	236	(1.5)	237	(1.6)

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									
		Chart		Significance Test		Gap Analysis			
Percentages for mathematics, grade 4 by year, jurisdiction and time per week on math instruction [T088001]: 2011									
Year	Jurisdiction	Less than 3 hours		3-4.9 hours		5-6.9 hours		7 hours or more	
		Percentage	Standard Error	Percentage	Standard Error	Percentage	Standard Error	Percentage	Standard Error
2011	Alabama	4	(1.1)	3	(1.2)	62	(3.3)	31	(3.0)
	Alaska	3	(0.5)	8	(0.9)	58	(2.2)	31	(2.1)
	Arizona	3	(0.8)	5	(1.1)	57	(3.5)	35	(3.5)
NOTE: Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Mathematics Assessment.									

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		
Strategy	Recommendation	Activities/Products
1.. <i>Select core and rotated clusters of questions around research-based theoretical frameworks</i>	1a, 1c	<ul style="list-style-type: none"> • 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. <i>Extend NAEP Background Questionnaires to monitor topics of current policy interest</i>	1b	<ul style="list-style-type: none"> • 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. <i>Launch a process for the continual examination of the validity, reliability, efficiency, and consistency of measures</i>	2a,2b,2c 1d, 2f	<ul style="list-style-type: none"> • 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. <i>Report on item sampling reforms to incorporate extended question sets and topics including eliminating duplicative and low-priority items</i>	3a, 3b	<ul style="list-style-type: none"> • 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

- Barton, P. (2002). *Perspectives on background questions in the national assessment of educational progress*. Report to The National Assessment Governing Board. Available December 2011 online: <http://nagb.org/publications/comment-03b.htm>.
- Borgers, N. and Hox, J. (2000). *Reliability of responses in questionnaire research with children*. Available January 2012 online: <http://joophox.net/papers/p021704.pdf>.
- Braun, H., Jenkins, F., and Grigg, W. (2006). *A closer look at charter schools using hierarchical linear modeling* (NCES 2006-460). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office.
- Chen, C. and Stevenson, H. (1995). Motivation and mathematics achievement, A comparative study of Asian American, Caucasian-American, and East Asian high school students. *Child Development*. 66, 1215-1234.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- ETS (2010). *Issues review memos: NAEP questionnaires- student core items; NAEP questionnaires- school core items; How to best examine teacher effectiveness: addressing background issues for the NAEP teacher core items; working drafts*. Last updated: October 15, 2010.
- Jones, F.L. and J. McMillan (2001) 'Scoring occupational categories for social research: a review of current practice, with Australian examples. *Work, Employment and society* 15(3): 539-563.
- Grissmer, D., Flanagan, A, Kawata, J. and Williamson, S. (2000). *What state NAEP test scores tell us about improving student achievement*. Rand Corp. Available December 2011 online: http://www.rand.org/pubs/monograph_reports/MR924.html#toc.
- Hauser, R. and Warren, J. (1997). Sociological indices for occupations: a review, update and critique. *Sociological Methodology*. 27, 177-298.
- Lacour, M. & Tissington, L. (2011). The effects of poverty on academic achievement. In *Educational Research and Reviews* Vol. 6 (7), pp. 522-527, July 2011 Available online at <http://www.academicjournals.org/ERR>.
- Levine, R., Huberman, M. Allen, J. and DuBois, P. (2001). *The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items* Working Paper No. 2001-19 September 2001. Available January 2012 online: <http://nces.ed.gov/pubs2001/200119.pdf>.
- Levine, R., Huberman, M. and Buckner, K (2002). *The measurement of instructional background indicators: cognitive laboratory investigations of the responses of fourth and eighth grade students and teachers to questionnaire items*. U.S. Department of Education, National Center for Education Statistics, NCES 2002-06. Available December 2011 online: <http://nces.ed.gov/pubs2002/200206.pdf>.
- Mead, N., Grigg, W., Moran, R., and Kuang, M. (2010). *National Indian education study 2009 - part II: the educational experiences of American Indian and Alaska Native students in grades 4 and 8* (NCES 2010-463). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington,

- D.C. Available January 2012 online:
<http://nces.ed.gov/nationsreportcard/pdf/studies/2010463.pdf>.
- NAGB (2003). *Background information framework for the National Assessment of Educational Progress*. National Assessment Governing Board. August 1, 2003.
 Available December 2011
 online: <http://www.nagb.org/publications/frameworks/backinfoframenew.pdf>.
- NAEP (2004). *NAEP 2004 trends in academic progress three decades of student performance in reading and mathematics*. National Center for Education Statistics.
 Available December 25, 2011
 online: <http://nces.ed.gov/nationsreportcard/pdf/main2005/2005464.pdf>.
- NAEP (2011). *Questionnaires for Students, Teachers, and Schools*. Available December 2011 online: <http://nces.ed.gov/nationsreportcard/bgquest.asp>.
- Noel-Miller and Hauser -- NAEP-SES Project Progress Report (August 2011)
- OECD (2009). *2009 Student questionnaire*. Program for International Student Assessment 2009. Available December 2011
 online: http://nces.ed.gov/surveys/pisa/pdf/quest_pisa_2009_student.pdf.
- TMSS 2011 assessment design. Available January 2012 online:
http://timss.bc.edu/timss2011/downloads/TIMSS2011_Frameworks-Chapter4.pdf.
- U.S. Congress. Healthy, Hunger-Free Kids Act of 2010. Available January 2012 online:
http://www.fns.usda.gov/cnd/Governance/Legislation/CNR_2010.htm.
- U.S. Department of Education: USED (2001). *Longitudinal evaluation of school change and performance: final report*. Prepared by Westat and Policy Studies Assoc..
 Available December 25, 2011
 online: http://www2.ed.gov/offices/OUS/PES/esed/lescp_highlights.html.

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

The 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

March 2012

The Background Variables Represent a Potentially Critical National Information Resource

- Describe **educational resources** available to support learning for students with different home backgrounds.
- Track progress in implementing key **curricular, technological and education policy changes**.
- Monitor **student motivation, out-of-school learning** and other research-based factors affecting learning.
- **Benchmark high-performing states and urban districts and those with high achievement growth** to identify factors differentiating performance. This domestic effort would parallel extensive international analyses of the background variables in PISA and TIMSS.

2

But NAEP Background Data Are Currently an Underused Resource

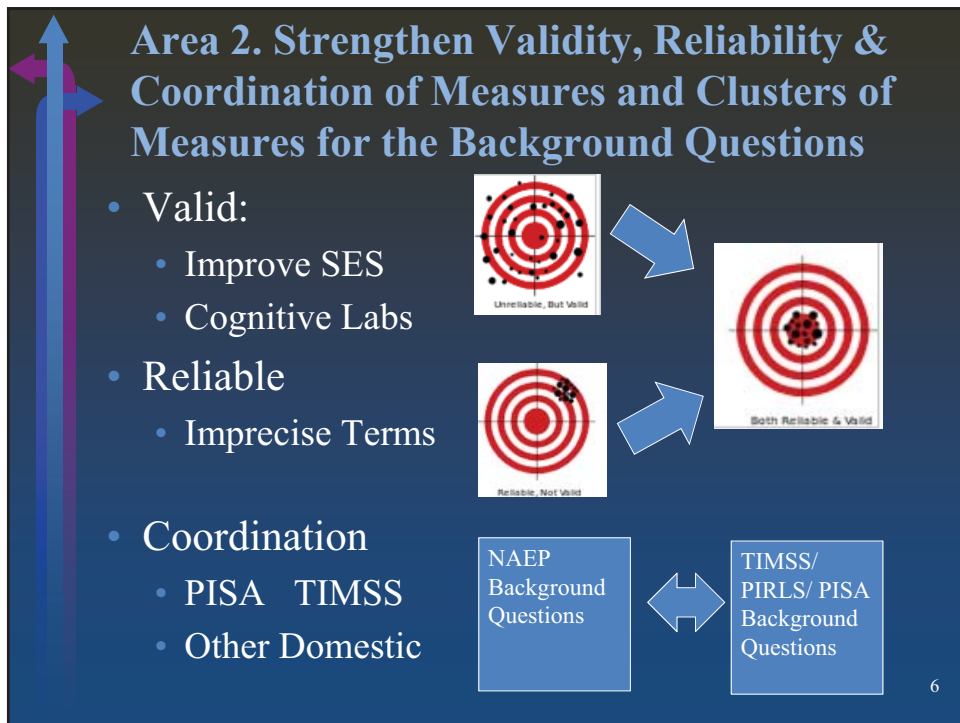
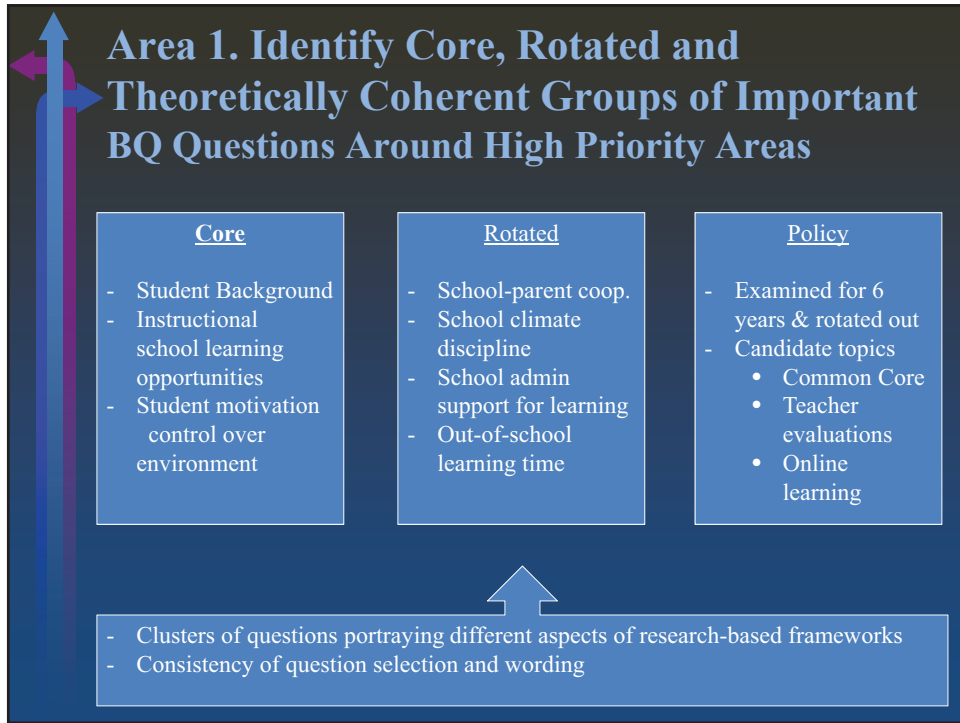
- NAEP no longer reports on student, teacher and school background responses when publishing main assessment results (except for student subgroups).
- In-depth NAEP special reports using the BQ data are rare. *The Educational Experiences of American Indian Students in Grades 4 and 8 is an exception (2010).*
- This leaves the primary access to the BQ data through the NAEP Data Explorer, which is great for the professional researcher but not for educators, policy makers, or the general public.

3

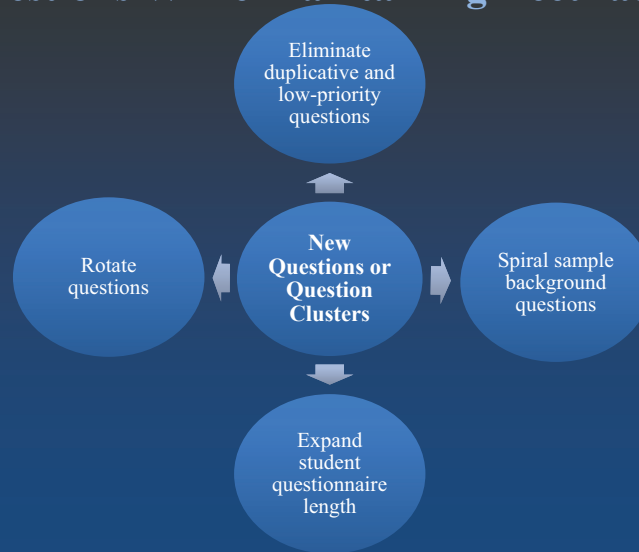
Expert Panel Recommendations to Strengthen NAEP BQ in Four Areas

1. Ask Important Questions	2. Improve the Accuracy of Measures	3. Strengthen Sampling Efficiency	4. Reinstitute Meaningful Analyses & Reporting
<ul style="list-style-type: none"> •Core questions •Rotated questions •Policy questions •Theoretical frameworks •Consistent questions over time •Delete duplicative & low-priority questions 	<ul style="list-style-type: none"> •Valid •Reliable •Coordinated (with domestic and international surveys) •Cognitive labs 	<ul style="list-style-type: none"> •Spiral sampling •Extended questionnaire time •Alternate surveys •Pooling item responses across successive surveys 	<ul style="list-style-type: none"> •Special BQ reports •Online compendium of responses •Report descriptive not causal findings •Externally conducted research •Improve online tools
<ul style="list-style-type: none"> • Establish a single NAGB committee overseeing background questions • Review budget including need for staff to implement recommendations 			

4



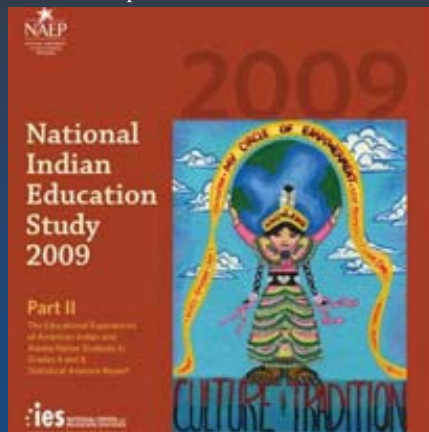
Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Accuracy



7

Area 4. Reinstitute the Analysis and Regular Reporting of the NAEP Background Questions

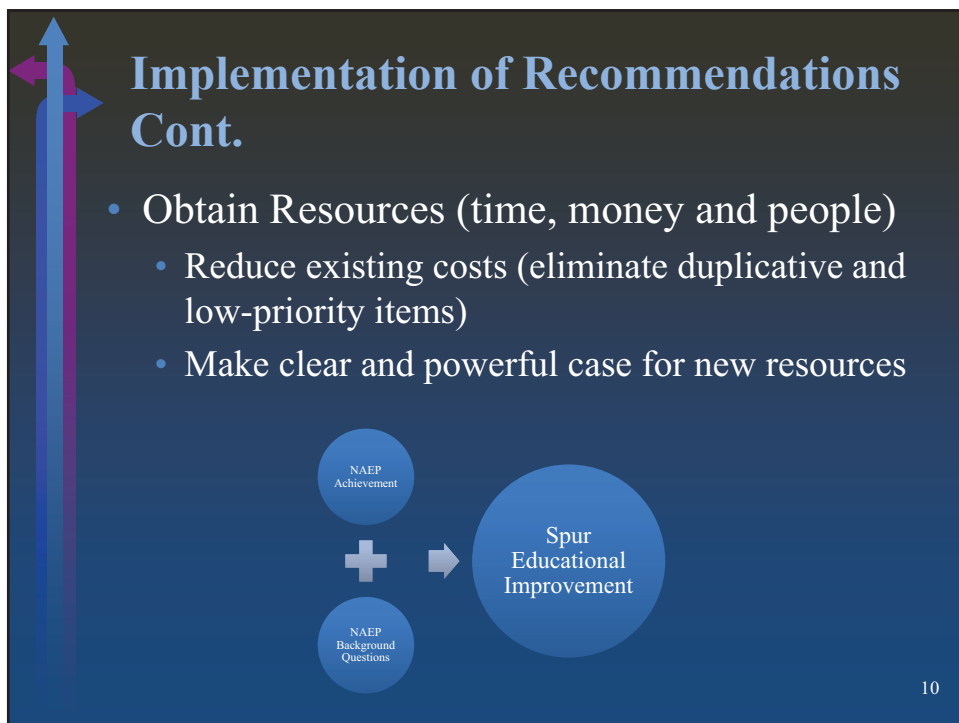
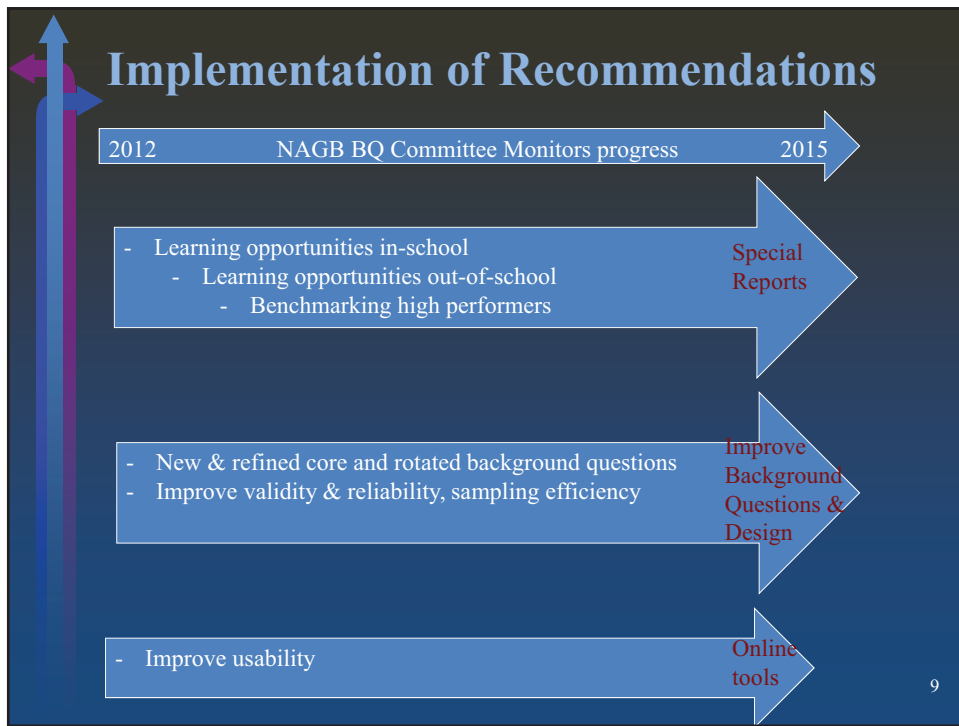
Education Experiences of Native Americans



Statistical Tables: Urban Districts in Math

Exhibit B. Illustrative Table of Background Question Indicators With a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment

	Grade 8 All Students	Grade 8 Eligible for National School Lunch Program	Grade 8 Students in Special Education	Grade 8 Students in Homeless Programs	Grade 8 Students in Migrant Education Programs	Grade 8 Students in Title I Programs	Grade 8 Students in English Language Learner Programs	Grade 8 Students in Gifted/Talented Programs	Grade 8 Students in Career/Technical Education Programs	Grade 8 Students in Alternative Education Programs	Grade 8 Students in Other Programs	Grade 8 Students in All Programs
Anders	284	44	7	42	37	17	21	17	38	17	16	41
Atlanta	275	60	8	57	65	13	20	27	33	32	46	59
Atlanta	266	62	5	27	75	26	57	57	95	41	59	31
Austin	287	59	8	25	61	27	30	42	57	39	53	52
Baltimore City	261	65	9	45	55	41	38	38	75	11	85	17
Boston	252	76	9	66	76	39	30	13	69	12	41	47
Charlotte	265	52	8	35	87	18	29	44	47	13	66	76
Chicago	270	84	4	32	67	47	57	23	84	20	45	65
Chicago	268	102	11	28	69	33	25	6	58	14	14	44
Delhi	274	65	7	32	46	27	39	61	68	13	45	24
Detroit	246	75	17	24	81	46	37	11	83	19	18	63
District of Columbia	255	70	12	43	65	29	39	57	68	40	33	30





Thank You

Panel Members

Marshall S. Smith (Chair)

Naomi Chudowsky

Alan Ginsburg (Secretary)

Robert Hauser

Jennifer Jennings

Sharon Lewis

Response to the Governing Board’s Expert Panel on Background Questions

NCES commends the Governing Board for their initiative to examine the background questions in their current usage and process. NAEP’s background questions are an important resource to the educational community. As such, it is critical that NCES and the Governing Board continue to evaluate processes and improvements to strengthen this important resource.

NCES supports many of the recommendations discussed in the paper. Obtaining policy relevant information and trends is an important goal of NAEP. In particular, NCES agrees with the following recommendations and through our current and future activities we will continue to support them and strive to enhance their presence in the NAEP program:

While we may not agree with all of the individual components of these recommendations, we endorse the principle behind the recommendation.

- **Recommendation Area 2** (*Strengthen the validity, reliability and coordination of the measures and clusters of measures for background questions*). NCES is interested in exploring additional ways to strengthen the reliability and validity of background questions. Currently, NAEP employs the following procedures to evaluate the reliability and validity of a questionnaire item:
 - Conduct expert panel reviews of items to confirm that the question covers the full range of the meaning of the construct.
 - Conduct cognitive interviews to check for consistent understanding of questions (and terms) and investigate the range of responses that respondents will report.
 - Examine missing rates and response patterns to assess whether there may be a problem that warrants changing an item or not including an item in a future administration.
 - Examine the relationship between survey responses with other variables (e.g., students average scale scores).

- **Recommendation 1d** (*Use consistency over time as a criterion to consider for question selection and wording*). The report calls for consistency in question wording over time as a consideration. Reporting trend information is of critical importance for NAEP and NCES concurs with the importance of this role. Over the last several years, changes to the wording of background questions have been made to ensure consistency across the questions (such as consistency across grades or consistency across subjects). As such, the program is actively adopting the philosophy of maintaining consistency in the background questions.

- **Recommendation Area 4d** (*NAEP should encourage others to conduct exploratory studies of the NAEP background variables*). External researchers utilize NAEP data to prepare reports focused on background question findings. It is important that researchers are provided opportunities to access NAEP data to help ensure the background questionnaire are not an underused resource. NCES would be supportive of refining the current grant process in order to give an increased emphasis for these types of activities.
- **Recommendation 4e** (*Further improve the powerful online NAEP tools for data analysis*). NCES is interested in allowing easier access to the wealth of background information to the general public. In addition, it may be worth noting that some of the report's recommendations related to expanding the current functionality of NDE to make exploration and exporting of background variable data more convenient are, at least in part, already implemented. For example, it is now possible to export data to Excel (rather than needing to key-enter the data) from the Build Reports tab. Also, it is possible to search for keywords in a variable's name on the Select Variables tab.

While we support the overall effort and the above-mentioned specific recommendations, NCES would like to respond to the following areas discussed in the report that are potentially more problematic:

- **Increase Student and School Burden (as suggested in Recommendations 2c and 3b):** We are greatly concerned about the report's recommendations that would significantly increase burden on students and schools. Specifically, the recommendation to expand burden on student and schools by requiring more time for answering background questionnaires will likely have a negative impact on participation and response rates. Unlike TIMSS, PISA, and PIRLS, there are many schools that are sampled by NAEP on a regular basis. NAEP is assessed every year, often at grades 4, 8, and 12, while the international assessments are assessed every three to five years, with only one (for PISA and PIRLS) or two (for TIMSS) groups of students. NAEP is a much more frequent presence, and consequently, must be more considerate of the time schools are asked to divert from teaching to testing. Because NAEP is administered with much greater frequency, one might predict school participation in NAEP would be noticeably lower than in the international assessments. However, the data indicate the opposite: NAEP's response rates are significantly better and this result may be attributed to testing time.

NAEP currently requires only 65 minutes for assessment time to answer the cognitive and background questions for paper and pencil assessments and only 75 minutes for computer-based assessments, with the additional 10 minutes being added to the cognitive portion of the assessment. The international assessments, however, require between 100 and 150 minutes of assessment time. Field reports suggest the amount of time required on the international assessments is a significant problem and the time burden reduces school participation. Conversely, the school response rates for NAEP are much higher than for any international assessment administered in the U.S., as indicated in the table below.

U.S. School Participation Rates for NAEP and International Assessments

Assessment	School Response Rate	Frequency of Assessments	Amount of Assessment Time (Cognitive and BQ)
NAEP (grade 4)	97%	Every year	65-75 minutes
NAEP (grade 8)	97%	Every year	65-75 minutes
NAEP (grade 12)	94%	Every year	65-75 minutes
TIMSS (grade 4)	70%	Every 4 years	102 minutes
TIMSS (grade 8)	68%	Every 4 years	120 minutes
PISA	68%	Every 3 years	150 minutes
PIRLS	57%	Every 5 years	100 minutes

Notes:

Participation rates are from the most recently published assessments: NAEP (2011), TIMSS (2007), PISA (2009), and PIRLS (2006).

NAEP is congressionally mandated for reading and mathematics for grades 4 and 8. These subjects are in the field every other year. NAEP 2010 school response rates, a year in which reading and mathematics were not in the field, were also superior to school responses rates for the international assessments. The 2010 NAEP school response rate was 96% for grade 4, 96% for grade 8, and 89% at grade 12.

- **Spiral Background Questions (as suggested in Recommendation 3a):** The report recommends implementing a procedure in which the background questions are spiraled so that no student receives all of the background questions, but that the full set of questions is administered across the entire sample. The intention behind this recommendation is to expand the number of background questionnaire items that are administered as part of any given subject-area assessment. NCES supports this effort; however several challenges must first be addressed.

For instance, in order to implement this effectively and so that the results are unbiased, all questions would need to be included in the analysis conditioning model. The exact approach that could be taken to implement this would need to be determined. Numerous options could be considered for modifying the conditioning model (such as employing multiple conditioning models or including all questions and treating the ones that were not administered as missing). NCES would need to investigate these different alternatives to determine the most appropriate methodology for the NAEP data, both in terms of the reliability and validity of the results and the reporting timeline and requirements.

In addition, it is important to recognize that spiraling the background questions will decrease the individual sample size for each question. In return, the standard errors associated with the results will increase. The exact impact of this increase in standard errors would need to be investigated to evaluate if it would have significant reporting implications for subgroups within jurisdictions. NCES wants to make sure that spiraling does not compromise our

ability to report background data due to a sampling (e.g., background data for urban districts that tracks progress in implementing instructional curricular, and technological changes).

- **Rotating Background Questions (as suggested in Recommendation 1a):** The report calls for the rotation of some background questions, such that they would only be assessed in every other administration. The theory of such an approach is that information on additional topics and questions can be collected, without increasing burden. However, it is important to note that this approach would yield longer periods of time between trend reporting and, thus, limited trend information would be available with each assessment. For the legislatively mandated reading and mathematics grades 4 and 8 assessments, states use the trend information to help explain changes from one administration to the next. In addition, for the assessments administered less frequently (i.e., every 4 or 6 years), very little trend data could be captured among rotated questions over the course of the framework.

- **Scope of Questions (as referenced in the Executive Summary and Recommendation 1b):** The report cites the importance of including policy relevant topics in the NAEP background questionnaires, such as opportunity-to-learn issues; key instructional, curricular, and technological changes; and out-of-school learning factors. The panel should be aware that there are examples of background questions that already address each of these topics in both the core and subject-specific student questionnaires, as well as in the teacher questionnaires. NCES has previously met resistance from some of our stakeholders who view certain topics, such as out-of-school learning, as potentially too intrusive. Moreover, NAEP legislation prohibits evaluating or assessing personal or family beliefs and attitudes.

No matter how much time the program decides to require of its questionnaire respondents, there will always have to be choices made based on priorities for reporting and information policy. Extending the NAEP background questions further to inform topics of current policy interest could put the program in a precarious position given the recent national conversation concerning the level of government involvement in schools. NAEP is not designed to serve as a program evaluator. The NAEP legislation stipulates, “The use of assessment items and data on any assessment authorized under this section by an agent or agents of the Federal Government to rank, 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.” NCES recommends exercising significant caution so school administrators, teachers, students, and their parents do not characterize NAEP background questions as overly intrusive.

- **Expanded Use of Cognitive Laboratories (as suggested in Recommendation 2f):** The report calls for expanded use of cognitive interviews in developing new background questions. It is not clear how NAEP can significantly improve upon current processes based on this recommendation. NAEP employs extensive application of cognitive interview procedures to ensure the language and terms used in background questions are clear to respondents, and elicit the type of information desired. Starting in early 2009, all newly developed and revised questions have undergone cognitive laboratory procedures. In addition to traditional cognitive interview techniques, NCES and its contractors use other

techniques to ensure high-quality items, including pre-testing in the form of focus groups and item tryouts, especially with many of the new computer-delivered background questionnaire development efforts.

- **Pooling Item Responses Across Surveys (as mentioned in Recommendation 3b):** It is not clear how item responses could be pooled across successive surveys in any meaningful way. Each assessment year is sampled independently to ensure the results represent the population of interest. As such the combined, or pooled, responses across years would not represent either target population. For, example, pooling the responses from the 2009 and 2011 grade 4 mathematics student questionnaires would not represent results from either 2009 or 2011. As such, pooling item responses would present a severe violation to data integrity.

Finally, NCES would like to acknowledge that most of the efforts discussed in the paper (both those that NCES endorses and those that would require additional consideration) have cost implications. Creating additional questions and assessing students for increased time, creating additional reports, spiraling the background questions, and conducting bridge studies would all incur additional costs. Without additional funds from the federal budget, the addition of these activities would place other NAEP activities in jeopardy.

American Educational Research Association Comments
to the National Assessment Governing Board
on the Expert Panel Report
NAEP Background Questions: An Underused National Resource
Response to April 11, 2012 Request from NAGB

The American Educational Research Association (AERA) is pleased to have been invited to comment on the expert panel report on *NAEP Background Questions: An Underused National Resource* (2012). We commend the leadership of the National Assessment Governing Board (NAGB) for undertaking a review of background questions at this critical time, when disparities in educational progress and achievement continue to challenge our educational and societal commitments.

AERA staff, leadership, and members have been engaged in NAEP/NAGB for many years through providing testimony on behalf of NAEP appropriations, participating in focus groups on future directions of NAEP, providing research forums for NAGB leadership, and participating in previous workshops about background questions. AERA is a co-publisher and distributor of a comprehensive history of NAEP, *The Nation's Report Card: Evolution and Perspectives* (2004). Also, as the national scientific association for approximately 25,000 education researchers, many of our members serve on NAEP/NAGB committees and boards, and most importantly they are active users of NAEP through undertaking scientific studies that have both added to knowledge and revealed areas where NAEP could be strengthened.

The AERA comments are organized in three parts as suggested in the NAGB request: a general comment on the overall direction provided by the report, comments highlighting specific areas of support, and areas where we have specific concerns or raise questions. We also offer a concluding comment regarding how best to proceed in the next generation of NAEP development, including in areas of background variables, during the years ahead.

General Comment on the Guidance Provided by the Report

Overall, this report constitutes a very important step in enhancing the largest U.S. educational survey. The background items have consistently been underdeveloped and thus minimize the opportunity for deeper and more comprehensive analyses, especially regarding trends in performance of different groups of students. This report not only suggests more detailed information on background characteristics but also highlights other family activities in the home shown to be associated with academic performance

that could and should be part of the NAEP background questionnaire. Particularly valuable is that this report makes recommendations regarding items that are replicated and should be deleted. While the specifics may merit further discussion and advice (including from experts within the National Center for Education Statistics [NCES]), the general orientation toward reducing redundancy and pressing for coordination and integration is wise. Perhaps most importantly the report suggests incorporating items from PISA and TIMSS that would enhance the usefulness of NAEP in terms of comparing results with other international surveys. The report also importantly recommends the need for further research—specifically noting the value of a grant program for researchers to extend the analytic properties of NAEP.

Presently NAEP is a descriptive survey of the nation’s academic performance. Several reports have suggested modifications to the sampling design of NAEP, including embedding an individual student longitudinal component that would increase the opportunities for estimating causal effects using observational data from random-sample surveys. The report appropriately cautions researchers that even with an enhanced background questionnaire the present design of NAEP limits what statistical inferences can be estimated. This point is well-taken; however, recent methodological advances may increase the efficiency of estimators that can be used to estimate causal effects. This could be one of the methodological problems that could be explored with grant support. Nevertheless, we continue to underscore the cautions associated with causal analysis with cross-sectional data and self-reported survey questions. This is another area where coordination with NCES could lead to strong strategic decisions and choices consistent with the spirit and objectives set forth in this report.

The report advances recommendations and offers sound counsel and directives related to background questions that merit further follow-up and consideration. Taken as a whole, the report can strengthen NAEP and the value of this survey for monitoring educational progress. AERA appreciates that implementation of many of these recommendations will be challenging in terms of the resource demands of time, money, and expertise, but we hope NAGB will find the means and mechanisms, and where appropriate working closely with NCES, to implement them. We are enthusiastic about the general direction and recommendations in this report; nevertheless, we note below some specific areas where we would urge NAGB to exercise caution in moving forward.

Comments Supporting Specific Recommendations

1. AERA supports the implementation recommendation (5d) to establish a single NAGB committee with responsibility for all background questions, and also supports the functions proposed for this committee with regard to monitoring, developing, and disseminating information about the background questions. It may seem out of sequence that our first point in support of the recommendations is the final recommendation in the report itself. However, such a committee can have major, sustained, and systemic value. AERA testified in favor of a similar strategy for moving

forward with background questions in an oral report at the 2002 hearings sponsored by NAGB. We believe that the issue of background questions is one requiring continuous review and fixed responsibility and that a standing committee of NAGB is a sound approach to providing capacity for the functions envisioned.

2. AERA also supports the major recommendations for providing background questions that are drawn from established research and that are useful for analysis leading to improved understanding of education phenomenon uniquely assessed by NAEP. The report suggests that “core” and “second tier” questions might be introduced and rotated across surveys periodically; that efforts be made to identify clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics; and encourages study of spiral sampling to permit additional questions without adding to time requirements for students.

3. We call attention to the importance of the report recommendation that special reports highlighting background question analysis be produced regularly. In addition to providing the public with richer information about the education status of the nation, such reports and the attending explanations of the limitations of cross sectional, non-longitudinal data bases, will educate the public about its limitations for causal analyses. In essence, the reports could model the sound and productive use of background data for policy makers, members of the media, and researchers.

4. AERA has been leading a grants program (the AERA Grants Program) under the aegis of a Governing Board of leading research scientists since 1989. The AERA Grants Program is dedicated to the analysis of large-scale federal data sets supported in particular by the National Center for Education Statistics (NCES), including NAEP, and the National Science Foundation. The program was begun and continues to advance knowledge, tools, and methods exactly for the reason recommended in the report—to enhance the use of data resources that otherwise are underutilized, especially at the level of their potential. We support the recommendation for NAGB to create a research program with regard to the use and development of background questions as vital to achieving the goals of the report. We encourage consideration of whether a new funding entity needs to be created or whether within IES/NCES or through entities such as the AERA Grants Program the same goals could be effectively and efficiently realized.

5. AERA endorses the report’s recommendation that NAGB and NCES continue to seek ways to develop an improved measure of socio-economic status (SES), either through development of composite items or by generating SES data from information available by linking with other instruments such as the American Community Survey. The recommendation to explore geocoding to aggregate data from the U.S. Census is worth exploring also.

Reliance on SES indicators such as school lunch eligibility is problematic at all times and especially now that the Department of Agriculture is considering modifications in its

policy for providing school lunch. Similarly weak measures of SES have limited researcher's ability to use NAEP data effectively, and AERA welcomes the priority of improved SES questions as central to improved background questions. As the lead federal agency for education statistics working in cooperation with the other federal statistical agencies of the federal government, NCES is and should be an invaluable resource and partner in this regard.

6. Finally, we think it likely that every background question currently in use had a purpose when introduced, but agree that they should be reviewed and modified with an eye toward making them useful in research and analysis. The report has a number of sound specific suggestions for doing so that merit the attention of NAGB. Having participated in some previous NAEP efforts at item development, we are mindful of the weight carried by each question and of the difficult trade offs involved, but also by the high level of expertise of current staff and consultants in performing this task. Nonetheless, we believe that several of the recommendations for additional research-based questions that might suggest explanations for cognitive achievement should be explored by NAGB. We further encourage that such efforts build on the use of cognitive interviewing techniques we believe are in use by NAEP in developing questions.

Comments on Raising Concerns Related to Specific Recommendations

1. We question the soundness of the recommendation to use NAEP to monitor adoption or impact of policy changes such as the Common Core standards. While we encourage NAEP enhancements in order to have policy relevant value, we think it is unnecessary and potentially politically risky for NAEP to be used to monitor implementation or compliance with federal or national reforms. The report fails to provide a compelling reason for seeking this type of information as part of NAEP background questions and this recommendation appears antithetical to goal of developing research-based questions that are theoretically sound. The Institute of Education Sciences is a federal agency that is evaluating the impact of federal intervention programs and is more suited for this work. Additionally, many research organizations and entities outside of government have been effective in conducting such work.

2. We have concerns about the recommendation in 1e to eliminate asking about race/ethnicity in the school questionnaire because it is "obtainable from student records." Self identification and multi-racial classification are important measures that can directly relate to students' perceptions and performance. We urge caution in considering student records on race/ethnicity, especially given the vagaries of how such data may be collected and reported, as a sufficient variable. Dropping such questions from NAEP seems inconsistent with the overall aims of the report in seeking to advance consideration of background attributes.

3. Issues about the misuse of NAEP background questions and assessment data to support causal statements resulting in misinformed policy development are a serious

matter. As noted above, we welcome the caution about over interpretation with regard to causality provided in the report at recommendation 4e, but note that the emphasis is misplaced. The recommendation calls on NAEP to “not publish causal interpretations of the factors determining performance differences based on NAEP data.” The problem of misuse is elsewhere and requires education of educators, media, and policy makers and to some extent also researchers about appropriate inferences from such data rather than admonition to NAEP.

A number of actions proposed in the report will indirectly address this problem (e.g., reports build on background questions, dissemination of information about them, small grants programs), but we believe the recommendation might be amended to include a full description of why NAEP is unsuitable for providing causal interpretation. The illustrations of the value of background questions provided in the report seem ambiguous on the question of causality. For example, the illustration of dramatically different degrees of afterschool math instruction in the south and northeast (p. 10) has no significance unless linked to differences in achievement in the same regions. This is true of many of the illustrations of application of background questions and without further explicit clarifications begs the question of avoiding causal interpretations.

Concluding Comment Looking Ahead

As noted at the outset, we applaud the leadership shown by NAGB in undertaking this investigation of background questions and seeking the advice of an expert panel. The February 2012 report is a valuable contribution. We believe that important directions and recommendations have been provided for improving the quality and use of background questions. We know that this will be steady and challenging work.

As this work is moving forward, we urge that further consideration be given to addressing the major challenges before NAEP as it seeks to be as relevant in the future as in the past. The report makes frequent reference to international studies such as PISA and the alignment of NAEP. Embedded in this important report are broader questions about the next generation of research monitoring and understanding educational progress and how best to undertake it. We urge NAGB to look to the future and consider such issues as: What will be the important and unique contribution of NAEP in 2022? What are the implications for NAEP of the statewide longitudinal data systems (SLDS) and how best to nest the use of administrative data systems with data collections such as NAEP? What are the implications for the development of NAEP of the growth of salient programs in NCES and other agencies, such as the National Children’s Study, or the growth and increased capacity of non-governmental groups engaged in reporting on education through new technologies? Finally, NAGB could usefully address the changed—and dramatically heightened—expectations for data use in education that have developed among policy makers over the past decades.

We anticipate that such an exploration will demonstrate a continued, valued role for NAEP in the U.S. education system. However, we also think it would be useful to consider vehicles (including statutory ones) for strengthening the relationship between NAGB and NCES. The report references activities of NAGB and NCES as if they were freestanding—where collaborations are underway or being urged. We fully recognize that the structure provided by current legislation encourages separation of responsibilities between NAGB and NCES and that there is already a high degree of collaboration in place. Nevertheless, a review of the core purposes of NAEP might suggest that its goals could be realized more effectively and efficiently through closer collaboration with NCES in planning, instrument development, data gathering, and analyses, and, over time, there may be compelling reasons to strengthen this relationship in legislation.

NCES as a federal statistical agency in the United States has a stature and purpose that can benefit NAEP in the year ahead. We urge that NAGB consider how NAEP can benefit from stronger connections to NCES and gain from the multiplier effects of NCES data systems and the work of NCES with other statistical agencies (e.g., Census Bureau, Bureau of Justice Statistics, Bureau of Labor Statistics), with the states and their administrative longitudinal data systems, and with other nations in the further development of international data collections.

The American Educational Research Association appreciates being asked to comment on the report on NAEP background questions. Please call on our Association if we can be of further help in pursuing the objectives and strategies for continuing to work on background questions or for a broader examination of the contributions of NAEP in the coming decades.

Respectfully submitted,

Felice J. Levine, Ph.D.
Executive Director
flevine@aera.net
202-238-3201

Policymakers Weigh Gathering More Data for NAEP

Goal is to improve understanding of performance

By **Erik W. Robelen**

As many experts raise questions about the future of "the nation's report card," the governing board for the assessment program is exploring changes aimed at leveraging the achievement data to better inform education policy and practice.



The core idea, outlined in a **report** to the board, is to expand and make far greater use of the background information collected when the National Assessment of Educational Progress is given. In doing so, the report suggests, NAEP could identify factors that may differentiate high-performing states and urban districts from low performers.

The effort, it says, would parallel the extensive reporting of background variables in global assessment systems, such as the **Program for International Student Achievement**, or PISA.

The report was released just weeks after the Obama administration proposed a fiscal 2013 budget that would cut the NAEP budget by \$6 million, while funding a pilot program of state participation in PISA.

"Currently, the NAEP background questions are a potentially important but largely underused national resource," says the report by a six-member expert panel commissioned by the **National Assessment Governing Board**, or NAGB, which sets policy for the testing program. "These data could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education and the context for understanding achievement and its improvement."

In addition, the report says NAEP background questions could help track policy trends, such as implementation of the Common Core State Standards or new teacher-evaluation systems.

The report, presented this month to NAGB at a meeting in New Orleans, was apparently well-received by many board members, including the chairman, former Massachusetts Commissioner of Education David P. Driscoll. But some of the ideas are generating pushback from current and former federal officials.

"NAGB has a tool that they want to use for everything," said Mark S. Schneider, a former commissioner of the National Center for Education Statistics, the arm of the U.S. Department of Education that administers the test. He argues that NAEP should stick to its core strengths, namely measuring student achievement and serving as a benchmark for state assessments.

"I find this just a distraction," Mr. Schneider said of the proposed plan.

Causation vs. Correlation

Although the report emphasizes the importance of not letting correlations between math achievement and rates of absenteeism, for instance, be confused for causation, Mr. Schneider argues that such distinctions would be lost on the public and risk damaging NAEP's reputation.

"They will make statements that will inevitably push the boundaries, and you will end up with questionable reports, in my opinion," said Mr. Schneider, who is now a vice president of the Washington-based American Institutes for Research. Other concerns raised about the proposals are the cost involved, especially given the president's proposed cut to NAEP, and what some experts say may be resistance to the federal government's collection and reporting of more information on students, given privacy concerns.

The new report, commissioned by NAGB, notes that complementing the NAEP tests is a "rich collection" of background questions regularly asked of students, teachers, and schools. But the collection and the public reporting of such information have been significantly scaled back over the past decade, the report says.

"NAEP should restore and improve upon its earlier practice of making much greater use of background data," the report says, "but do so in a more sound and research-supported way."

It offers recommendations in four areas related to the background questions: asking "important questions," improving the accuracy of measures, strengthening sampling efficiency, and reinstating what it calls "meaningful analysis and reporting."

It's the fourth area, analysis and reporting, that is proving especially controversial.

Marshall S. "Mike" Smith, a co-author of the report and a former U.S. undersecretary of education in the Clinton administration, notes that the report comes at a time when NAEP's long-term relevance is at issue. He cites the work to develop common assessments across states in English/language arts and mathematics, as well as the growing prominence of international exams, like PISA.

"The future of NAEP is somewhat in doubt," Mr. Smith said.

PISA's use of extensive background questions, he said, has enabled it to have wide influence.

"They've built narratives around the assessments: Why are there differences among countries" in achievement, he said. "We can't do that with NAEP. We're not able to construct plausible scenarios or narratives about why there are different achievement levels among states. And we've seen that can be a powerful mechanism for motivating reform."

Mr. Driscoll, the chairman of NAGB, said the next step is for board staff members to draft recommendations on how the proposed changes could be implemented.

"I have challenged the board to think about how NAEP and NAGB can make a difference and have an impact," he said. "There is some very valuable information that we can lay out ... that would be instructive for all of us."

The report makes clear that NAEP should not be used to assert causes for variation in student achievement, but that a series of "descriptive findings" could be illustrative and help "generate hypotheses" for further study. For example, it might highlight differences in access to 8th grade algebra courses or to a teacher who majored in math.

"A valid concern over causal interpretations has led to a serious and unjustified overreaction," the report says.

But some observers see reason for concern.

"It's a mistake to present results that are purely descriptive," said Grover J. "Russ" Whitehurst, a senior fellow at the Brookings Institution in Washington who was the director of the federal Institute of Education Sciences under President George W. Bush. "It is misleading, and it doesn't make any difference if you have a footnote saying these results should not be considered causally."

Jack Buckley, the current NCES commissioner, expressed reservations about some of the suggestions, especially in the analysis and reporting of the background data.

"The panel is looking toward PISA as an exemplar," he said. "Folks at [the Organization for Economic Cooperation and Development, which administers PISA] write these papers and get a broad audience, but it's not always clear that the data can support the conclusions they reach about what works."

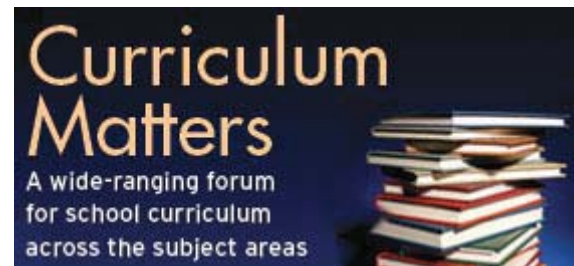
Mr. Buckley said he understands NAGB's desire to be "policy-relevant," but he cautioned that "we have to carefully determine what is the best data source for measuring different things."

Mr. Driscoll said he's keenly aware of not going too far with how the background data are used.

"I agree ... that we have to be careful about the causal effects," he said. "I think we've gone too far in one direction to de-emphasize the background questions, and the danger is to go too far in the other direction."

Vol. 31, Issue 24, Pages 20-21

RELATED BLOG



[Visit this blog.](#)

The Hidden Gem of NAEP: Contextual Variables

Laura Egan, Westat NSSC

Paula Hutton, Maine NAEP State Coordinator

Angie Mangiantini, Washington NAEP State Coordinator

Jan Martin, South Dakota NAEP State Coordinator

Paul Stemmer, Michigan NAEP State Coordinator

CCSSO NCSA

June 29, 2012

Session Overview

- NAEP Contextual Data: Background and Overview
- Using NAEP Contextual Data at the State Level: Effective Practices and Lessons Learned
 - Raising the Rigor in Maine
 - STEM Education and Science in Washington
 - A Tale of Two Issues: Using Contextual Variables to Inform Different Audiences in South Dakota
 - Cognitive Mediational Strategies and Student Performance: A Cautionary Tale from Michigan

NAEP Contextual Data: An Overview

- Non-cognitive items included in student, school, and teacher questionnaires
- Focus on a variety of topics related to the context of learning
 - General and subject-specific items
 - Some stable items
 - Some items are added/removed based on timely topics

NAEP Contextual Data: An Overview

- History:
 - Limited contextual data was collected in early NAEP
 - gender, race/ethnicity, literacy materials at home
 - Expansion in 1980s
 - define a more extensive array of subgroups of the student population for reporting purposes
 - inform educational policy by describing the contexts for learning, sometimes called opportunities to learn
 - support research into factors that may be related to student achievement

NAEP Contextual Data: An Overview

- Statutory Requirements
 - The 2001 ESEA Reauthorization (P.L. 107-110) requires NAEP to collect information on gender, race/ethnicity, socio-economic status, disability, and limited English proficiency. It must report test data on these groups, whenever feasible, that is cross-tabulated, compared, and reported according to the categories required.
 - NAEP may only collect information that is “directly related to the appraisal of academic achievement”
 - concentrate on non-cognitive variables that are known from other research to have such a relationship
 - NAEP cannot ask about personal or family beliefs and attitudes

Background Information Framework for the National Assessment of Educational Progress, 2003

NAEP Contextual Data: An Overview

Three types of background data :

1. General Student Reporting Categories
 - Since the first NAEP assessment in 1969, achievement results have been disaggregated by subgroups of the population
 - Gender
 - Race/ethnicity,
 - SES - parental education, literature in home
 - Type of school location: disadvantaged urban, advantaged urban, and rural

NAEP Contextual Data: An Overview

Three types of background data :

1. General Student Reporting Categories (cont.)
 - NAEP expanded these after 2002 and now reports on a broader range of subgroups, such as
 - Eligibility for free/reduced-price lunch
 - SD or ELL status
 - Type of location: urban, suburban, town, rural
 - Type of school: public, private, Catholic, other private

NAEP Contextual Data: An Overview

Three types of background data :

2. Contextual/Policy Information

- In every assessment, NAEP collects data on basic characteristics of the school and student body in the school; teacher background, qualifications, and experience; and several student characteristics. These variables provide a basic context for achievement.
- Timely policy/contextual issues may be rotated across assessments.

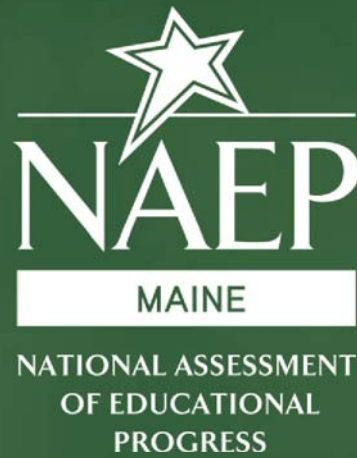
NAEP Contextual Data: An Overview

3. Subject-Specific Information

- The subject-specific items in NAEP are focused and limited. A set of key issues within each subject area will be addressed in a focused and in-depth manner across the life of each assessment framework.
- When a new assessment framework is approved, NCES reviews the recommendations for background data made by the framework committee.

NAEP Contextual Data: Process

1. NAGB oversees the development of framework and item specifications for the background items.
2. NCES develops an issues paper to reflect the new priorities, identify the data needed to address the issues, and propose an item rotation plan. NCES, in conjunction with a panel of experts, develops an issues paper.
3. NAEP contractors draft and revise background items.
4. NCES reviews the background items.
5. The items are piloted, and the results are analyzed.
6. Based upon pilot data results, some items are revised.
7. The background items once again undergo reviews by item development contractors and then by NCES.
8. NCES presents items to the Governing Board for its approval.
9. The items are submitted for clearance by NCES to the Office of Management and Budget.
10. Once clearance is received, each background item is typeset into the respective student, teacher, school, and/or SD/ELL questionnaires.



Raising the Rigor in Maine

Raising the Rigor in Maine

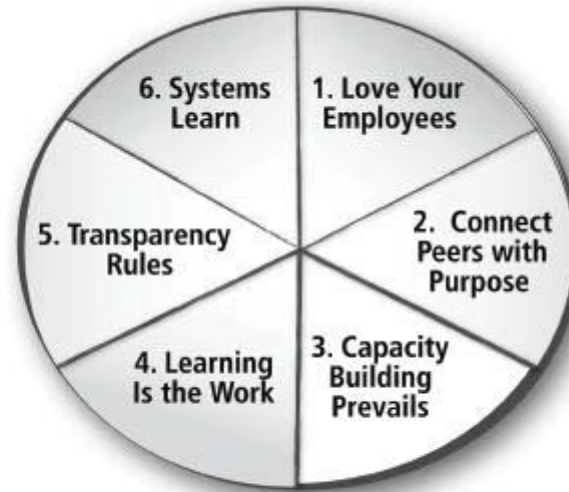
What is Change?

(according to Michael Fullan)

- New materials
- New behaviors/practices
- New beliefs/understanding

•Process is Key

The Six Secrets of Change



— Fullan, 2008a

•Looked at the contextual data on the NAEP Data Explorer to develop questions

5									
6	2011	Feel math classwork is too hard	Never, hardly ever, sometimes	87%	Often, always, almost always			13%	
7									
8	2011	Feel math work is too hard	Never, hardly ever, sometimes	87%	Often, always, almost always			13%	
9	2009	Feel math work is too hard	Never, hardly ever, sometimes	87%	Often, always, almost always			12%	
10									
11	2011	Like what is done in math class?	Never, hardly ever, sometimes	43%	Often, always, almost always			56%	
12	2009	Like what is done in math class?	Never, hardly ever, sometimes	44%	Often, always, almost always			56%	
13									
14	2011	Feel classwork is too easy?	Never, hardly ever	20%	Sometimes			51%	Often, almost always 29%
15	2009	Feel classwork is too easy?	Never, hardly ever	18%	Sometimes			52%	Often, almost always 30%
16									
17	2011	Like what is done in math class?	Never, hardly ever, sometimes	44%	Often, almost always, always			56%	
18	2009	Like what is done in math class?	Never, hardly ever, sometimes	43%	Often, almost always, always			56%	
19									
20				2011				2009	
21		Do a good job on math tests?	Never, Hardly ever	4%	Never, Hardly ever			5%	
22			Sometimes	24%	Sometimes			26%	
23			Often	33%	Often			32%	
24			Always, almost always	39%	Always, almost always			37%	
25									
26	Q1	If 68-71% of the students indicate that they feel they can do a good job on math assignments, does that indicate about rigor?							
27	Q2	If 87% of the students say that the classwork is never, hardly ever or sometimes too hard, what's the implication for the rigor of the classwork?							
28	Q3	What does could it mean when 43-44% of students say they never, hardly ever or just sometimes like what they do in math class?							
29	Q4	What does it mean with regard to mass customized learning when 29-30% find the math work too easy?							
30	Q5	What does it mean with regard to mass customized learning when 28% - 31% say they never, hardly ever or sometimes do a good job on math tests?							
31									

•NAEP Contextual Data in Mathematics (A look at student responses.)

In Maine, I looked at grade 4 and grade 8 student responses in the area of difficulty of the mathematics work being asked of students. For both grades I looked at questions that were asked of students in the 2009 and 2011 NAEP background surveys.

Grade 4 Questions

- How often do you feel you can do a good job on math assignments?
- Do you feel math class work is too hard?
- Do you feel math work is too hard?
- Do you like what is done in math class?
- Do you feel class work is too easy?

Based on this information, these are the queries I would share with teachers of grade 4 students:

1. If 68-71% of students indicate that they feel they can do a good job on math assignments, what does that indicate about the rigor of the work?
2. If 87% of the students say that the class work is never, hardly ever or only sometimes too hard, what's the implication for the rigor of the class work?
3. What could it mean when 43-44% of students say they never, hardly ever or just sometimes like what they do in math class?
4. What does it mean with regard to Mass Customized Learning when 29-30% find the math work too easy?
5. What does it mean with regard to Mass Customized Learning when 28-31% of students say that they never, hardly ever, or sometimes do a good job on math tests?

•NAEP Contextual Data Questions in Mathematics (A look at teacher responses.)

I looked at grade 4 and grade 8 teacher responses in the areas of amount of time for class instruction in mathematics, amount of daily homework assigned, emphasis on domains, ability grouping for classes, and computer use/availability.

I also looked to the Common Core State Standards for Mathematics to see the emphasis on domains in grades 4 and 8. Note: There are domains that begin and end in the Common Core State Standards as well as different domain titles at the different grades. These are changes from our current Maine Learning Results. These changes/differences should be kept in mind as teachers begin their work with aligning current curriculum to the Common Core State Standards. |

Numbers and Operations

- Why is it teachers in both grades 4 (100%) and 8 (98%) have a moderate to heavy emphasis on Numbers and Operations?
- Common Core State Standards has a moderate to heavy emphasis on Numbers and Operations in grade 4, but in grade 8 there is little emphasis on this domain...
- Does this emphasis reflect in student scores? Are we seeing the results expected with this level of emphasis on the domain?
- What does this imply for 8th grade teachers as they work to implement the Common Core State Standards in their classrooms?



Purpose and Design of the NECAP Mathematics Assessment

Distribution of Emphasis for Mathematics

Mathematics Content Standards (GLEs)	Grade 2(3)*	Grade 3(4)	Grade 4(5)	Grade 5(6)	Grade 6(7)	Grade 7(8)
Numbers and Operations	55%	50%	45%	40%	30%	20%
Geometry and Measurement	15%	20%	20%	25%	25%	25%
Functions and Algebra	15%	15%	20%	20%	30%	40%
Data, Statistics and Probability	15%	15%	15%	15%	15%	15%

* 2(3) indicates end-of-grade 2 GLEs are tested at the beginning of grade 3



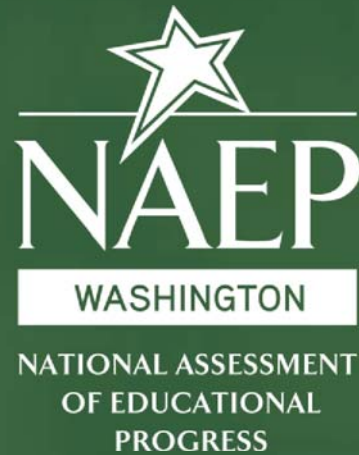
Purpose and Design of the NECAP Mathematics Assessment

Distribution of Emphasis for Mathematics						
Mathematics Content Standards (GLEs)	Grade 2(3)*	Grade 3(4)	Grade 4(5)	Grade 5(6)	Grade 6(7)	Grade 7(8)
Numbers and Operations	55%	50%	45%	40%	30%	20%
Geometry and Measurement	15%	20%	20%	25%	25%	25%
Functions and Algebra	15%	15%	20%	20%	30%	40%
Data, Statistics and Probability	15%	15%	15%	15%	15%	15%

* 2(3) indicates end-of-grade 2 GLEs are tested at the beginning of grade 3



Maine
Department of
Education



STEM Education and Science in Washington



Science & STEM Education in Washington State: Background & Current Landscape



Randy L. Dorn
State Superintendent of
Public Instruction

July 2011

Purpose

The purpose of this report is to describe the current state of science and STEM education in Washington State and the policies and programs supporting science and STEM education. The intent is to inform decision makers, educators, and the public on potential funding and programs to support a statewide system of science education. The majority of the information for this paper was collected between August and October 2010.

Question:

How much time do teachers spend on Science Instruction by state?

NAEP DATA USED FOR REPORT:

- *Responses from the Teachers Background Questionnaire on time spent teaching Science in Grade 8 and Grade 4 as compared to other jurisdictions*
- Scale Score Comparisons by State and Grade Level
- Disaggregated data by ethnicity and poverty

Date obtained from the NAEP Data Explorer <http://nces.ed.gov/nationsreportcard/naepdata/>

MAIN NDE 1. Select Criteria 2. Select Variables 3. Edit Reports 4. Build Reports

STEP 4: View each report table by selecting the report name from the drop-down menu. Create report types to edit and preview, each tab created represents one report type to export. Double-click report tabs to rename. NDE Help

Subject, Grade: Science, Grade 4
Jurisdictions: National public, Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming, DoDEA
Measure: Composite scale
Variable: Time per week on science instruction
Year: 2005

Select Report: Report 1 Export Reports

Table Chart Significance Test Gap Analysis

Average scale scores and percentages for science, grade 4 by year, jurisdiction and About how much time in total do you spend with this class on science instruction in a typical week? (teacher-reported) [T090101]: 2005

Year	Jurisdiction	Less than 1 hour		1-1.9 hours		2-2.9 hours		3-3.9 hours		4 hours or more	
		Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage
2005	National public	140	6	144	17	150	33	151	26	154	18
	Alabama	130	3	135	6	140	20	145	32	144	39
	Arizona	129	12	141	28	140	32	143	20	147	7
	Arkansas	145	6	142	17	146	30	147	30	157	18
	California	128	13	136	33	141	33	139	15	138	5
	Colorado	152	10	153	20	155	29	158	29	159	12
	Connecticut	151	3	152	24	156	45	154	20	158	8
	Delaware	#	1	150	6	153	32	152	37	154	24
	Florida	144	7	148	17	149	37	153	22	155	17
	Georgia	141	3	139	7	147	27	148	33	150	30
	Hawaii	138	17	139	33	148	33	144	13	142	4
	Idaho	151	15	155	34	156	34	160	14	152	4
	Illinois	#	1	129	8	147	29	152	42	151	20
	Indiana	142	7	148	16	153	44	154	25	159	8
	Iowa	—	—	—	—	—	—	—	—	—	—
	Kentucky	#	1	#	2	#	1	162	13	158	82
	Louisiana	#	2	140	5	129	11	144	38	145	44
	Maine	#	1	160	16	160	39	160	34	163	9

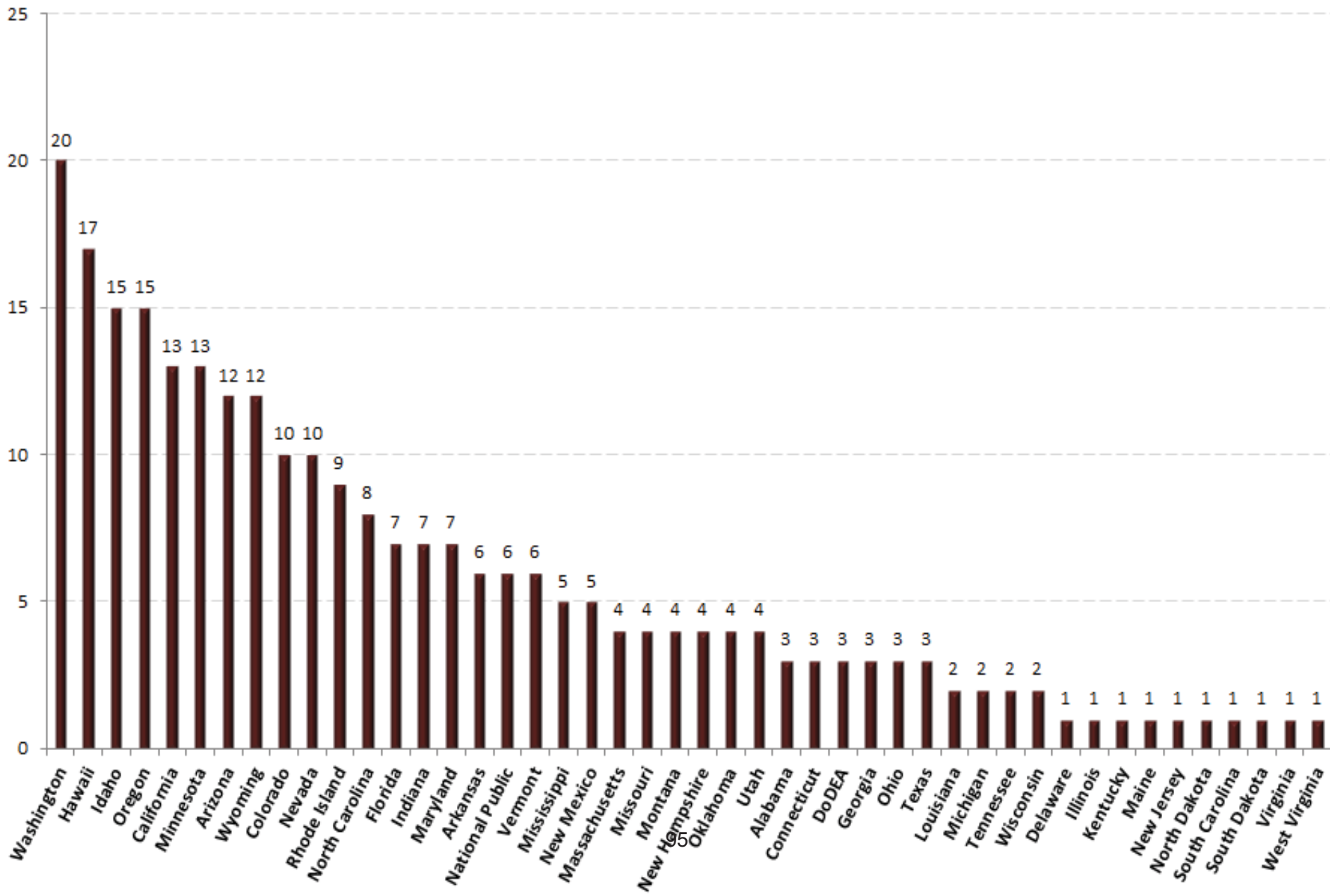
A guide (handout) for replicating the data for the study from the NAEP Data Explorer is available.

NAEP GRADE 4 SCIENCE 2005

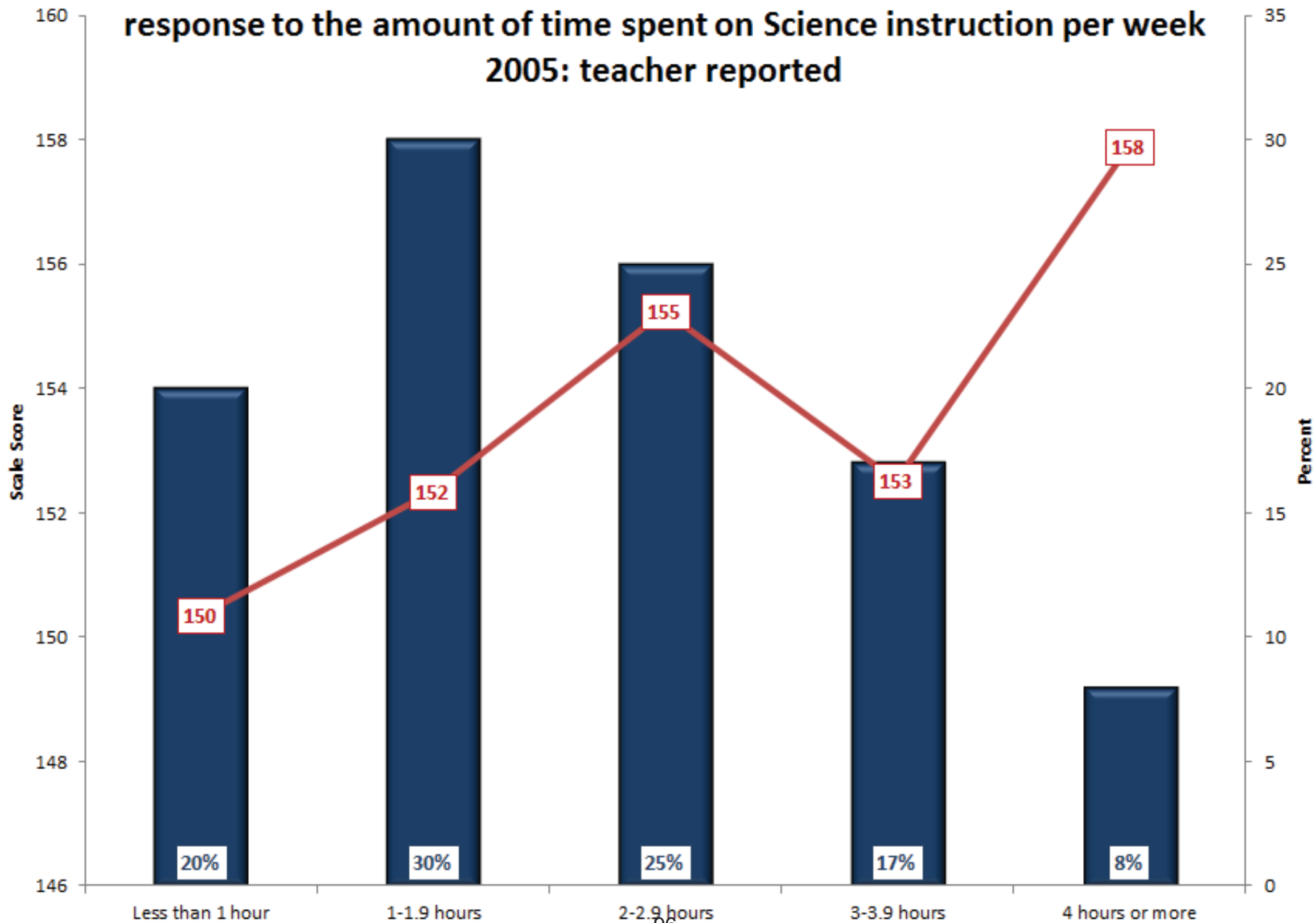
About how much time in total do you spend with this class on science instruction in a typical week? (teacher-reported) : 2005

Year	Jurisdiction	Less than 1 hour		1-1.9 hours		2-2.9 hours		3-3.9 hours		4 hours or more		2005 Average Scale Score
		Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage	Average scale score	Percentage	
2005	Alabama	130	3	135	6	140	20	145	32	144	39	142
	Arizona	129	12	141	28	140	32	143	20	147	7	139
	Arkansas	145	6	142	17	146	30	147	30	157	18	147
	California	128	13	136	33	141	33	139	15	138	5	137
	Colorado	152	10	153	20	155	29	158	29	159	12	155
	Connecticut	151	3	152	24	156	45	154	20	158	8	155
	Delaware	+	1	150	6	153	32	152	37	154	24	152
	DoDEA	+	3	157	13	157	34	157	33	156	17	156
	Florida	144	7	148	17	149	37	153	22	155	17	150
	Georgia	141	3	139	7	147	27	148	33	150	30	148
	Hawaii	138	17	139	33	148	33	144	13	142	4	142
	Idaho	151	15	155	34	156	34	160	14	152	4	155
	Illinois	+	1	129	8	147	29	152	42	151	20	148
	Indiana	142	7	148	16	153	44	154	25	159	8	152
	Kentucky	+	1	+	2	+	1	162	13	158	82	158
	Louisiana	+	2	140	5	129	11	144	38	145	44	143
	Maine	+	1	160	16	160	39	160	34	163	9	160
	Maryland	142	7	146	17	145	28	151	24	157	24	149
	Massachusetts	147	4	159	29	161	40	163	21	160	6	160
	Michigan	+	2	140	6	152	27	153	40	155	24	152
	Minnesota	153	13	152	25	159	36	162	18	158	7	156
	Mississippi	132	5	131	16	136	31	138	26	128	22	133
	Missouri	156	4	157	12	155	41	160	29	159	14	158
	Montana	146	4	157	22	161	42	163	25	159	6	160
	National Public	140	6	144	17	150	33	151	26	154	18	149
	Nevada	136	10	138	32	142	37	143	16	136	5	140
	New Hampshire	155	4	162	23	160	39	163	26	164	8	161
	New Jersey	+	1	154	15	156	34	153	34	154	17	154
	New Mexico	135	5	141	19	142	41	143	26	136	10	141
	North Carolina	149	8	145	22	150	41	152	22	145	6	149
	North Dakota	+	1	150	4	158	29	160	48	164	18	160
	Ohio	136	3	157	7	153	25	155	40	163	25	157
	Oklahoma	144	1	147	10	151	24	151	26	150	10	150

Percentage of Grade 4 students having teachers spending less than 1 hour per week on Science instruction by state (2005)



Percent and Scale Scores of Grade 4 students based on teachers' response to the amount of time spent on Science instruction per week 2005: teacher reported



Reporting Contextual Variables

Draft Version of Report

Time Spent Teaching Science

A comprehensive survey regarding time spent teaching science has not been conducted for Washington State. However, according to a 2005 NAEP survey on the percentage of time spent on science in 4th grade, **20% of Washington 4th grade teachers surveyed reported teaching less than one hour of science per week** (this was the highest proportion of the 45 states surveyed). Only 8% of the 4th grade teachers reported spending more than four hours per week teaching science. Clearly, elementary students are not receiving adequate time for science instruction.

¶
According to the NAEP survey, for 8th grade science teachers surveyed 4% reported teaching science 1-2.9 hours per week; 62% reported teaching science 3-4.9 hours per week; 17% reported teaching science 5-6.9 hours per week; and 16% reported teaching science 7 or more hours per week. Because science is taught by subject area teachers in middle and high school, time spent teaching science at the secondary level is likely not a significant factor in students' opportunity to learn science.

Comment [am1]: Twenty-percent of fourth-grade students taking the 2005 NAEP Science assessment had teachers who reported teaching Science less than one hour per week. Only eight percent of fourth grades had teachers reporting spending more than four hours per week teaching science.

Comment [am2]: According to the NAEP survey, 4% of the eighth graders tested had teachers reporting they taught science 1-2.9 hours, 62% of the eighth grade students had teachers reporting teaching science 3-4.9 hours per week, 17% of the students had teachers who taught science 5-6.9 hours per week, and 16% of the eighth graders had teachers providing science instruction 7 or more hours per week.

FINAL REPORT

Time Spent Teaching Science

A comprehensive survey regarding time spent teaching science has not been conducted for Washington State. However, **twenty percent of fourth grade students taking the 2005 National Assessment of Education Progress (NAEP) Science assessment had teachers who reported teaching science less than one hour per week.** Of the 45 states surveyed, Washington had the highest proportion of teachers reporting teaching less than one hour of science per week. Only eight percent of the fourth grade teachers reported spending more than four hours per week teaching science. Clearly, relative to other states, Washington's students are receiving significantly less science instruction in fourth grade and possibly in other elementary grades.

ISSUE IDENTIFIED FROM REPORT BASED ON CONTEXTUAL VARIABLE:

In preparing this report a number of issues were identified that warrant further consideration:

- Time spent on science in elementary grades.

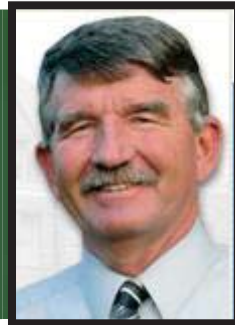
Resulting Actions:



Washington to Lead Effort to Develop New Science Standards

Twenty states will work together to develop the Next Generation Science Standards

"Since I took office, I have stressed the need for more quality science instruction at every grade, not just the grades where we test science as a state. We need to hire teachers with science backgrounds and place the same emphasis on science education that we do with reading, writing and math, especially in elementary school."



Randy Dorn: State Superintendent of Public Instruction



A Tale of Two Issues: Using Contextual Variables to Inform Different Audiences in South Dakota

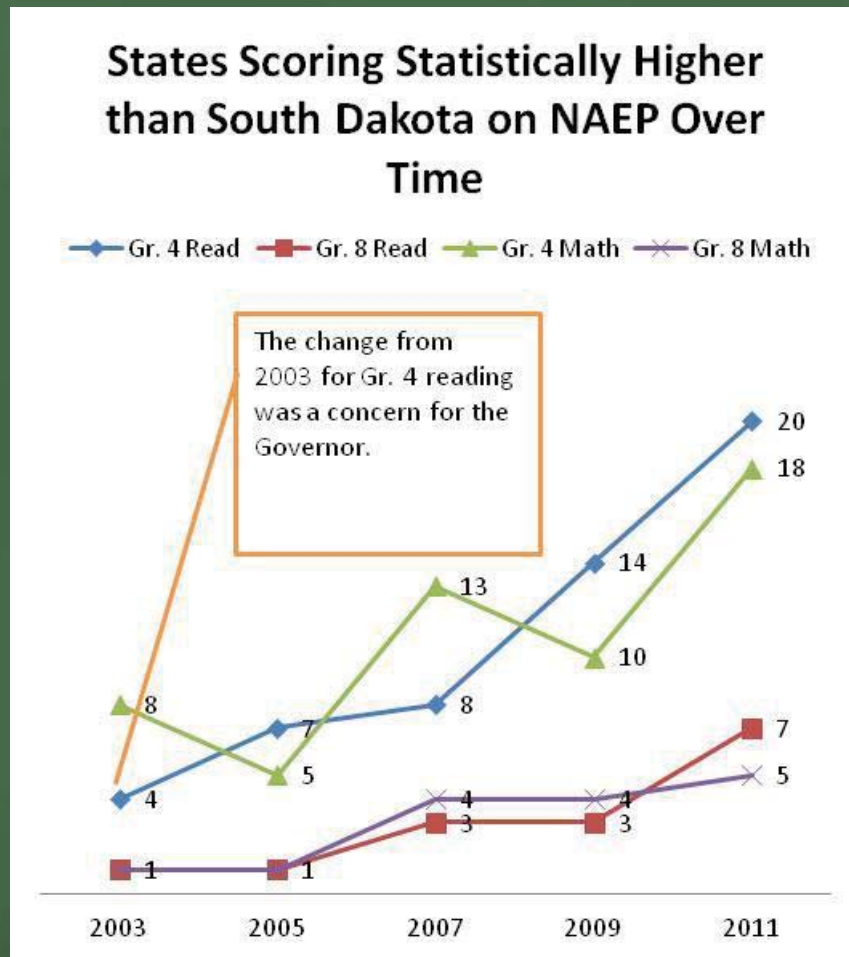
•A tale of two issues – using contextual variables to inform different audiences in South Dakota

Issue: Grade 4 Reading Scores declining

- Focus on demographics and school variables
- Governor honed in on the grade 4 decline as part of his education reform platform
- Inherent bias among administrators in the state regarding cause of decline
- Investigation into the school variables revealed several issues

- *Issue:* Implementation of Common Core ELA standards in grade 9-12
- Grade 8 Reading Data to inform practice
- School, teacher and student perceptions of practice
- In 2011, all public schools with 8th grade part of the state sample. School and teacher data provides a comprehensive overview of current status in state
- Data will be used in professional development for grades 7-12

•Grade 4 Reading Decline



• Response to Governor's concern

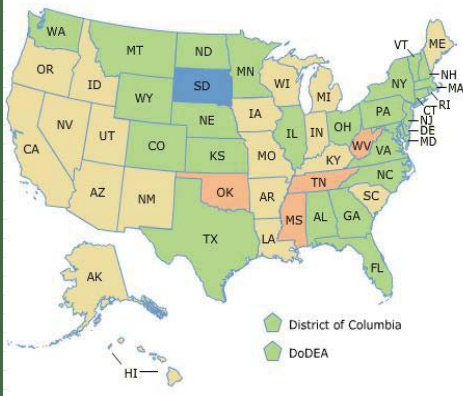
- Was it the Native American students?
- Was it special education?

While the gap between White and Native American students is greater than the national gap, that is not the whole story.

Title I status and location as critical as race/ethnicity gaps.

•Grade 4 Reading Investigation – Race/Ethnicity

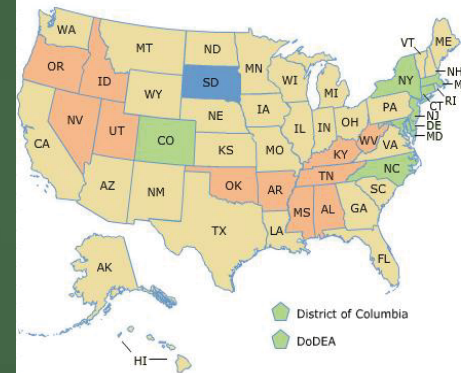
Reading, grade 4
Difference in average scale scores between jurisdictions, for race/ethnicity used to report trends, school-reported [SDRACE] = White
2011



2011 Gr. 4 White Students

2003 Gr. 4 White Students

Reading, grade 4
Difference in average scale scores between jurisdictions, for race/ethnicity used to report trends, school-reported [SDRACE] = White
2003



•Grade 8 Literacy Study

How are teachers structuring English Language Arts curriculum in grade 8?

- As part of baseline data for a literacy integration project at the middle and high school levels, NAEP data were analyzed to create a snapshot of the state of language arts in South Dakota's middle schools.
- NAEP survey data from students, teachers and administrators were used to provide the context to better understand the need to integrate literacy strategies across the curriculum, in particular for secondary Career and Technical Education courses.
- A major finding was the need for greater emphasis/integration of Informational Reading across the curriculum.
- Another finding of the study was that even in the English Language Arts classrooms, there is often a mismatch between students and teachers understanding what is happening instructionally .
- These data will be used as part of literacy integration workshops.
- The focus was on state data using percentages so teachers could think about where they are instructionally.

•Gr. 8 Curriculum Structure – Administrator Perception

To what extent is your school's English/language arts program structured according to the following resources?

	Not at all	Small Extent	Moderate extent	Large extent
State standards or frameworks	1%	0	2%	97%
District standards or frameworks	4%	5%	5%	71%
Results from state/district assessments	5%	5%	25%	65%
In-school frameworks and standards	18%	22%	27%	33%
Results from school assessments	2%	12%	42%	44%
Recommendations from reading/lang arts department	5%	27%	35%	33%
Discretion of individual teachers	4%	31%	50%	14%
Commercially designed programs	28%	44%	20%	8%

How are the state standards aligned to the NAEP Framework ?

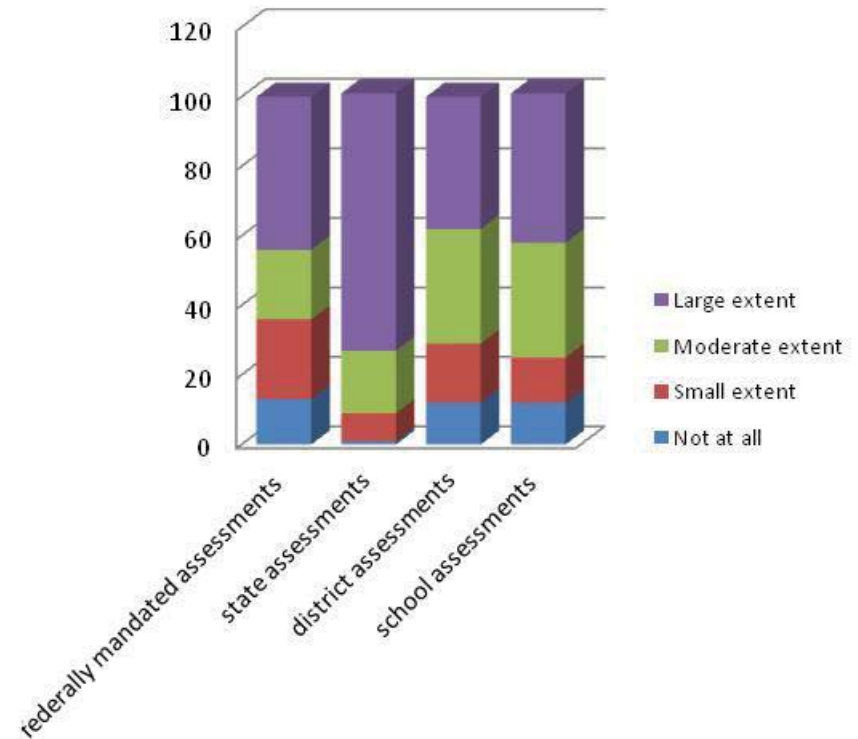
In conversations with teachers, their perception is that they must follow the textbooks which is different than the administrators' perception of what is happening in the classroom.

Administrators' perception of assessment focus

To what extent does your school's 8th grade English/language arts curriculum focus on preparation for the following types of assessments?

	Not at all	Small extent	Moderate extent	Large extent
federally mandated assessments	13	23	20	44
state assessments	1	8	18	74
district assessments	12	17	33	38
school assessments	12	13	33	43

Focus on preparing for assessments



•Teacher - Instructional Emphasis

	Not at all	Small extent	Moderate extent	Large extent
Fiction	1%	2%	46%	51%
Literary nonfiction	3%	17%	61%	19%
Poetry	8%	42%	41%	10%
Exposition	5%	35%	50%	10%
Argumentation & persuasion	15%	52%	28%	5%
Procedural texts and documents	14%	59%	24%	4%

The NAEP Reading Framework for grade 8 divides text type 45% Literary Texts and 55% Informational Texts.

- ✓ Not surprising that the greatest emphasis was on fiction with some attention paid to literary nonfiction and exposition.
- ✓ The lack of emphasis on procedural texts supports the need for more work on integrating literary skills across the curriculum.
- ✓ The data supports the need for more work across the state to assist schools so they can incorporate Informational Texts into the curriculum as CCSS for ELA is implemented.

Teacher – Instructional Practices

To what extent have you emphasized the following cognitive processes when reading informational and literary texts in class?

	Not at all	Small extent	Moderate extent	Large extent
Locate/Recall	1%	21%	42%	36%
Integrate/Interpret	1%	9%	50%	40%
Critique/Evaluate	1%	18%	53%	28%

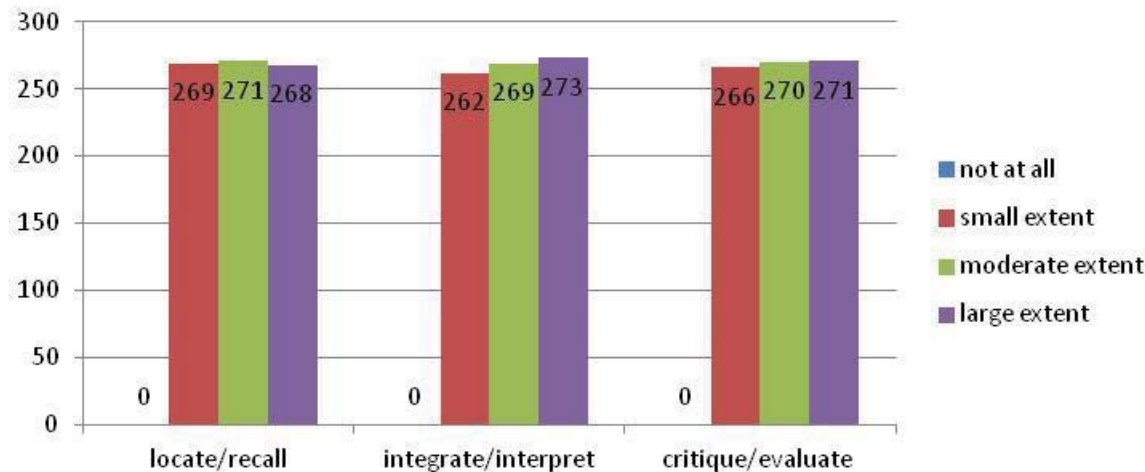
NAEP Reading Framework:

20% Locate/recall

50% Integrate/Interpret

30% Critique/Evaluate

Average Scale Scores across Cognitive Processes



•Teacher – Instructional Practices

	Never or hardly ever	Once or twice a month	Once or twice a week	Everyday or almost everyday
ask students to read aloud	14%	26%	39%	21%
ask students to write about something they have read	1%	25%	46%	28%
give students time to read books of own choosing	3%	11%	49%	38%
group activity or project about what they have read	10%	60%	24%	6%
ask students to explain or support understanding of what read	#	11%	49%	40%
watch movies, videos; or listen to tapes, compact discs	37%	52%	9%	2%
ask students to make predictions about what they have read as they are reading	#	16%	47%	36%

•Student Perceptions of Reading Instruction

	Never or hardly ever	Once or twice a month	Once or twice a week	Every day or almost every day
Read aloud in class	22%	27%	31%	20%
Read books of own choosing in class	22%	20%	25%	33%
Discuss interpretation of what read	31%	32%	27%	10%
Discuss vocabulary in class	10%	23%	45%	23%
Group projects about what was read in class	29%	46%	20%	5%
Explain what was read in class	16%	24%	36%	24%
Read silently in class	6%	14%	39%	40%
Write about what was read	19%	38%	31%	12%
Identify main themes of passages	12%	29%	39%	20%

•Comparing Student and Teacher Perceptions about Instructional Practices

	Teachers	Students
Read aloud in class	50%	51%
Read books of own choosing in class	87%	58%
Write about what is read	84%	43%
Explain what was read	89%	60%

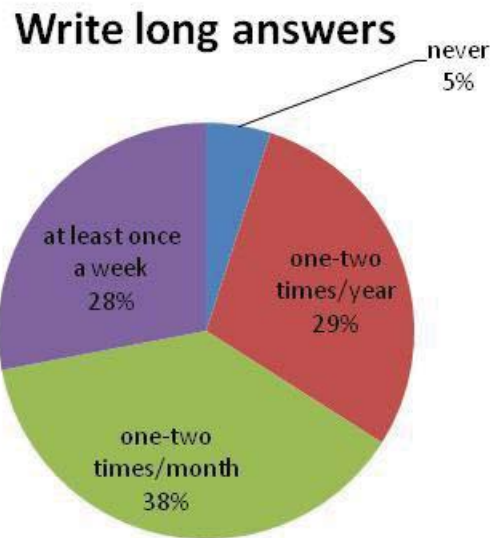
Table combines once or twice a week with daily or almost every day – the activities occur at least once a week.

•Students – How much do they read?

Pages read in school and for homework		
	% of students	average scale score
5 or fewer	20%	263
6 -10	22%	267
11-15	17%	271
16-20	14%	273
more than 20	26%	272

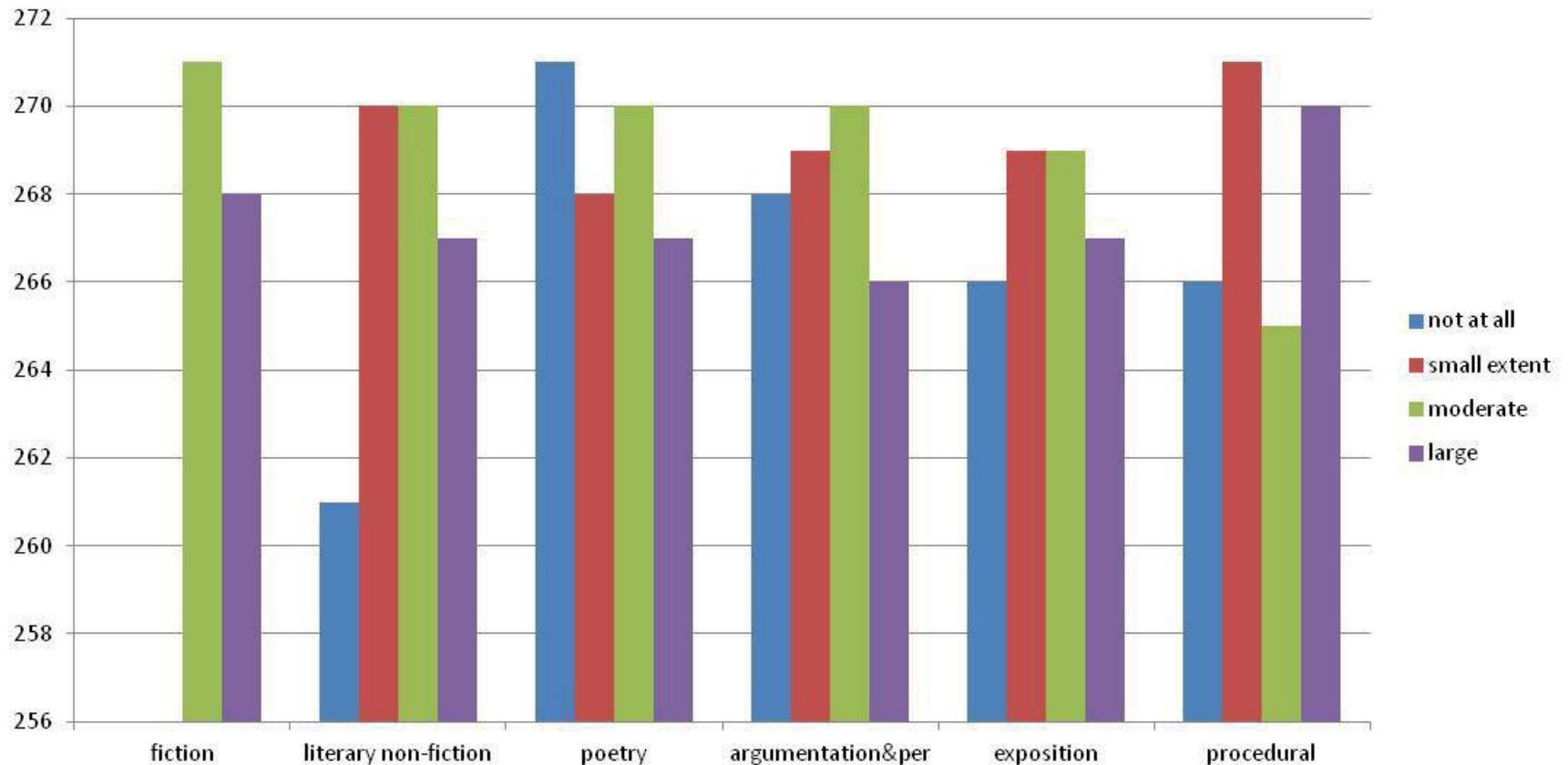
Students who reported having to read 11 or more pages a day had significantly higher scale scores than the students reading 10 pages or less.

Students – Writing Long Answers on Tests



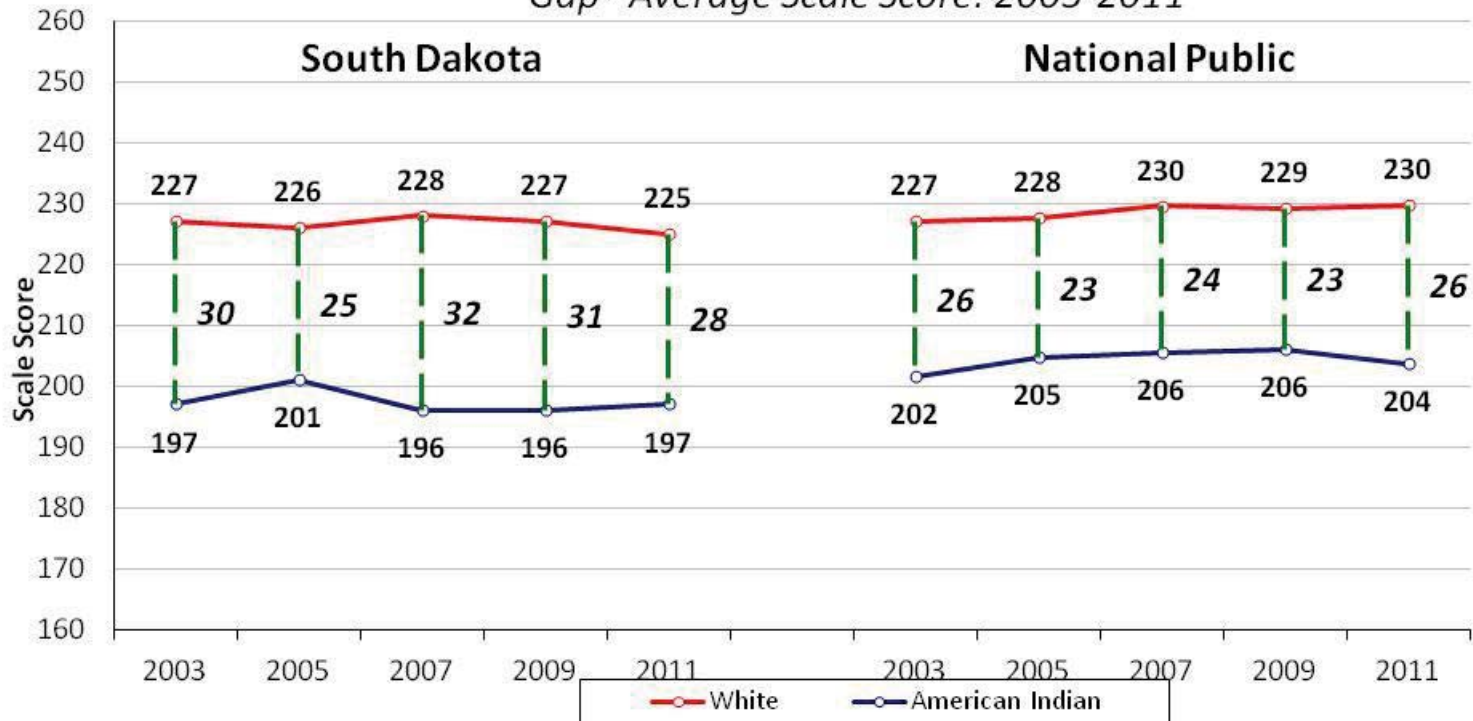
Frequency	average scale score
never	250
one-two times/year	266
one-two times/month	273
at least once a week	270

Distribution of Scale Scores by Instructional Emphasis



NAEP Reading Grade 4 – White - American Indian

Gap - Average Scale Score: 2003-2011

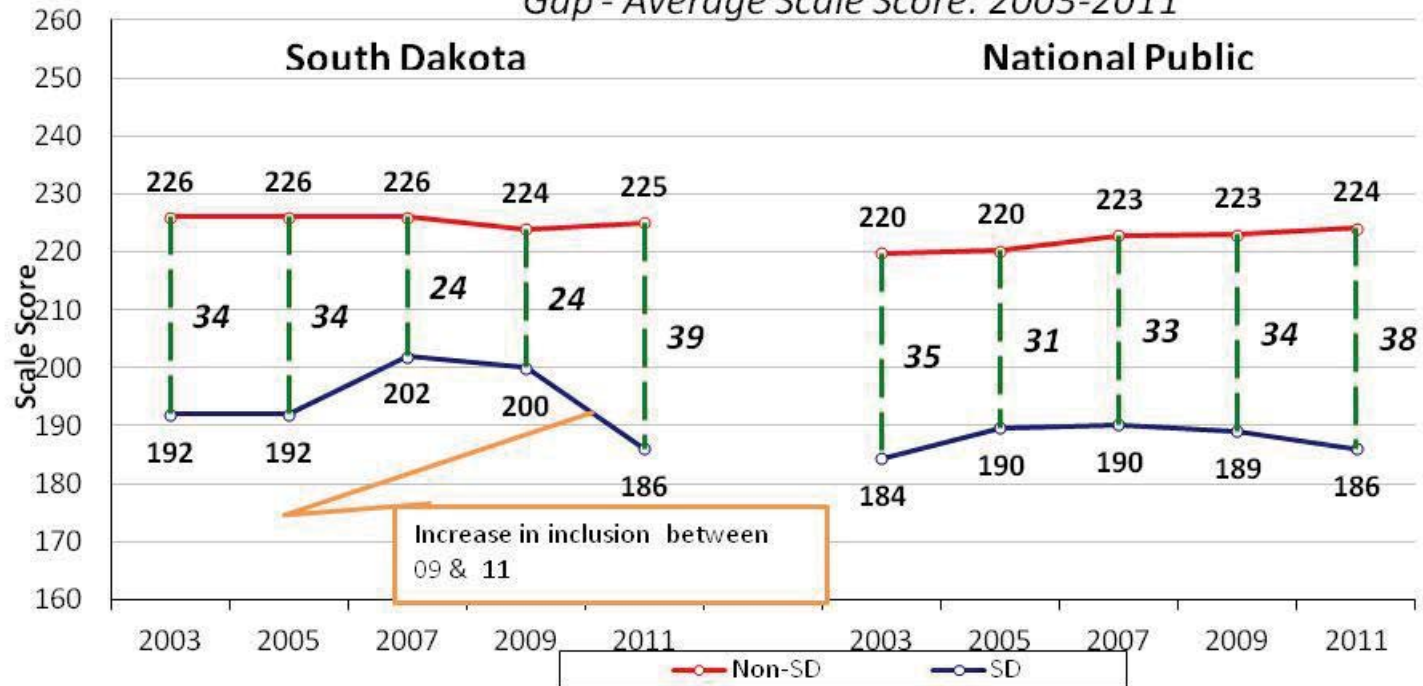


NOTE: The NAEP Reading scale ranges from 0 to 500. Observed differences are not necessarily statistically significant.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress

NAEP Reading Grade 4 – Students with Disabilities

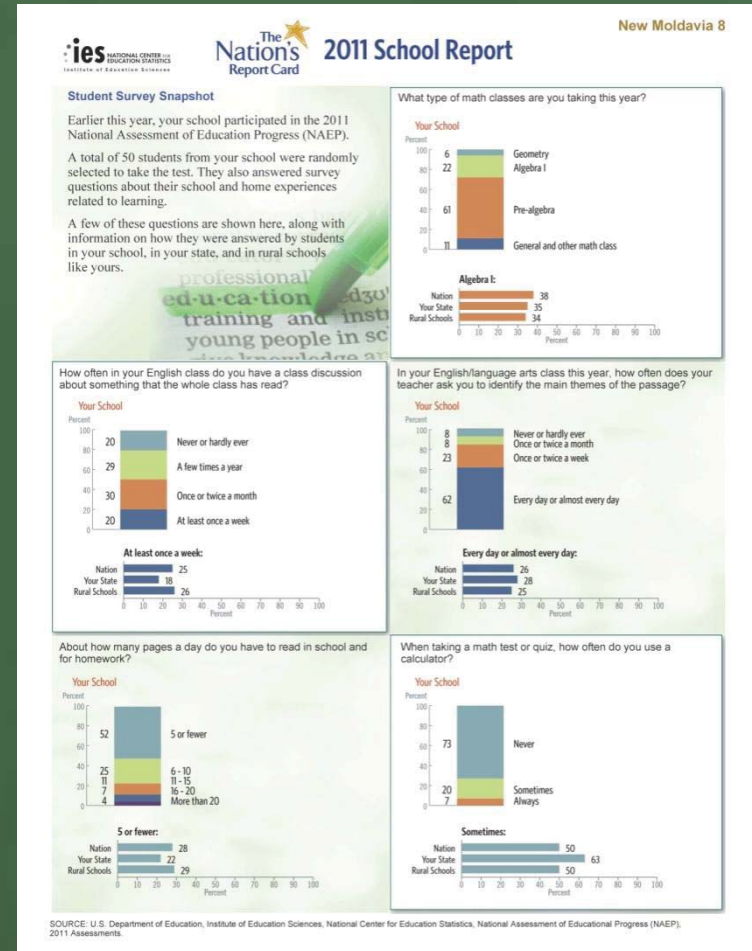
Gap - Average Scale Score: 2003-2011



NOTE: The NAEP Reading scale ranges from 0 to 500. Observed differences are not necessarily statistically significant.
 SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational

• Possible Future Uses

- Additional uses at state-/district-level
- Possibility of collecting this information at the state level
 - Pros
 - Cons



Questions? Comments?

lauraegan@westat.com



**THE NATION’S REPORT CARD
SCIENCE IN ACTION:
HANDS-ON AND INTERACTIVE COMPUTER TASKS
FROM THE 2009 SCIENCE ASSESSMENT
JUNE 19, 2012**

Overview

The public release of *Science in Action: Hands-On and Interactive Computer Tasks from the 2009 Science Assessment* took place on June 19, 2012 at 10 a.m. EDT as a live event and webcast. The event took place within a unique “Science + You” interactive exhibit space sponsored by the Living Classrooms Foundation at Foundry Lofts in Washington, D.C. John Dillow, executive director of Living Classrooms D.C., offered welcoming remarks on behalf of the organization.

For this release, there were 30 in-person release attendees and 196 webcast participants (NAEP staff and contractors were removed from registration lists and pie charts illustrating the types of audiences represented that follow this summary). The release received national media coverage from print, broadcast, and online outlets, most notably publications and blogs specializing in science. Selected media clips are presented in this document.

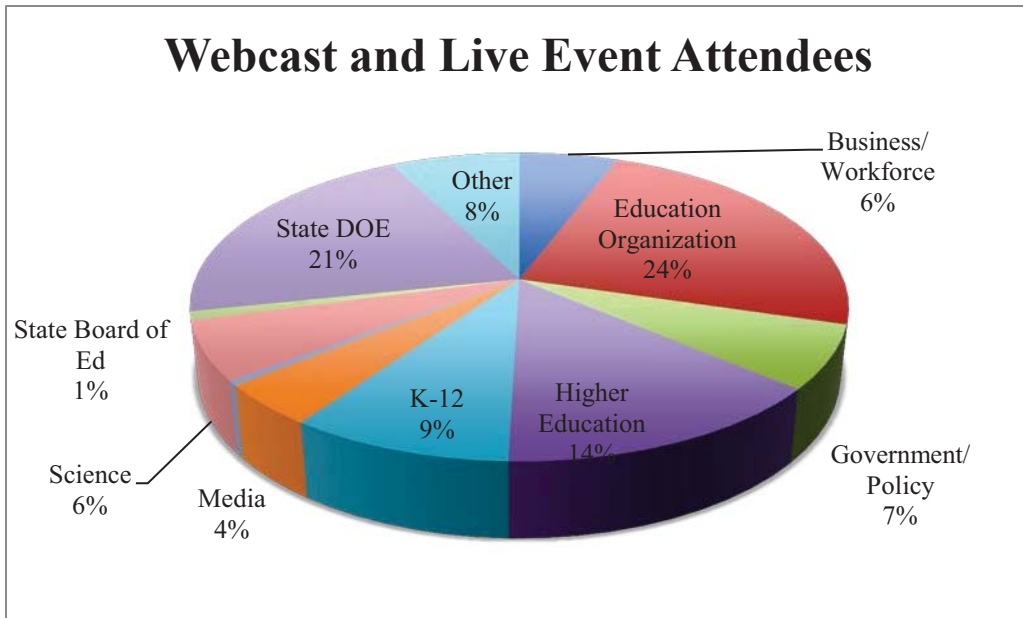
Release Event

Live event and webcast to release *Science in Action: Hands-On and Interactive Computer Tasks from the 2009 Science Assessment* on June 19 at 10 a.m. EDT.

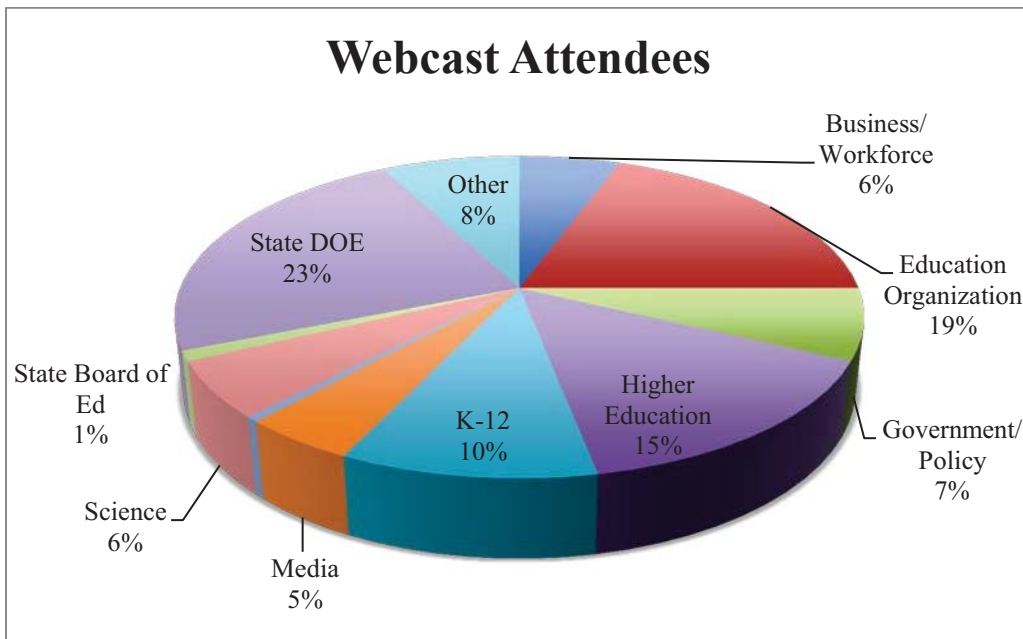
Panelists included:

- Eileen Weiser, Member, Michigan State Board of Education; Governing Board Member (Moderator)
- Alan Friedman, Consultant, Museum Development and Science Communication; Governing Board Member
- Jack Buckley, Commissioner, National Center for Education Statistics
- Chris Dede, Wirth Professor in Learning Technologies, Harvard Graduate School of Education

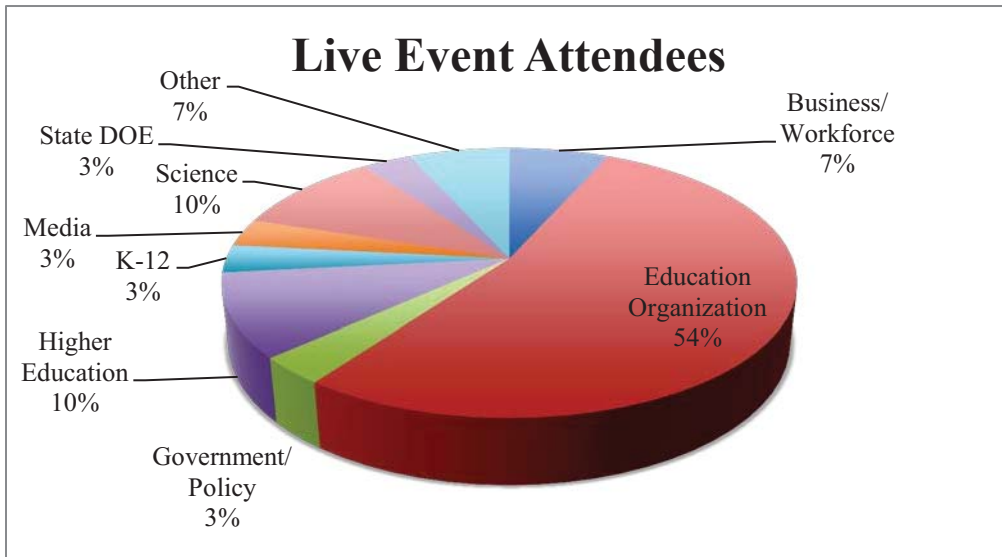
Webcast and Live Event Attendees- 226 combined total (excluding NAEP staff and contractors)



Webcast Attendees - 196 total (excluding NAEP staff and contractors)



Live Event Attendees – 30 total (excluding NAEP staff and contractors)





The Nation's Report Card: Science in Action: Hands-on and Interactive Computer Tasks from the 2009 Science Assessment

Select Media Clips

Interactive NAEP Tests Show Students Answer Simple Science Questions Well, Struggle to Explain

Associated Press, June 19, 2012 - Sarah Parnass

NAEP Reveals Shallow Grasp of Science

Education Week, June 19, 2012 - Nora Fleming

NAEP Results Good News, Bad News on Student's Science Transfer

Denver Post, June 19, 2012 - Kevin Simpson

Measuring How Well Kids Do Science

Science News, June 19, 2012 - Janet Raloff

US Students Can Do Basic Science Experiments but Many Cannot Explain Results New NAEP Results Show

Orlando Sentinel, June 19, 2012 - Leslie Postal

Should Hands-On Experiments Replace Bubble Tests

KQED, National Public Radio, June 19, 2012 - Lillian Mongeau

NAEP Science: Students can do experiments and get answers but can't explain or justify their results

Atlanta Journal Constitution, Get Schooled Blog, June 19, 2012 - Maureen Downey

Report: U.S. Science Students Run Successful Science Experiments, But Can't Explain Results

CNN, Schools of Thought Blog, June 19, 2012 - Sally Holland

U.S. Students Need Help on More Complex Science Tasks, NAEP 2009 Results Reveal

Tampa Bay Times, June 19, 2012 - Jeff Solocheck

For Most U.S. High Schoolers, STEM Knowledge is Only Skin Deep

US News and World Report, June 20, 2012 - Kelsey Sheehy

Students Fail at Scientific Thinking

UPI, June 19, 2012 - Staff Reports

'Nation's Report Card:" Students Struggle To Explain Scientific Principles

Education Writers Association's *Educated Reporter Blog*, June 20, 2012 – Emily Richmond

Why Aren't Americas Students Smart Enough to Handle Science?

Forbes Magazine, June 21, 2012 – J. Maureen Henderson

Associated Press

Interactive NAEP Tests Show Students Answer Simple Science Questions Well, Struggle to Explain

By Sarah Parnass

American children do much better identifying the correct answers to simple scientific tasks than using evidence from their experiments to explain those answers.

The National Assessment of Educational Progress, often called the Nation's Report Card, asked students in grades four, eight and 12 to perform actual experiments to apply principles they learn in the classroom on a practical level. The results of the 2009 tests were released Tuesday

“That tells us that our science teaching isn't getting us as far as we need to go,” said Chris Dede, professor from Harvard Graduate School of Education.

Katherine Carroll, an 11th- and 12th-grade chemistry teacher in Waterboro, Maine, said even her best students struggle to explain their conclusions in the lab reports they turn in for her class. She found them more accustomed to questions with one right answer.

“Teachers have moved towards teaching more knowledge, as opposed to the understanding behind that knowledge,” Carroll said.

Like Carroll, Dede said kids' difficulty explaining is old news to most teachers and parents, but this is the first time they have concrete evidence demonstrating the problem. “Having something that is more than just anecdotes, that is rigorous research across a wide range of students, is very helpful, because it's a better form of evidence on which to make decisions,” Dede said.

The first test, called Hands On Tasks (HOTs), allotted students 40 minutes to conduct experiments with physical objects. This allowed for a richer analysis of their understanding of the subject than pencil and paper tests can provide, according to Alan Friedman, chairman of the National Assessment Governing Board's Assessment Development Committee.

HOTs, however, are nothing new. NAEP tests used them as far back as 1996.

Friedman said the second type of test, Interactive Computer Tasks (ICTs), went beyond what had previously been measured, testing how students ran their own experiments in simulated natural or laboratory environments with the ability to go back, adjust variables and correct their mistakes on a computer.

“This is a set of skills which in the real world is invaluable,” Friedman said, “and which before this we'd never been able to know if students could do this or not.”

Though Friedman said the computer tests are “dramatically more expensive” to design, traditional assessments cannot measure these same skills.

During ICTs, just over a quarter of high school seniors could both select and explain their correct answers about heating and cooling. Double that amount - 54 percent - in the eighth grade group could support correct conclusions with evidence, but only 15 percent of fourth grade students could do the same in their experiment.

The computer tasks eliminated limits of geography and time, so students could virtually

see, for example, how a plant given a certain amount of sunlight would grow without waiting days or weeks to see the actual process.

Though the tests raised significant questions about students' abilities to apply scientific knowledge to the real world, they at least seemed to enjoy taking them, according to Peggy Carr, associate commissioner at the National Center for Education Statistics. Carr usually observes students losing interest in the traditional NAEP tests. "Not so with these assessments," Carr said.

In the hands on tasks, female students in every grade outdid their male counterparts by 2 to 4 percentage points, on average. Girls also scored slightly better than male students in grades eight and 12 on interactive computer tasks.

This gender gap shows a reversal from the traditional NAEP tests in which eighth-grade boys scored at least four points higher on average than their female peers in 2009 and 2011.

White and Asian-Pacific Islander students outperformed black and Hispanic students in the hands on tasks, and Asian/Pacific Islander students achieved higher scores on average than other students in all grades' computerized assessments.

The lowest scoring group in both assessments was 12th grade black students. They answered 19 percent of computerized questions correctly, whereas their Asian-Pacific Islander counterparts passed 33 percent.

Education Week

NAEP Reveals Shallow Grasp of Science

By Nora Fleming

Elementary, middle, and high school students failed to demonstrate a deep understanding of science concepts when they performed activity-based science tasks and investigations, concludes a study released today from the first national assessment of both hands-on and interactive computer-based science activities.

The hands-on tasks, which required students to use materials and laboratory equipment to perform science experiments, and the new, interactive computer tasks, which simulated an environmental or laboratory setting and asked students to solve scientific problems, were administered as part of the 2009 National Assessment of Educational Progress in science for 4th, 8th, and 12th graders. The report follows on the heels of the 2011 traditional pencil-and-paper science NAEP results released last month.

Both the hands-on and computer tests asked students to predict what might happen in a particular scientific scenario, make observations about what occurred in the scenarios, and explain the findings of the experiments or investigations they launched. These questions examined how well students could conduct and reason through “real life” science situations and grasp the scientific concepts of what occurred in their investigations, according to the report from the National Center on Education Statistics, the U.S. Department of Education division that administers NAEP.

About 2,000 students at each grade level were given each test and asked to complete two, 40-minute hands-on tasks or three interactive computer tasks, 20 to 40 minutes in length. In an 8th grade interactive computer task, for example, students could have been asked to plan a new, simulated recreation area for a town using part of an existing wildlife area, evaluate the impact different locations for the recreation space could have on local wildlife, and determine which space would be best to build on.

“Increasingly, graduates are called on to do things in today’s world that require more than rote memory and how to follow instructions,” Alan J. Friedman, a member of the National Assessment Governing Board, which sets policy for NAEP, said during a conference call yesterday about the tests. “There was no way to memorize for this test and no amount of rote drill and practice that could prepare students for it; these tests test what students can do in more complex environments and the richness of what students can do with real stuff.”

On average, the students were able to accurately report what was happening in scenarios with limited data, but were challenged by manipulating multiple variables and making decisions as part of running an experiment, according to the findings. Additionally, the numbers of students able to draw the right conclusions in experiments was much higher than the numbers of students who were able to provide an explanation or justification for their answer based on the findings.

Seventy-one percent of 4th graders could accurately select how volume changes when ice melts, for example, but only 15 percent could explain why that happened using evidence from the experiment.

The findings were fairly consistent across grade levels, other than 12th grade students' scoring some 15 percent lower than the younger students on the interactive computer tasks. Differences in test results were more pronounced instead between race, class, and gender groups. Disadvantaged and minority students performed lower than white and Asian students on both tests, and females performed better than males on hands-on tasks, but lower on the pencil-and-paper 2009 tests.

“While I’m happy to see the vast majority of students [tested] were able to make straightforward observations, I’m not particularly happy to see a smaller number know what data to collect in an experiment,” Jack Buckley, the NCES commissioner, said during the briefing. “This points to something we need to work on in the future.” In 2014, a technology and engineering-literacy NAEP is also expected to be administered.

New Standards

Last month, NAEP also released the results of its 2011 science tests, which found fewer than a third of 8th graders performing at “proficient” levels in science. Though there were small improvements in performance for all groups from the previous administration in 2009, on average, disadvantaged, black, and Latino students performed below basic level. The new results from the interactive science tests and 2011 results arrive as state and other education leaders work on finalizing a set of voluntary, national science standards aimed at improving the quality of science education in the United States, with the goal of shifting from rote memorization of subject matter to building students’ deeper understanding of core science concepts, how they connect, and how they can be applied to the real world.

Just last month, a draft of the new standards, which are being developed by a cadre of 26 states and a team of writers led by Achieve, a Washington-based nonprofit, were released to the public for comment. Focused around scientific and engineering practices, cross-cutting concepts across science disciplines, and core subject matter in physical, life, earth, and space sciences, and engineering and technology, the standards are expected to be finalized by early next year.

According to Mr. Friedman, the findings of the science-activity NAEP are right in line with what the new standards aim to improve: depth versus breadth in the understanding and practical application of science.

“The new tests are tailor-made to the types of skills listed in the new [draft] science standards,” he said. “We’re in a really good position to be models for assessments and provide the kind of information called for by the new standards.”

Nancy Butler Songer, a professor of science education and learning technologies at the University of Michigan and a longtime researcher on improving science education, said that while the NAEP results were disappointing, the future is not completely dismal.

Ms. Butler Songer, who is also one of many advisers providing feedback on the development of the new national science standards, said she finds it promising that NAEP and national organizations like Achieve are continuing to recognize the need to change science education and build “fused knowledge,” or content knowledge plus science

practices. These current efforts are part of the necessary “pieces coming together” to improve science education, she said, which include professional development to help teachers teach science better, curriculum and standards to guide teaching, and tests to measure how well students are understanding these concepts.

“We’ve maintained a misconception in what it meant to know science,” she said. “While it’s taken awhile to uproot this idea, what we know now is that you can’t get to a deeper level of understanding in science without working in science in a sophisticated way. You have to use models or gather and apply evidence from experiments to that concept in order to really know science. It’s no longer enough to settle for memorizing facts.”

Denver Post

NAEP Results Good News, Bad News on Student's Science Transfer

By Kevin Simpson

Science students across the country showed promise in performing hands-on tasks with basic data, but they faltered as variables became more complex and struggled to explain how they reached even correct conclusions, according to a new set of tests.

The National Assessment of Educational Progress on Tuesday reported its first results from the 2009 exams designed to measure how well students apply basic science knowledge by asking them to conduct actual experiments and manipulate computer models.

The assessments, given to students in fourth, eighth and 12th grades, involved both hands-on tasks — a pair of 40-minute

Colorado Classroom covers local and state education issues affecting K-12 and higher education students in the state of Colorado.

science experiments — and interactive computer problems of either 20 or 40 minutes simulating conditions that would be impossible or unsafe to perform in a lab.

The assessment revealed three key overall findings, said Jack Buckley, commissioner at the National Center for Education Statistics.

High percentages of students succeeded in dealing with limited sets of data and making straightforward observations about them — 80 percent in fourth grade, 84 percent in eighth grade and 75 percent in 12th grade.

But those success rates dropped precipitously when more variables were introduced — to 35 percent in fourth grade, 24 percent in eighth grade and 25 percent in 12th grade. And finally, the percentage of students in all three grades who reached correct conclusions from data far exceeded the percentage that could support those findings with evidence.

Buckley explained the disconnect using the example of one computer task given to fourth-graders that involved a simulated greenhouse. Students were asked to locate virtual plants in different levels of sunlight to see how well they grew, as measured by differences in height and number of blossoms, and determine whether the plants craved sun or shade.

"We found that the vast majority was able to reach the right conclusion," he said. "But when they got to the part of the assessment where they had to write why they reached that conclusion, to use the scientific evidence they'd just gathered from the simulation, this is where things started to go awry."

While 93 percent of those fourth-graders could determine that "Plant A" was a sun-loving plant, only 36 percent could use the data to explain why.

The two tests measured nationally representative samples of 2,000 students in each grade. But this was a sample separate from those who took the regular "pencil and paper" NAEP science tests.

Among other findings, while males scored higher than females on primary science assessments, females scored higher in hands-on tasks. On the computer models, there was no significant gender difference in scores.

Scores revealed an achievement gap in fourth and eighth grade between students from high- and low-income families.

Science News

Measuring How Well Kids Do Science

By Janet Raloff

Since 1969, the National Assessment of Educational Progress has issued report cards on how well America's youth perform on classroom tasks. Previously, they have assessed what kids know or can calculate. Two new components have now been developed to gauge a child's performance in hands-on and research-oriented interactive computer tasks. On June 19, NAEP released the first scores for these tests. And the overall grades: Well, they show plenty of room for improvement.

The new data from pilot-scale assessments of hands-on and computers-on research come from tests in 2009. Some 2,000 children took each test at each of three grade levels: 4th, 8th and 12th. "Across the 9 interactive computer tasks, we found that 42 percent of 4th graders, 41 percent of 8th graders and 27 percent of 12th graders gave correct answers on the steps they attempted," reports Jack Buckley, commissioner of the National Center for Education Statistics, which administers NAEP tests.

Overall, students were likely to be successful on parts of the testing "that involved limited sets of data and making straightforward observations from those data," he observes. Where kids tended to stumble — sometimes badly — was in using those data to extrapolate a general trend or justify a conclusion. For instance, Buckley notes, on one computer simulation for 4th graders of plants growing in a greenhouse, kids could move the plants around and identify, based on growth patterns, which were sun- versus shade-loving plants, and which fertilizer application rate proved most effective.

But when asked to explain in writing how they reached those conclusions, "this is where things started to go awry, Buckley said. Many simply couldn't "back up their conclusions effectively with the evidence they had just collected from the simulation."

Last month, NAEP issued 2011 science achievement stats for kids in middle schools across the nation. The science score was middling. Literally. On a 300-point scale, 8th graders collectively scored 152 points — up a mere 2 points from 2009. Two percent of the 122,000 surveyed children scored at an advanced level, no differently than two years earlier.

Nothing to brag about, such scores should come as no surprise. On one international survey after another, U.S. students fail to lead the pack. For instance, scores for 8th graders in the 2007 Trends in International Mathematics and Science Study (issued in 2009 and the most recent data available) averaged 508 points for math and 520 for science — hovering around the average (500 points) for this yardstick.

How did that compare with scores elsewhere around the world? "At eighth grade, the average U.S. science score was higher than the average scores of students in 35 of the 47 other countries, lower than those in nine countries (all located in Asia or Europe), and not measurably different from those in the other three countries," TIMSS reported. Ten percent of U.S. kids met or exceeded the advanced international benchmark in science — a smaller share than in Singapore, Taiwan, Japan, England, Korea or Hungary.

But the rote memorization of facts, formulas or rules that can lead to high scores on such tests do not a good 21st century scientist or engineer make, notes Alan Friedman, a member of an independent, bipartisan board established by Congress to set policy for NAEP. Important as those skills are, he says, in today's climate they simply aren't sufficient. So NAEP developed research-performance based tasks, he says, to measure "what students know and can do in more complex, real-world situations. (And this physicist is familiar with science achievement and outreach to the nation's youth: For 22 years he directed the New York Hall of Science.)

Regarding the newly reported scores, Buckley says that "As a citizen and a parent, I was not particularly happy — although pleased to see that the vast majority of students was capable of making straightforward scientific observations from data." He expressed far less satisfaction that a much smaller share could "either use strategy to actually decide what data to collect, or to arrive at the correct conclusions and be able to back them up with the evidence that they had just collected. I think that points to something that we need to work on."

Friedman was a bit more charitable. "The fact that we didn't bomb on it" — at least the initial, simpler elements of these tests, "that's very satisfying." As a science educator, he said: "I was relieved, frankly, that students didn't do really badly." Keep in mind, he pointed out, "No amount of rote drill and practice" — of memorizing formulas, words and scientific laws — "would help you to any significant extent on these tests. You really had to think on your feet."

The new research report card raises a big question for the nation's education elite: how to raise those scores, because they point to shortfalls in developing, synthesizing and using data — the essence of science. The issue isn't how poorly kids elsewhere around the world might do this (and we don't know that they do it poorly), it's only important that U.S. schools ensure their students do it well. At issue? Only the future economy and health of the nation.

Orlando Sentinel

US Students Can Do Basic Science Experiments but Many Cannot Explain Results New NAEP Results Show

By Leslie Postal

When the NAEP science test — the exam dubbed the “nation’s report card” — was administered back in 2009, some students were also asked to do hands-on or computer-based experiments. The goal was to see how well students could apply what they’d learned in science lessons to “real-life ” problems.

These activities showed that students can do the basics but stumble with more complex tasks — and with explaining a correct conclusion.

“The report shows that students were challenged by parts of investigations requiring more variables to manipulate, strategic decision-making in collecting data, and the explanation of why a certain result was the correct conclusion,” the NAEP folks said.

The hands-on tasks were given to a sampling of students nationwide in grades 4, 8 and 12. There are no state results for this report.

“Science is fundamental to education because it is through scientific inquiry that students understand how to solve problems and ultimately how to learn,” said David Driscoll, chairman of the National Assessment Governing Board, in a statement. “So it’s tragic that our students are only grasping the basics and not doing the higher-level analysis and providing written explanations needed to succeed in higher education and compete in a global economy.”

These hands-on tasks included having fourth graders use a computer simulation to figure out best sunlight conditions for different plants.

Another asked 12th graders to do a lab that involved testing water quality. Seventy five percent of students could test the water samples, 64 percent could recommend a site for a new town based on their findings but only 11 percent “were able to provide a valid recommendation and support their conclusions with details,” NAEP found.

You can find the full report, and more examples of what students were asked to do, [here](#).

KQED, National Public Radio

Should Hands-On Experiments Replace Bubble Tests

By Lillian Mongeau

To get a better understanding of how well students can solve complex problems and apply science to real-life scenarios, the National Assessment for Education Progress recently used hands-on experiments as a way to test 4th, 8th, and 12th grade students, and found that this kind of assessment gives a much more accurate reflection of student comprehension.

Results from a 2009 round of testing called The Nation’s Report Card Science in Action: Hands-On and Interactive Computer Task, examined 6,000 students—2,000 at each grade level—from across the country. Students performed tasks like testing water samples (12th grade) and assembling electric circuits (4th grade). They also participated in interactive computer tasks that simulated longer term experiments, like observing plant growth. In both scenarios, students were evaluated on their ability to perform the tasks, observe the results and draw conclusions.

“The bottom line is, we learned so much more that we couldn’t have learned from those paper and pencil tests,” said Jack Buckley, commissioner at the National Center for Education Statistics, which creates the annual “Nation’s Report Card” based on the results of tests like this one administered by the National Assessment for Educational Progress (NAEP).

But what they learned was a mixed bag.

A majority of students at all grade levels (76 percent) were able to perform the simpler experiments correctly and accurately observe the results. However, when experiments involved more complicated data sets, students’ ability to execute and observe fell sharply — only 36 percent of students tested across grade levels were able to complete the tasks under these conditions.

The test also revealed a disconnect between observation and explanation. Even though a majority of students (71 percent) were able to draw the correct conclusions from the results of their experiments, less than a third (30 percent) were able to explain their results.

For example, one of the hands-on tasks for 12th grade students was to determine the best location for a new town based on water quality. The students were expected to test various water samples for specific pollutants and then compare those levels to a chart put out by the Environmental Protection Agency. A whopping 75 percent of students were able to do this accurately. But when it came time to make a recommendation for where the new town should be built, only 11 percent of students were able to explain their recommendation using the data they’d collected.

The conclusion? “[Students] can conduct science investigations using limited data sets, but many students lack the ability to explain results. The report shows that students were challenged by parts of investigations requiring more variables to manipulate, strategic decision-making in collecting data, and the explanation of why a certain result was the

correct conclusion,” the report states.

For the most part, student performance broke down as it usually does along ethnic and economic lines. Low-income students performed worse than their wealthier peers and black and Latino students performed worse than their white and Asian counterparts. However, there were a few notable exceptions.

On some parts of some tests, black and Latino students did as well or nearly as well as white students. For example, on a computer task that required 4th grade students to observe plant growth, 80 percent of students came to the correct conclusion. Eighty-one percent of white students got the right answer, 79 percent of black students did and 74 percent of Hispanic students did. (Eighty-six percent of Asian and Pacific Islander students got that one right.)

What’s more, though male students generally outperform female students on the national science assessment, female students beat male students on the hands-on tasks. Alan Friedman, a physicist and the chair of the committee in charge of developing national assessments, said that as a scientist he was relieved that students did well on the first section of the test. “There’s no way for them to memorize for this test. You really had to think on your feet,” he said.

Still, Friedman said, he wasn’t shocked that students struggled to explain their results. “Unfortunately, that’s not surprising,” he said.

Though hands-on standardized tests aren’t brand new, they have historically been too expensive and complicated to use on a wide scale. And the technology needed for interactive computer tasks has not been up to snuff until recent years.

Officials at NAEP said tests like these are more accurate and provide far more detailed results. Buckley said they must become the norm to keep up with new curriculum standards meant to keep pace with the changing world of science and technology. “We’re in a really good position to provide models for assessment,” Buckley said, that can “provide information on what students can know and do that’s called for in the new standards.”

Atlanta Journal Constitution

NAEP Science: Students can do experiments and get answers but can't explain or justify their results

By Maureen Downey

NAEP — known as the Nation’s Report Card — released results today of how American students fared on a new component of its science test that included hands-on, interactive experiments and virtual labs.

The new component was added to the 2009 science assessment. In one example, 12th graders were asked to determine a location for a new town based on an assessment of water quality flowing near that site. Students were asked to test water samples, determine levels of pollutants and then justify the decision where they would locate the new town using the data from the experiment they conducted.

Overall, students could conduct the experiments but were not as skilled in using their data to justify conclusions or writing reports. In one example cited in a webinar this morning on the results, 93 percent of fourth graders got the right answer in a science experiment, but only 32 percent could use the evidence from the experiment to justify their answer.

On the webinar announcing the results, National Center for Education Statistics Commissioner Jack Buckley said NAEP learned three key things through this new testing component:

- ”Students are pretty good at doing some parts of science. The vast majority could use simulated laboratories to do the tests.”
- However, “students overall across all the task and across all grade levels were challenged by the parts of the test that required them to consider more than one variable at a time or if they had to make strategic decisions about how to collect the data.”
- ”Students could select correct conclusions, but didn’t do so well when we asked them to explain their conclusions using the evidence from the data tables.”

Buckley noted that girls outscored boys in hands-on tasks, although boys outscored girls in the traditional NAEP science test for which results were released last month. There was no gender gap in the interactive computer segment.

Buckley said students loved these new assessment items, adding, “Kids said it was fun. It was hard to pry the computer and the tasks away from the kids. They really wanted to keep doing them. It was hard to get them to stop.”

Also taking part in the webinar was National Assessment Governing Board member Alan J. Friedman, a Georgia Tech graduate.

“With technology so close to the center of our society, we reward response to change and innovation,” he said. “So, testing to see how much students can memorize and how well they can follow instruction is no longer good enough. We need to know that students have the so-called higher order, 21st century skills...It is crucial to know if students know how an experiment or engineering task is designed, how data is analyzed and how to draw the best of multiple, possible solutions. These are all critical to innovation.”

CNN, Schools of Thought Blog

Report: U.S. Science Students Run Successful Science Experiments, But Can't Explain Results

By Sally Holland

Washington (CNN) - American students can successfully conduct simple science experiments at school, but aren't able to explain the results, a new report from the National Assessment of Educational Progress shows.

Results released today reveal that America's fourth-, eighth-, and 12th-graders struggled when investigations had more variables to manipulate or required strategic decision-making while collecting data. Many weren't able to explain why certain results were correct.

It's the first time the National Assessment of Educational Progress, known as the Nation's Report Card, measured how students performed on hands-on and interactive computer tasks like a professional scientist might. While traditional standardized tests grade students on what they know, people in the workforce are measured on how they apply what they've learned in school. This analysis moves away from "paper and pencil" tests and should allow for a different type of analysis by education experts.

The testing involved more than 2,000 students from public and private schools in fourth, eighth and 12th grades during the 2009 school year. In one task, 12th grade students were asked to determine the best location to build a town based on the quality of the water supply. The results show 75% of students could perform tests on water samples and tabulate data, but only 11% could "provide a valid final recommendation by supporting their conclusions with details from the data," according to the report.

"It's tragic that our students are only grasping the basics and not doing the higher-level analysis and providing written explanations needed to succeed in higher education and compete in a global economy," NAEP chairman David Driscoll said in a statement.

The Nation's Report Card results reflect the experiences of K-12 teachers, said Patricia Marsteller, director of Emory University's Center for Science Education, but she sees reason to be optimistic. Problem-based learning requires the support and education of teachers, school boards, administrators, education professors and others; measuring how students succeed and struggle at it might help solidify hands-on learning in the curriculum, despite the obstacles.

"It's moving away from, 'Which of the following are mammals?' the little memorization questions, to something that really does address how science is done," said Marsteller, an Emory biology faculty member. "This is a way, way better way to investigate whether students are learning what we think they need to learn to be active citizens and scientifically literate.

"Real world problems are messy. Research is messy. Having a real investigation, where you design something and you iterate, 'What else could be going on? What else could I do? That's where you learn something.'"

The report, "Science in Action: Hands-On and Interactive Computer Tasks from the 2009 Science Assessment," found that female students scored higher than their male counterparts on the hands-on on tasks, but male students scored high on the more traditional "paper and pencil" tests. No gender gap was found for interactive computer tasks.

CNN's Jamie Gumbrecht contributed to this story.

Tampa Bay Times

U.S. Students Need Help on More Complex Science Tasks, NAEP 2009 Results Reveal

By Jeff Solochek

Florida political and education leaders lately have made the case that students need more relevant exposure to science, technology, engineering and math. The latest NAEP results suggest that such a direction is wise.

Noting that science is much more than book learning, but also about exploration and application, the National Center for Education Statistics writes that the 2009 hands-on science assessment indicates that U.S. students do better at simple science than in more complex problems. They're also better able to arrive at conclusions than explain them.

From the report summary:

Students were successful on parts of investigations that involved limited sets of data and making straightforward observations of that data.

Students were challenged by parts of investigations that contained more variables to manipulate or involved strategic decision making to collect appropriate data.

The percentage of students who could select correct conclusions from an investigation was higher than for those students who could select correct conclusions and also explain their results.

It's interesting to note that female students at all three tested grade levels outperformed males in hands-on tasks, while male students did better on paper and pencil testing. Read more about the details and findings [here](#). What are your schools doing to improve science education, including hands-on lessons?

U.S. News and World Report

For Most U.S. High Schoolers, STEM Knowledge is Only Skin Deep

By Kelsey Sheehy

American teens are adept at conducting scientific experiments, but only if they don't stray beyond the basics, according to assessment results released Tuesday by the National Center for Education Statistics.

Seventy-five percent of high school seniors successfully completed straightforward experiments as part of the 2009 National Assessment of Educational Progress (NAEP) science exam. When tasked with more complicated experiments, only 25 percent came to the correct conclusion.

[See photos of U.S. News's Best High Schools for STEM.]

Students have even more trouble explaining their results and drawing conclusions from the data they collected during the experiments. Only 11 percent of the 12th-grade students were able to do so, according to "The Nation's Report Card: Science in Action," which detailed results for students in grades 4, 8, and 12.

In a competitive, technology-dominated society, simply following instructions will not cut it, Alan Friedman, a member of the National Assessment Governing Board, the policy arm of the NAEP, said at a panel discussion of the results on Tuesday.

"Testing to see how much students can memorize and how well they can follow the instructions is no longer good enough," Friedman said. "It's crucial to know if students understand ... how to draw the best of all possible solutions."

[Find out why STEM education is vital to the U.S. economy.]

Approximately 2,000 students at each grade level were selected to conduct either hands-on or interactive computer-based experiments as part of the annual NAEP science assessment. Experiments for students in grade 12 included testing and analyzing water quality, investigating the heat capacities of different metals, and classifying stars.

The new tasks allow stakeholders to get a more complete picture of students' problem-solving abilities than traditional paper-and-pencil exams, David Driscoll, chairman of the National Assessment Governing Board, said in statement Tuesday.

Improving students' abilities to perform in-depth scientific analysis requires more than testing, Jack Buckley, commissioner of the National Center for Education Statistics, said at Tuesday's panel.

"It's not enough just to have a day at the end of the year when you walk in and drop these tests on kids," Buckley said. "It has to be part of the curriculum."

UPI

Students Fail at Scientific Thinking

Staff Reports

An attempt to assess U.S. students' ability to do hands-on science found they did a good job of getting the right answer, officials said Tuesday.

But students participating in the 2009 National Assessment of Educational Progress were not so good at explaining the connection between their test data and the conclusion, National Commissioner for Education Statistics Jack Buckley said.

"Students are pretty good at doing some parts of science. The vast majority could use simulated laboratories to do the tests," Buckley said in a "webinar" Tuesday reported by

The Atlanta Journal and Constitution. "Students overall across all the task and across all grade levels were challenged by the parts of the test that required them to consider more than one variable at a time or if they had to make strategic decisions about how to collect the data."

The NAEP tests a sample of 4th, 8th and 12th graders with tests given most frequently in math, reading, writing and science. The results are nicknamed "the nation's report card." The science component included hands-on and interactive computer sections for the first time in 2009.

Education Writers Association's *The Educated Reporter Blog*
'Nation's Report Card:' Students Struggle To Explain Scientific Principles
By Emily Richmond

American students are more successful at correctly completing simple scientific tasks than they are at explaining how they used evidence to draw their conclusions, according to the latest results from the National Assessment of Educational Progress, known as "The Nation's Report Card."

The NAEP results released Tuesday represent a sampling of U.S. students in grades 4, 8, and 12 who participated in interactive science assessments in 2009. Some of the questions involved what are known as "hands on tasks," while others were "interactive computer tasks." The assessments measure the students' grasp of prior knowledge – material covered in class – as well as their ability to predict, observe and explain outcomes based on the evidence provided.

That cognitive leap -- from choosing the right answer and being able to articulate how it was reached -- matters, said Alan Friedman, chairman of the assessment development committee for the National Assessment Government Board, which sets policy for NAEP.

"Science and technology would be easy if all our challenges could be solved with simple memorization of accepted facts, and purely procedural application of known principles and laws," Friedman said at a press conference Tuesday in Washington, D.C. to announce the NAEP results. "In the real world, things are messy and one size does not fit all."

The hands-on tasks for fourth graders included having them assemble a simple electrical circuit and then determine the conductivity of various objects. The 12th graders were asked to test samples from two different water sources to determine the better location for a new town. (The interactive computer tasks for all grade levels are available online via the NAEP Web site.)

The interactive activities used for the assessments were hugely popular among the students, said Jack Buckley, commissioner for the National Center for Education Statistics, which oversees the NAEP administration. Teachers reported that their students said they had actually tried harder on the activities "because it didn't seem like a test," Buckley said.

When it came to the achievement gap for minorities and students from low-income households, the science assessment results mirrored prior NAEP findings in core subjects including math, reading, and history. The group of students qualifying for free and reduced-price meals scored significantly below their more affluent peers. The group of black students had the lowest percentage of correct answers at every grade level.

Interestingly, the gender gap that typically favors boys on science assessments was reversed in some areas of the new results. By a margin of between 2 and 4 percentage points, girls' group outscored the boys' at every grade level on the hands on tasks. For the interactive computer tasks, girls also outscored the boys in grades 8 and 12, although the

margin was just 1 percentage point.

Part of the problem is that even in schools where hands-on science learning is commonplace, students are not being asked to write or explain how they reached their conclusions, NAEP officials say. Just 39 percent of fourth graders and 57 percent of eighth graders had teachers who said they put at least a moderate emphasis on cultivating scientific writing skills. At the high school level, 28 percent of 12th graders said they had to complete a written science report at least once a week.

Indeed, it's one thing to be able to give the correct answer to a question or follow directions for a lab experiment, and another to use that knowledge as a springboard to deeper levels of intellectual inquiry. Since that rarefied air is where innovation happens, the new NAEP results are of particular interest to Change the Equation, a coalition of over 100 leading companies that have teamed up to help transform how STEM (science, technology, engineering and mathematics) courses are taught in public schools.

The NAEP results are "further confirmation that it's those broader kinds of skills, the ability to draw appropriate inferences to understand and explain, is where we show weakness," said Claus von Zastrow, Change the Equation's chief operating officer and director of research. "That's very much the area where we need to show strength."

Change the Equation's corporate partners want students to have greater access to the kinds of interactive learning opportunities used in the NAEP assessment, so they are doing more than just rote memorization, von Zastrow said. However, students still need to master basic skills and core subject matter.

"There's been a long-raging, and somewhat phony, debate between knowledge and application -- could you have 21st century skills without knowledge," von Zastrow said. "These (NAEP) results are further proof that's a false distinction."

Change the Equation is using a \$1.5 million grant from the Carnegie Corporation of New York (as part of its campaign to add 100,000 high-quality STEM teachers to public schools in the next 10 years) along with private donations to scale up successful programs in underserved schools. The long-term goal is to cultivate students who will be the next generation of scientists and engineers, so that corporations can look closer to home for talent instead of recruiting outside of the United States, von Zastrow said.

"Innovation means creating something new out of what you know," von Zastrow said. "If students don't have the ability to move from facts to understanding to explanation to higher orders of thinking, there's not much hope for them to be particularly innovative."

Forbes Magazine

Why Aren't Americas Students Smart Enough to Handle Science?

By J. Maureen Henderson

The results are in and America's elementary, middle and high school students are stumped by science. The National Center for Education Statistics released the findings of their National Assessment of Educational Progress science exam this week and it doesn't bode well for the state of STEM (Science, Technology, Engineering and Mathematics) education.

While the majority of students at the fourth, eighth and twelfth grade levels could successfully complete straightforward hands-on or computer-based tasks and arrive at the correct conclusions, once additional variables or more complex calculations were introduced, their performance declined dramatically.

For example, 75% of high school seniors could successfully use test strips to test water samples for the levels of four pollutants, record the data and interpret whether the results exceeded EPA standards, but only 25% of students were able to design and conduct an investigation using a simulated calorimeter and related patterns in temperature changes in two different metals to determine which metal has the higher specific heat capacity.

Results were the same at the lower grade levels, where only 24% and 35% of eighth and fourth graders respectively were able to handle the more difficult experiments. Students also had difficulty in explaining how they arrived at a correct conclusion, with only 27% of twelfth graders able to both select a correct answer and explain why they did so in one section of the test. And in another section, only 11% were able to make a final recommendation that was supported by the data they had worked with in the experiment.

These results are particularly worrisome in light of the fact that four out of five students tend to make up their minds about whether or not to pursue college-level STEM studies during high school or earlier. Currently, only about a third of bachelor's degrees awarded in the US are in the STEM fields – by contrast, over half of Chinese and Japanese college students are specializing in STEM subjects.

The economic and career benefits of STEM education are well-documented. STEM occupations are forecasted to grow faster than non-STEM occupations through to 2020. Over the course of the recession, unemployment in STEM fields has been almost half that of non-STEM fields. And STEM professionals earn, on average, approximately 26% more than non-STEM counterparts.

The need for beefed-up STEM education is already a hot political issue. In 2011, President Obama promised that 100 000 new STEM teachers would be trained over the next decade and in January 2012, he called on Congress to pass legislation to fund the retraining of two million unemployed workers for more technical careers.

As of 2010, the total federal investment in STEM education across all agencies was \$3.4B, which represents approximately 0.3% of the US's total education budget of \$1.1T. According to the National Center for Educational Statistics, only 53% of high school seniors reported that they are currently enrolled in a science class.

If there's a bright spot in the NAEP report, it's the fact that female students are matching or exceeding the performance of their male peers in both hands-on and interactive tasks. Currently, less than 25% of STEM jobs are filled by women. But positive exposure to science in the classroom could make all the difference – 68% of female college students studying STEM subjects say a class or a teacher was what initially piqued their interest in these fields.



Upcoming NAEP Reports as of July 2012

Report **Expected Release Date**

Initial NAEP Releases

<i>2011 Writing Report Card: Grades 8 and 12</i>	August 2012
<i>2011 Meaning Vocabulary: Grades 4 and 8</i>	October 2012
<i>2012 Long-Term Trend Reading and Mathematics</i>	April 2013
<i>2012 Economics Report Card: Grade 12</i>	May 2013
<i>2013 Reading and Math Report Cards: Grades 4 and 8</i>	October 2013
<i>2013 TUDA Reading and Math: Grades 4 and 8</i>	December 2013

Other NAEP Reports

<i>2005 HSTS Math Curriculum Study</i>	November 2012
<i>Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade</i>	December 2012

NAGB Reports

<i>Mega-States Report: Grades 4, 8, and 12</i>	November 2012
--	---------------

Other Related Reports from NCES

<i>New Americans in Postsecondary Education: A Profile of First – and Second- Generation Undergraduates</i>	July 2012
<i>Trends Among Young Adults Over Three Decades, 1974-2006</i>	July 2012
<i>Postsecondary Institutions of Price of Attendance in the United States: Degrees and Other Awards Conferred: 2010-2011 and 12 Month Enrollment: 2010-2011</i>	July 2012
<i>First Findings from the Kindergarten Rounds of Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011</i>	July 2012

Mega-States: An Analysis of Student Performance in the Five Most Heavily Populated States in the Nation

This report will provide NAEP results for the five Mega-States—the most populous states in the nation. The Mega-States are California, Florida, Illinois, New York, and Texas. Close to 40 percent of the nation’s students attend schools in these states. Additionally, eight of the ten most heavily populated cities are located in the Mega-States and they represent distinct regions of the country. They also have the highest number of English language learners (ELL) in the nation. Given the scope of these school systems and the challenges they face, outcomes in these states inform and influence decision makers regionally and nationally. An NCES report that tabulates, organizes, and discusses these specific results (though published previously) provides a needed service to the educational community.

Content:

The NAEP 2011 Mega-States report will discuss NAEP results for fourth- and eighth-graders in reading, mathematics, science, and writing in California, Florida, Illinois, New York, and Texas. Demographic changes in these states between 1990 and 2011 will be presented graphically, showing how vastly their populations have shifted over time and how the demographics vary across the Mega-States. Additionally, a summation of how many times the Mega-States performed statistically higher than, no different from or lower than the nation in eight assessments will be displayed. Trends in average scores for each of the Mega-States will be shown. Gains made by Mega-States overall and by selected student groups over time will be displayed graphically, especially in mathematics and reading where a substantial trend line is available. The report will highlight score gains made within the Mega-States that outpaced gains made nationally. In addition, there will be graphs and discussions of student performances at or above *Proficient* for a variety of major NAEP reporting groups (e.g., racial/ethnic groups and National School Lunch Program eligibility groups). Percentages at or above *Proficient* that are significantly different from the nation and/or all other Mega-States will be highlighted.

Publication Plans:

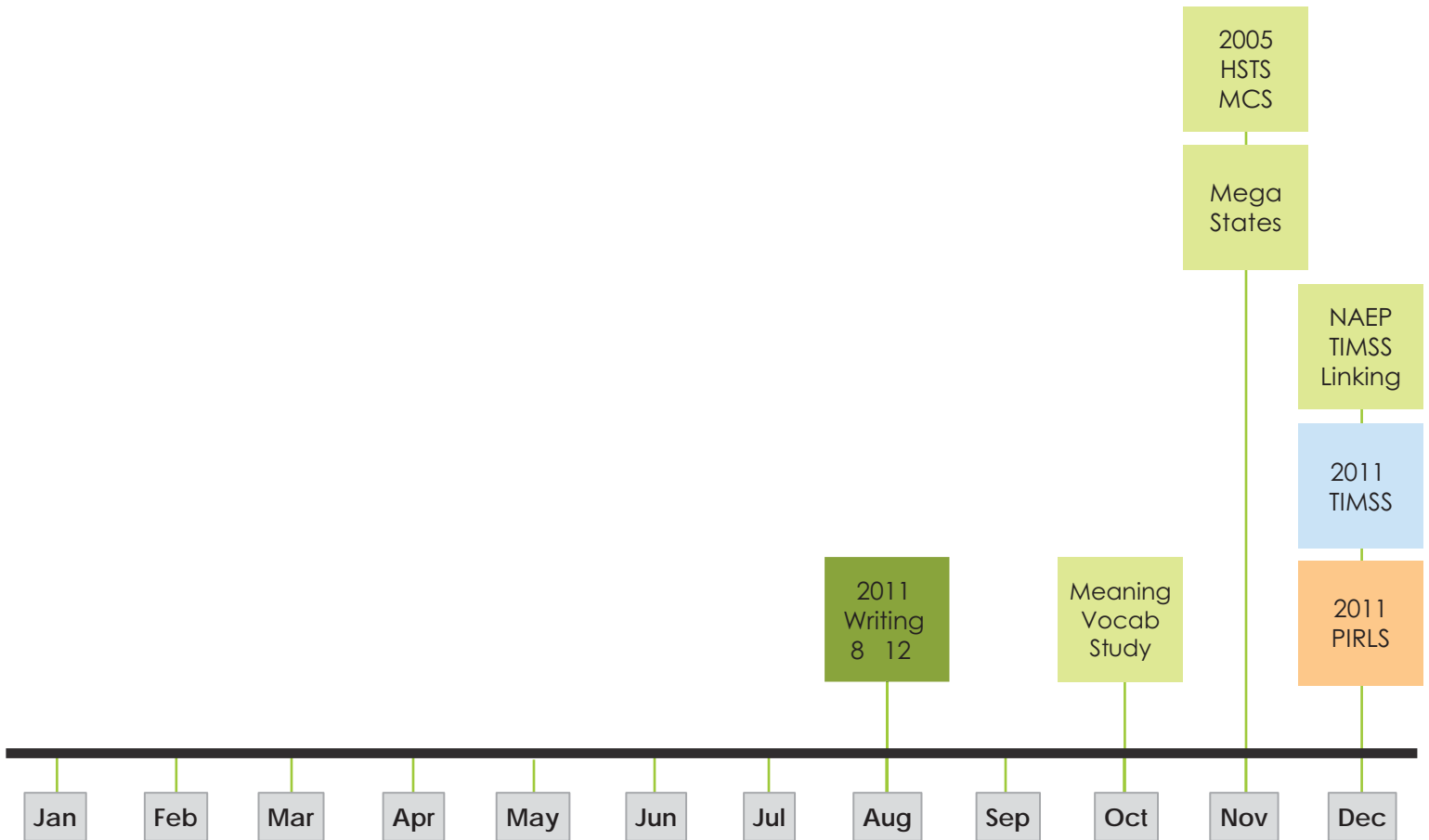
The report will be issued in both printed and electronic formats. The printed report will contain the main findings, comparisons, and trends. Links will be embedded in the electronic form of the report to more detailed findings on the NAEP web sites. This companion Web report will complement this presentation with interactive displays and state ranking tools, and will also include contextual variables and additional demographic data.

Projected Date for Governing Board Review: 9/10/2012

Projected Release: November 2012

2012

NCES Assessment Data Release Timeline



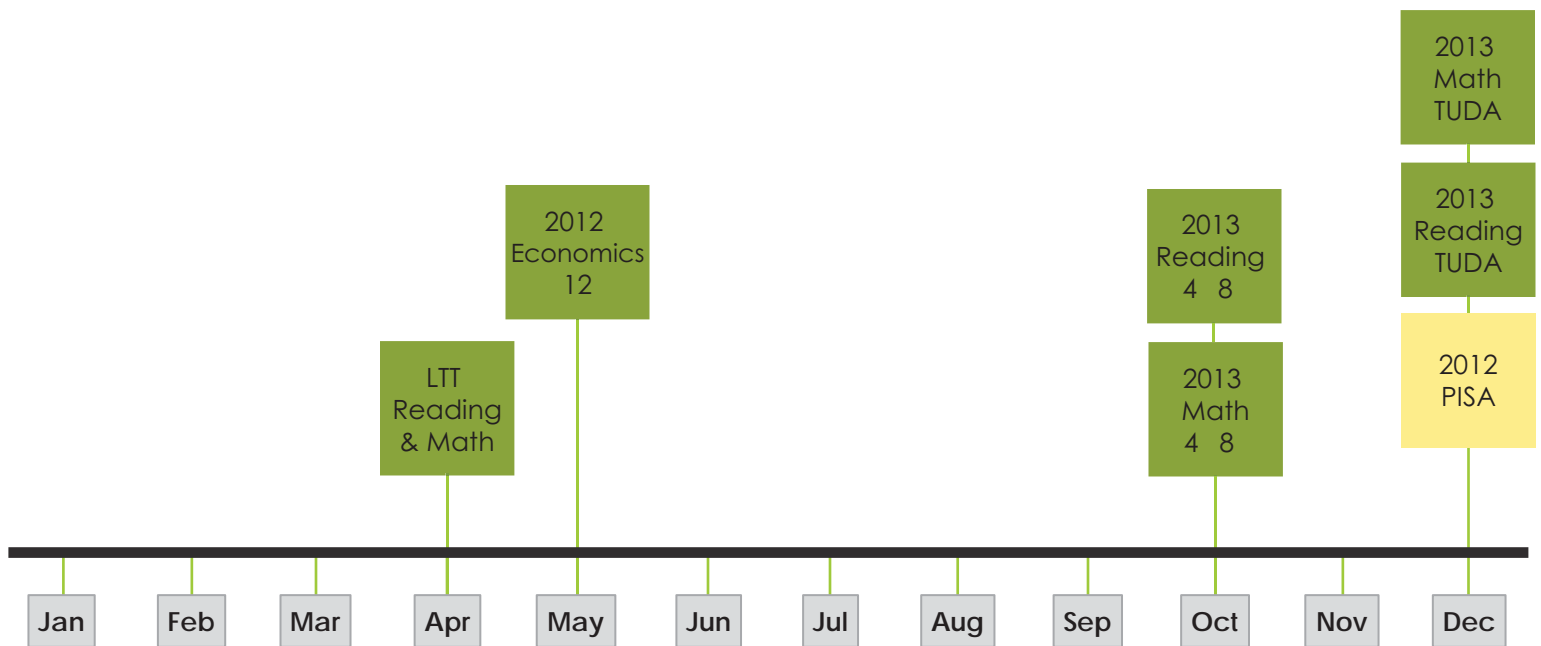
LEGEND

- NAEP Report Cards
- NAEP Studies
- TIMSS
- PIRLS
- PISA

DRAFT
July 12, 2012

2013

NCES Assessment Data Release Timeline



LEGEND

- NAEP Report Cards
- NAEP Studies
- TIMSS
- PIRLS
- PISA

DRAFT
July 12, 2012

Releases in 2012

- 2011 Science Report Card: Grade 8 (Web only)
- 2011 Writing Report Card: Grades 8 and 12 (National only)
- 2011 TIMSS : Grades 4 and 8 (National only)
- 2011 PIRLS : Grade 4 (National only)

Releases in 2013

- 2012 Long-term Trend (LTT) Reading Math: Ages 9, 13, and 17 (National only)
- 2012 Economics Report Card: Grade 12 (National only)
- 2013 Reading Report Card: Grades 4 and 8
- 2013 Mathematics Report Card: Grades 4 and 8
- 2013 Mathematics Report Card: Trial Urban Districts (TUDA): Grades 4 and 8
- 2013 Reading Report Card: Trial Urban Districts (TUDA): Grades 4 and 8
- 2012 PISA: Age 15 (National only)

Assessment Data Collection Schedule 2012 -13

NAEP

International

2012

- Economics: Grade 12 (National only)
- Long-term Trend: Ages 9, 13, 17 (National only)

- PISA USA (Age 15)

2013

- Reading: Grades 4, 8, 12
- Math: Grades 4, 8, 12

**NATIONAL ASSESSMENT GOVERNING BOARD
RELEASE PLAN FOR
NAEP MEANING VOCABULARY 2011 REPORT**

The Nation's Report Card in Meaning Vocabulary 2009 and 2011

The Nation's Report Card in Meaning Vocabulary 2009 and 2011 will be released to the general public during October 2012. Following a review and approval of the report's results, the release will be arranged as an online webinar. The release event will include a data presentation by the Commissioner of Education Statistics, with moderation and comments by at least one member of the National Assessment Governing Board. Full accompanying data will be posted on the Internet at the scheduled time of release.

With the 2009 National Assessment of Educational Progress (NAEP) Reading Framework, NAEP introduced a more systematic assessment of vocabulary, in which students were tested on their ability to use words to comprehend the sentence or paragraph in which the word occurs. These meaning vocabulary questions measure students' ability to apply word knowledge to develop and interpret meaning. This Report Card explains the new focus on meaning vocabulary and how it fits with the comprehension assessment and illustrates what students were asked to do with specific examples. It also shows the relationship of performance on meaning vocabulary to performance on reading comprehension.

Results are presented for all three grades, for the nation and the states, and for student groups including gender and race/ethnicity and span two years – 2009 and 2011. Vocabulary results for 2009 are based on nationally representative samples of 178,800 fourth-graders, 160,900 eighth-graders, and 51,700 twelfth-graders. Results for 2011 are based on samples of 213,100 fourth-graders and 168,200 eighth-graders. (There was no twelfth-grade assessment in 2011.)

DATE AND LOCATION

The release event for the media and the public will occur in October 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 National Assessment Governing Board member
- 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 or mailings 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, will work with Board staff to coordinate an in-person or online event designed to extend the life of the NAEP Meaning Vocabulary results by featuring current topics that would be of great interest and relevance to stakeholders with an interest in student achievement in reading and vocabulary. The event would be designed for organizations, officials, and individuals in the fields of education and policy whose work involves reading education and assessment.

NOTE TO Reporting and Dissemination Committee on Progress Report on 12th Grade Preparedness

Based on the Governing Board's discussions in May 2012, Board staff, with the assistance of Widmeyer Communications, has reconceptualized the public report on the preparedness of 12th graders for postsecondary education and job training. It is now being developed as a progress report for a broad policy audience describing the research conducted, the main research findings, and plans for future research based on the 2013 NAEP.

The report will discuss the complex issues involved, describe the Board's research—including both strengths and limitations, and present the full range of research findings relative to the NAEP scale. The report will also discuss research findings about college preparedness in relation to the Proficient achievement level in reading and mathematics. Unlike the version discussed in May, there will be no firm conclusions about preparedness reference points and no extended presentation of the percentage of students reaching them. Such conclusions cannot be adequately supported by the evidence assembled to date.

Under separate cover, Board members will receive a draft outline of the report and several key chapters to review. These will be discussed separately by all three of the Board's standing committees and by the full Board on Saturday morning. Members are being asked for feedback on the report to guide further staff work.

The preparedness report currently is scheduled for release in October or November. It will be accompanied by full texts of all research studies, which will be made available online at the time of release. This online compilation is being referred to as the technical report. Since the release should take place before the Board convenes in December, a draft release plan will be presented for action at the August Board meeting.

**NATIONAL ASSESSMENT GOVERNING BOARD
RELEASE PLAN FOR
NAEP 12TH GRADE PREPAREDNESS REPORT**

Governing Board Report on NAEP 12th Grade Preparedness Research

A report describing the studies and key findings of research commissioned by the Governing Board to enable the National Assessment of Educational Progress (NAEP) to report on the academic preparedness of 12th graders for postsecondary education or job training will be released to the general public during October or November 2012. The release will be in Washington, DC, in the form of a live, in-person event similar to a press conference. The release event will include a presentation on the research findings as well as comments and moderation from members of the National Assessment Governing Board and the NAEP 12th Grade Preparedness Commission. Accompanying materials, including technical reports, press release and statements, will be posted on the Governing Board's website at the scheduled time of release.

BACKGROUND

The Governing Board established a special commission on 12th grade issues in 2002 and subsequently arranged focus groups of educators, students, technical advisors, and business representatives to study this area. This commission recommended that NAEP be transformed to report on the preparedness of 12th grade students for college, job training and the military. The Board appointed a technical panel of experts in 2006 to identify research studies that could examine the links between NAEP and academic preparedness and in 2009, formally approved a program of preparedness research based on results from 12th grade NAEP. In 2011, the Board appointed the NAEP 12th Grade Preparedness Commission to increase awareness of the importance of preparing students academically for postsecondary education or training for employment after high school, as measured by NAEP. The Commission has conducted a series of symposia across the country with a two-fold purpose: to inform stakeholders of the Board's preparedness initiative and to receive input from stakeholders, including ways they can utilize the NAEP research.

The preparedness release event will serve as a culminating 12th Grade Preparedness Commission symposium, incorporating the ideas and issues raised at the more locally focused symposia and serving as a forum in which the final report of the preparedness research, involving more than 30 studies, is shared with key stakeholders and the general public, including media.

DATE AND LOCATION

The release event for the media and the public will occur in October or November 2012 at a suitable venue in the Washington, DC area. The exact date and location will be determined by the Chair of the Reporting and Dissemination Committee in consultation with the Chair of the 12th Grade Preparedness Commission.

RELEASE EVENT FORMAT

- Introductions and opening statement by Governing Board Chair David Driscoll
- Presentation of research and findings by Executive Director Cornelia Orr
- Comments by a Governing Board member; Gov. Ronnie Musgrove, chair of the NAEP 12th Grade Preparedness Commission; and other Commission members
- Questions from members of the press followed by comments or questions from the general audience
- The release event will last approximately 60-90 minutes. The symposium program may extend beyond the release and conclude later.
- Release may be Webcast live over the Internet. Viewers will be able to submit questions electronically to panelists.

REPORT RELEASE

At the live event, the Governing Board will provide attendees with packets of relevant materials, including a press release, panelist statements, and copies of the public report. On the Board's web site at www.nagb.org, these materials, including full technical reports from the research program, will be simultaneously available. The site will also feature links to social networking sites, key graphics, and audio and/or video material related to the event.

ACTIVITIES BEFORE RELEASE

In the days preceding the release, the Governing Board will offer embargoed access to the preparedness report as well as a national conference call for appropriate media who have signed the Governing Board's embargo agreement.

ACTIVITIES AFTER THE RELEASE

The Governing Board's communications contractor, Reingold, will assist the Governing Board to extend the life of this release by identifying opportunities, such as conferences, webinars, and seminars, where the Governing Board and the Commission can discuss the preparedness research and its implications for education and the economy.



Implementation of National Assessment Governing Board Inclusion Policy for Students with Disabilities (SD) and English Language Learners (ELL)

In March 2010, the Governing Board adopted a new policy, *NAEP Testing and Reporting on Students with Disabilities and English Language Learners*. The goals of the new policy include:

- Maximize participation of sampled students in NAEP;
- Reduce variation in exclusion rates for SD and ELL students across states and districts;
- Develop uniform national guidelines for including students in NAEP; and
- Ensure that NAEP is fully representative of SD and ELL.

The timeline for inclusion policy implementation is as follows:

- March 2010 – Governing Board approved the inclusion policy.
- NAEP 2011 – Report jurisdictions meeting the 95% and 85% inclusion goals in the Nation’s Report Card.
- NAEP 2012 – Pilot of the new decision tree for including SD and ELL.
- NAEP 2013 – Implement the new decision tree for including SD and ELL.

In order to meet the goals of the policy and achieve uniform national rules for including students, a new decision tree for including SD and ELL students was field tested for NAEP 2012 and will be implemented for NAEP 2013. According to the new decision tree, the only students that should be eligible to be excluded from NAEP are:

- Student with disabilities who take the state alternate assessment based on alternate achievement standards; and
- English Language Learners who have been enrolled in U.S. schools for less than one year.

All other students should be “encouraged” to participate in NAEP with or without accommodations allowed by NAEP.

In the past, inclusion decisions were made based on how the student was tested on the state assessment in the subject assessed by NAEP. Therefore, if a student required an accommodation that was not allowed by NAEP, such as read-aloud of the reading assessment, they were excluded from NAEP and counted in the jurisdiction’s exclusion rate. However, according to the new policy, these students should be counted as refusals rather than exclusions.

“Students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.”

NCES has conducted an analysis of the impact of implementing this part of the policy. The analysis re-examines the assessment data from 2011, assuming that this new reporting policy were applied, but assuming that the participating states, districts, and students would not have included any additional

students in the assessment (so that the only changes are for some students who were classified as exclusions to be classified as refusals).

Some results of this analysis will be presented at the meeting, showing the unintended impact on exclusion rates, response rates, and average scale scores (overall and for subgroups). The aspects of the data collection plan for 2013 which are related to the new decision tree will also be described.

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:

NAEP Decision Tree for Students with Disabilities

BACKGROUND CONTEXT

1. NAEP is designed to measure constructs carefully defined in assessment frameworks adopted by the National Assessment Governing Board.
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.

STEPS OF THE DECISION TREE

3. In deciding how a student will participate in NAEP:
 - 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.
 - b. If the student's IEP or 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.
 - c. If the student's IEP or 504 plan specifies an accommodation or modification that is not allowed on NAEP, then the student is encouraged to take NAEP without that accommodation or modification.

3. Students should be considered for exclusion from NAEP only if they have previously been identified in an Individualized Education Program (IEP) as having the most significant cognitive disabilities, and are assessed by the state on an alternate assessment based on alternate achievement standards (AA-AAS). All students tested by the state on an alternate assessment with modified achievement standards (AA-MAS) should be included in the National Assessment.
4. Students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.
5. NAEP should report separately on students with Individualized Education Programs (IEPs) and those with Section 504 plans, but (except to maintain trend) should only count the students with IEPs as students with disabilities. All 504 students should participate in NAEP.

At present the National Assessment reports on students with disabilities by combining results for those with an individualized education program (who receive special education services under the Individuals with Disabilities Education Act [IDEA]) and students with Section 504 plans under the Rehabilitation Act of 1973 (a much smaller group with disabilities who are not receiving services under IDEA but may be allowed test accommodations).^{*} Under the Elementary and Secondary Education Act, only those with an IEP are counted as students with disabilities in reporting state test results. NAEP should be consistent with this practice. However, to preserve trend, results for both categories should be combined for several more assessment years, but over time NAEP should report as students with disabilities only those who have an IEP.

6. Only students with an IEP or Section 504 plan are eligible for accommodations on NAEP. States are urged to adopt policies providing that such documents should address participation in the National Assessment.

For English Language Learners

1. All English language learners selected for the NAEP sample who have been in United States schools for one year or more should be included in the National Assessment. Those in U.S. schools for less than one year should take the assessment if it is available in the student's primary language.

One year or more shall be defined as one full academic year before the year of the assessment.

^{*} NOTE: The regulation implementing Section 504 defines a person with a disability as one who has a physical or mental impairment which substantially limits one or more major life activities, has a record of such an impairment, or is regarded as having such an impairment. 34 C.F.R. § 104.3(j)(1).

2. Accommodations should be offered that maximize meaningful participation, are responsive to the student's level of English proficiency, and maintain the constructs in the NAEP framework. A list of allowable accommodations should be prepared by NAEP and furnished to participating schools. Such accommodations may be provided only to students who are not native speakers of English and are currently classified by their schools as English language learners or limited English proficient (LEP).
3. Bilingual versions of NAEP in Spanish and English should be prepared in all subjects, other than reading and writing, to the extent deemed feasible by the National Center for Education Statistics. The assessments of reading and writing should continue to be in English only, as provided for in the NAEP frameworks for these subjects.
4. Staff at each school should select from among appropriate ELL-responsive accommodations allowed by NAEP, including bilingual booklets, those that best meet the linguistic needs of each student. Decisions should be made by a qualified professional familiar with the student, using objective indicators of English proficiency (such as the English language proficiency assessments [ELPA] required by federal law), in accordance with guidance provided by NAEP and subject to review by the NAEP assessment coordinator.
5. Schools may provide word-to-word bilingual dictionaries (without definitions) between English and the student's primary language, except for NAEP reading and writing, which are assessments in English only.
6. NAEP results for ELL students should be disaggregated and reported by detailed information on students' level of English language proficiency, using the best available standardized assessment data. As soon as possible, NAEP should develop its own brief test of English language proficiency to bring consistency to reporting nationwide.
7. Data should be collected, disaggregated, and reported for former English language learners who have been reclassified as English proficient and exited from the ELL category. This should include data on the number of years since students exited ELL services or were reclassified.
8. English language learners who are also classified as students with disabilities should first be given linguistically-appropriate accommodations before determining which additional accommodations may be needed to address any disabilities they may have.

RESEARCH AND DEVELOPMENT

The Governing Board supports an aggressive schedule of research and development in the following areas:

1. The use of plain language and the principles of universal design, including a plain language review of new test items consistent with adopted frameworks.
2. Adaptive testing, either computer-based or paper-and-pencil. Such testing should provide more precise and accurate information than is available at present on low-performing and high-performing groups of students, and may include items appropriate for ELLs at low or intermediate levels of English proficiency. Data produced by such targeted testing should be placed on the common NAEP scale. Students assessed under any new procedures should be able to demonstrate fully their knowledge and skills on a range of material specified in NAEP frameworks.
3. A brief, easily-administered test of English language proficiency to be used for determining whether students should receive a translation, adaptive testing, or other accommodations because of limited English proficiency.
4. The validity and impact of commonly used testing accommodations, such as extended time and small group administration.
5. The identification, measurement, and reporting on academic achievement of students with the most significant cognitive disabilities. This should be done in order to make recommendations on how such students could be included in NAEP in the future.
6. A study of outlier states and districts with notably high or low exclusion rates for either SD or ELL students to identify the characteristics of state policies, the approach of decision makers, and other criteria associated with different inclusion levels.

The Governing Board requests NCES to prepare a research agenda on the topics above. A status report on this research should be presented at the November 2010 meeting of the Board.

***Technical Advisory Panel on Uniform National Rules
for NAEP Testing of Students with Disabilities***

Report to the National Assessment Governing Board

July 22, 2009

Chair: Alexa Posny

***Members: Louis Danielson, George Engelhard,
Miriam Freedman, Claire Greer, Robert Linn,
Debra Paulson, and Martha Thurlow***

Technical Advisory Panel on Uniform National Rules for NAEP Testing of Students with Disabilities

Executive Summary of Report to NAGB - July 2009

Chair: Alexa Posny

*Members: Louis Danielson, George Engelhard, Miriam Freedman,
Claire Greer, Robert Linn, Debra Paulson, and Martha Thurlow*

The panel believes the National Assessment of Educational Progress (NAEP) is an important tool for understanding academic achievement among students with disabilities. To ensure that NAEP samples are fully representative and to maintain the comparability of state and district NAEP results, the panel recommends that NAEP

1. Encourage as many students as possible to participate in NAEP, and provide for the use of allowable accommodations that are necessary to enable students with disabilities to participate.
2. Clarify and expand NAEP's guidance to schools, encouraging maximum participation of students with disabilities so at least 95% of those drawn for the NAEP sample participate.
3. Report separately on students who have individualized education programs (IEPs) and those with Section 504 plans, but (except to maintain trend) only count the students with IEPs as students with disabilities.
4. Provide incentives for schools to include students with disabilities, including additional outreach and public reporting of participation rates below 95% of students with disabilities.
5. Support research efforts to develop targeted testing for students at both the top and bottom levels of achievement, with sound procedures to identify students to receive targeted test booklets on the basis of their performance on some standard indicator of achievement.
6. Encourage and review research on the identification and progress of students who have a significant cognitive disability but in the short term do not test this 1% of students on NAEP.
7. Assess the English language proficiency of students with disabilities who are English language learners and are drawn for the NAEP sample and provide linguistically appropriate accommodations for those who need them before determining whether additional accommodations may be needed to address any disabilities those students may have.

Although NAEP can establish rules for students to be tested in the same way, individual students participate in NAEP on a voluntary basis, and it is their schools that normally make the decision about whether a student drawn for the NAEP sample participates or not. Therefore, the cooperation of schools and parents is essential to ensure that NAEP samples in every jurisdiction are fully representative and that test results are comparable among the states and districts assessed. The recommendations in this report are intended to be of practical use in determining NAEP testing procedures and in working with states and districts to continue the assessment's tradition of producing comparable results and useful information.

Technical Advisory Panel on Uniform National Rules for NAEP Testing of Students with Disabilities

Report to National Assessment Governing Board

July 22, 2009

Chair: Alexa Posny

*Members: Louis Danielson, George Engelhard, Miriam Freedman, Claire Greer,
Robert Linn, Debra Paulson, and Martha Thurlow*

BACKGROUND AND INTRODUCTION

The National Assessment of Educational Progress (NAEP) was established in 1969 to measure the academic achievement of a representative sample of elementary and secondary students in the United States. It is sometimes called the Nation's Report Card. Subsequently, the assessment was expanded to provide representative-sample results for states and large urban school districts.

NAEP is designed to produce valid, comparable data on large groups of students. It is prohibited by law from providing results for individual children or schools. Scores are not intended and (because no student takes the entire test) cannot be calculated for individual students. Because NAEP measures change over time, it can provide participating states and districts with reliable, independent information about the success of their efforts to improve education. It is an important common measure of student performance.

Recently, concern has arisen about the wide variation among states and districts in the rates at which students with disabilities participate in NAEP. Confusion can arise when in some states almost all students with disabilities who are selected for the NAEP sample take the test, and in others many do not. Some advocates for students with disabilities believe that having good information on the achievement of the full population of students with disabilities is a critical tool in improving services for them. The purpose of this report is both to increase the uniformity of NAEP participation rates among states and districts and to make participation rates high and participation procedures uniform.

Specifically, the National Assessment Governing Board (NAGB) convened a technical advisory panel to recommend a uniform set of rules for testing students with a disability on NAEP. The eight-member group held an all-day meeting in Washington, DC, on April 23, 2009, for initial briefings and discussion. The panel conducted four conference calls and exchanged numerous drafts and e-mails between May and July.

The Governing Board charged the panel to make recommendations that:

- provide that students with similar disabilities be tested on NAEP the same way, regardless of where they live;
- maximize student access and meaningful participation;
- ensure that the constructs on NAEP frameworks be measured and that all students may be placed on the same scale;
- permit only accommodations that maintain the validity, reliability, and comparability of NAEP results; and
- are feasible, logistically and financially, and without detrimental consequences.

RECOMMENDATIONS

1. Encourage as many students as possible to participate in NAEP, and provide for the use of allowable accommodations that are necessary to enable students with disabilities to participate.

The panel recommends that all students with disabilities participate in NAEP with appropriate accommodations that they need, which are approved by NAEP. The panel understands that some students will not be allowed to use on NAEP some of the accommodations or modifications that are permitted on tests administered by the state or district.

The panel defines an appropriate accommodation as:

- a change to the way NAEP is normally administered, and
- a change that does *not* alter the construct being measured, and
- a change that is needed to enable a student to take the test.

If a proposed accommodation alters the construct being measured, the panel considers it a modification. The panel defines a modification as:

- a change to the way NAEP is normally administered, and
- a change that does alter the construct being measured.

The panel recommends *against* the use of any change that would alter the construct NAEP is designed to measure, as defined by the NAEP frameworks.

The panel understands that the Governing Board defines the construct underlying the NAEP reading test as “an active and complex process that involves understanding written text.” Because the Governing Board defines this construct to include the ability to decode written text, the panel reaffirms the current NAEP practice of not allowing “read aloud” as an accommodation on the reading test.

The panel understands that the Governing Board defines the construct underlying the NAEP mathematics test as involving five elements, one of which is “Number Properties and Operations (including computation...)” Because this construct includes computation, the panel reaffirms current NAEP practice of not allowing the use of calculators on those parts of the NAEP math test that assess computation.

2. Clarify and expand NAEP’s guidance to schools, encouraging maximum participation of students with disabilities.

As stated previously, the panel recognizes that the testing rules NAEP adopts will not yield comparable state and local results if jurisdictions vary in their participation practices. The panel therefore recommends changes to the guidance given school personnel in deciding whether students drawn for the NAEP sample are to be tested. The panel recommends advising schools on the purpose and nature of NAEP and the desirability of high participation rates, and setting the clear expectation that at least 95% of all students with disabilities drawn for the NAEP sample are expected to take the test.

In a departure from past guidance, the panel recommends state and local decision makers begin with the expectation that almost all students with disabilities will take the test, and then make decisions regarding the accommodations that individual students will be allowed to have. Specifically, the panel recommends this revised Decision Tree be provided to schools:

NAEP Decision Tree for Students with Disabilities

BACKGROUND CONTEXT

1. NAEP is designed to measure constructs carefully defined by frameworks adopted by the Governing Board. Those frameworks include a definition of reading as “an active and complex process that involves understanding written text,” (including the ability to decode text) and include in its definition of mathematics five elements, one of which is “Number Properties and Operations (including computation...).”
2. NAEP provides a list of accommodations that are and are not allowed in reading, mathematics, and other subjects. [See Column B of appendix for accommodations allowed and not allowed on NAEP.]

STEPS OF THE DECISION TREE

3. In deciding how this student will participate in NAEP:
 - a. If the student has an IEP or 504 plan and is tested without accommodation, then he or she takes NAEP without accommodation.
 - b. If the student’s IEP or 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.
 - c. If the student’s IEP or 504 plan specifies an accommodation or modification not allowed on NAEP, then the student takes NAEP without that accommodation or modification.

Students should be *excluded* from participating in NAEP *only* if they have previously been identified in an IEP as having a significant cognitive disability, and are assessed by the state on an alternate assessment based on alternate achievement standards (AA-AAS). Students should be *included* if tested on an alternate test with what is called modified achievement standards (AA-MAS).

The panel recommends that guidance to school decision-makers include:

- i) a short, clear account of the purpose and value of NAEP, why the inclusion of virtually all selected students is needed to provide representative samples, and the steps to determine how a selected student should participate, and
- ii) the target for the percentage of students appropriately to be excluded from participating in NAEP would be 1% of the sample.

The panel also recommends that a broader effort at public information be undertaken to explain the value of NAEP and of securing high participation rates in the assessment.

3. Report separately on NAEP results for IEP and 504 students.

The panel recommends that NAEP report results for both IEP and 504 student groups, but report them separately, and calculate state scores for students with disabilities using IEP results only. At present the National Assessment reports on students with disabilities by combining the results for students with an individualized education program (who receive special education services under the Individuals with Disabilities Education Act [IDEA]) and those with Section 504 plans under the Rehabilitation Act of 1973 (a much smaller group who are not special education students but may be allowed test accommodations).

Under the Elementary and Secondary Education Act, only students with an IEP are counted as students with disabilities in reporting state test results. NAEP should be consistent with this practice. However, the panel recognizes the usefulness of maintaining NAEP trends, and therefore recommends reporting both sets of data and combining results for IEP and 504 students only to preserve the trend line. The panel recommends over time defining students with disabilities for NAEP as only those who have an IEP. All 504 students should participate in NAEP.

4. Provide incentives for schools to include students with disabilities.

The panel recommends that NAEP make enhanced efforts to provide a short clear description of the purpose and value of NAEP and of full student participation in the assessment. These materials should be aimed at school personnel, state officials and the general public, including the parents of students with disabilities.

The panel recommends that upon release of each new set of NAEP results, information indicating the states and districts with more or less than 95% participation rates of students with disabilities with IEPs be among the information bullets highlighted for the

public and the press. All students with 504 plans are expected to participate. Participation rates should be reported both as a percentage of the total sample and as a percentage of the students identified with disabilities within the sample.

The panel further recommends undertaking special studies to look at any outlier states, with unusually high or low exclusion rates, and to continue work previously done for NCES to probe whether there is a cut point beyond which exclusion rates appear suspect.

Some members of the panel noted that there is significant variation among the states in the rate at which they identify students with disabilities for IEPs. While on average states identify about 12-13% of their students as having a disability and needing special education services, some states identify only 9% of their students, and others identify twice that percentage. The differences result mostly from state and local policy rather than the incidence of disability itself. Generally, jurisdictions with high identification rates include more students with mild disabilities. Those with low identification rates include only the more severe, which would make it more difficult to achieve 95% SD participation even though, overall, more of their students may be taking the assessment.

As an alternative to the 95% participation guideline for students with disabilities, some members of the panel recommend that NAEP study the possibility of developing a uniform SD participation guideline based on a percentage of the total student population, regardless of the percent identified as SD. If more than the selected percentage were excluded on the basis of disability, that would be noted in NAEP reports as indicating that the sample was not fully representative. For example, a maximum of 0.6% of the total sample not tested, or 99.4% participating, would correspond to a SD participation rate of 95% where 12% of the sample is identified as having a disability.

5. Support research efforts to develop targeted testing for all students at both the top and bottom levels of achievement, with sound procedures to identify students to receive targeted test booklets on the basis of their performance on some standard indicator of achievement.

The panel recommends that research and development efforts be pursued for NAEP to test all students, not only students with disabilities, at the top and bottom levels of achievement on targeted booklets with a high concentration of difficult or easy items that can be placed on the existing NAEP scale.

Currently all students are tested by NAEP with two 25-minute blocks of items covering a broad range of difficulty, some easy, some difficult, many in the middle. Any student might be randomly assigned any of the various booklets covering the complete range of difficulty for the grade and subject in which he or she is being tested.

The National Center for Education Statistics (NCES) is now developing booklets with a concentration of existing easy items that could be targeted for low-performing students. The panel recommends building upon this research effort, if successful, to create targeted tests at *both* the top and bottom of the achievement spectrum. High-performing students, those doing work well above grade level, would encounter more challenging items that

allow them to demonstrate knowledge at the advanced level. Likewise, low-performing students would encounter more items that allow them to demonstrate knowledge at the below basic level. This would allow NAEP to measure and report more accurately and in greater detail the knowledge and skills of those students scoring below basic and those scoring advanced. At both ends of the continuum, standard errors would be reduced, and better information would be available about student performance and improvements over time. If needed, additional easy and difficult items should be developed that test NAEP constructs on the existing NAEP scale.

The panel recommends that NAGB attend closely to NCES' on-going research in this area, and base future decisions on this work and similar research by others. If targeted testing becomes part of future NAEP operations, this information should be described carefully for state and local decision makers. Efforts should be made to explain how these innovations enable students with disabilities who are studying at below basic levels and those who are studying at advanced/above grade levels to engage with NAEP at all points of the continuum of achievement.

The panel recommends that NAEP find an objective and psychometrically sound method to identify which students take any targeted tests that are developed. It recommends consideration of the following possibilities:

a) a universal 2-stage process, the system proposed by R. Darrell Bock, in which all students receive a comprehensive block first (a locator test), and then receive either a booklet with a concentration of easy items, a test with a concentration of difficult items, or the usual full-range test in the second block, depending upon their performance on the initial locator test.

While this option was the preference of many panel members, it entails major issues of test administration that need to be taken into account before the technique would become feasible.

b) a specially constructed new NAEP screener.

This would entail new development work.

c) student performance near the top or bottom percentile rank of the state's previously administered state assessment.

While several panel members were hesitant to use results of varying state assessments, existing research shows that even the widely different tests used by states produce scores that correlate well enough with NAEP to be useful in identifying top and bottom performers who would be assigned high or low blocks of items.

d) a new or different method that may emerge, which is psychometrically sound and easy to administer.

The panel wants to see the adoption of a method that is fair, feasible, objective and effective, but recognizes that considerable technical development would be required before targeted testing can become a regular part of NAEP.

The panel recommends that the assignment of a targeted test to a student be based on how the student performs on some standard indicator of achievement (such as a test), and NOT upon a student's label, such as having a disability or being in advanced placement classes. The panel intends that the availability of the easy form of the test assure participating schools that low-performing students, including students with disabilities, are able to participate without altering NAEP standards. Likewise, high-performing students could be challenged on items in the assessment at the greater level of difficulty.

6. Encourage and review research on the identification and progress of students who have a significant cognitive disability but in the short term do not test this 1% of students on NAEP.

The Panel recommends that NAGB form a panel of experts and stakeholders to review research and best current practices for identifying, measuring and reporting the progress of students who have a significant cognitive disability, and to make recommendations to NAGB for how emerging findings can and should be applied to NAEP in the future so such students could be included in NAEP.

The panel believes that NAEP should encourage the appropriate assessment of all children, but recommends that for the near future students with a severe cognitive disability—about 1% of the student population—be excluded from NAEP. The exclusion of these students should not be considered in determining whether a jurisdiction meets participation rate guidelines.

7. Assess the English language proficiency of students with disabilities drawn for the NAEP sample and provide NAEP-approved, linguistically appropriate accommodations for them before determining whether additional accommodations may be needed to address any disabilities these students may have.

Some students drawn for the NAEP sample will be both English language learners and students with disabilities. For these students it is important first to determine the level of their English proficiency, and the accommodations allowed for them on NAEP. If these students have also been identified as having a disability and are eligible to receive special education services, they should receive whatever accommodations are allowed by NAEP that they need to participate in the NAEP assessment.

APPENDIX A

LIST OF MEMBERS AND AFFILIATIONS



Technical Advisory Panel on Uniform National Rules for NAEP Testing of Students with Disabilities

- Alexa Posny, Kansas Commissioner of Education (Chair)
Former Director, Office of Special Education Programs
U.S. Department of Education
- George Engelhard, Jr.
Professor of Educational Studies (Educational Measurement and Policy)
Emory University, Atlanta, GA
- Louis Danielson, Managing Director, American Institutes for Research
Former Director, Research to Practice Division, Office of Special Education
Programs, U.S. Department of Education
- Miriam Freedman, attorney and author
Stoneman, Chandler & Miller, Boston, MA
- Claire Greer, Consultant for Autism, Severe, and Multiple Disabilities
Exceptional Children Division
North Carolina Department of Public Instruction
- Robert Linn, Professor of Education (Emeritus)
Research and Evaluation Methods Program
University of Colorado
- Debra Paulson
Middle school math and special education teacher
El Paso, TX.
- Martha Thurlow, Director
National Center on Educational Outcomes
University of Minnesota

APPENDIX B

ACCOMMODATIONS ALLOWED ON NAEP

On state assessment this student:	COLUMN A	COLUMN B			
		Accommodations allowed on NAEP			
		Reading	Math	Science	U.S. history or geography or civics
Presentation Format					
Has directions read aloud/repeated in English or receives assistance to understand directions	<input type="radio"/>	Standard NAEP practice			
Has directions only signed	<input type="radio"/>	Y	Y	Y	Y
Has test items signed	<input type="radio"/>	N	Y	Y	Y
Has occasional words or phrases read aloud	<input type="radio"/>	N	Y	Y	Y
Has all or most of the test materials read aloud	<input type="radio"/>	N	Y	Y	Y
Uses a Braille version of the test	<input type="radio"/>	Y	Y	Y*	Y*
Uses a large print version of the test	<input type="radio"/>	Y	Y	Y*	Y
Uses magnifying equipment	<input type="radio"/>	Y	Y	Y*	Y
Response Format					
Responds in sign language	<input type="radio"/>	Y	Y	Y	Y
Uses a Braille typewriter to respond	<input type="radio"/>	Y	Y	Y	Y
Points to answers or responds orally to a scribe	<input type="radio"/>	Y	Y	Y	Y
Tape records answers	<input type="radio"/>	N	N	N	N
Uses a computer or typewriter to respond	<input type="radio"/>	Y	Y	Y	Y
		Spell/grammar check not allowed			
Uses a template to respond	<input type="radio"/>	Y	Y	Y	Y
Uses a large marking pen or special writing tool	<input type="radio"/>	Y	Y	Y	Y
Writes directly in the test booklet	<input type="radio"/>	Standard NAEP practice			
Setting Format					
Takes the test in a small group ²	<input type="radio"/>	Y	Y	Y	Y
Takes the test one-on-one ²	<input type="radio"/>	Y	Y	Y	Y
Takes the test in a study carrel	<input type="radio"/>	Y	Y	Y	Y
Receives preferential seating, special lighting, or furniture	<input type="radio"/>	Y	Y	Y	Y
Must have test administered by familiar person	<input type="radio"/>	Y	Y	Y	Y
Timing Accommodations (Note: NAEP takes only 90 minutes.)					
Receives extended time ⁴	<input type="radio"/>	Y	Y	Y	Y
Is given breaks during the test	<input type="radio"/>	Y	Y	Y	Y
Must be allowed to take subject test over several days	<input type="radio"/>	N	N	N	N
Other Accommodations					
Uses a calculator, including talking or Braille calculator for computation tasks	<input type="radio"/>	NA	N*	NA	NA
Uses an abacus, arithmetic tables, graph paper	<input type="radio"/>	NA	N	NA	NA
Uses dictionary, thesaurus, or spelling/grammar-checking software or devices	<input type="radio"/>	N	N	N	N
Receives the following accommodation(s) not listed above.	<input type="radio"/>	Check with your NAEP representative			



Reporting Puerto Rico Results: Knowledge and Skills Accessible Study (KaSA)

As part of the 2011 assessment, the Knowledge and Skills Appropriate (KaSA) study was conducted in Puerto Rico and the mainland U.S. to evaluate recent efforts to increase the measurement precision in the estimation of student ability at the lower end of the NAEP scale, while still administering an assessment that is consistent with the NAEP mathematics framework. As part of the study, KaSA items were developed to address a targeted subset of the NAEP mathematics framework, based on the appropriateness of subtopics and objectives. While KaSA items are written to address framework objectives, the pool of items does not span the breadth of the framework. In terms of item types, the number of multiple-choice items is relatively large in the KaSA item pool and approximately 70 percent of the items are of low mathematical complexity, while the remainder are of moderate complexity. In comparison, operational assessments have a target of 25 percent low complexity. For each grade, 60 KaSA items were developed and placed in four 15-item KaSA blocks. The KaSA items were translated into Puerto Rican Spanish for administration in Puerto Rico.

The KaSA items and blocks functioned well in 2011 and were, on average, easier than operational items. However, it is critical to evaluate the success of KaSA in terms of trend results. Therefore, a repeat of the study, using the same KaSA blocks and instrument and sample design, is planned for 2013. The hope and expectation of the 2013 study is that the encouraging results of 2011 can be replicated in Puerto Rico in 2013, which would establish the level of confidence and validity desired to report (trend) results for Puerto Rico.

If results from the 2013 KaSA study are successful, then Puerto Rico could receive reports that are similar to those that other states receive, including:

- A one-page, user-friendly state snapshot report that provides scores for students overall and selected student groups as well as score changes between 2011 and 2013 for Puerto Rico
- The ability to populate the state report generator with data for Puerto Rico so representatives from the jurisdiction's department of education could use the tool to assemble customized reports

While scores can be reported for Puerto Rico (if the study indicates that score estimation is reliable), it should be noted that there is a necessary delay in reporting scores for Puerto Rico. This delay in reporting is due to the fact that estimating the scores for Puerto Rico requires the establishment of the operational scores for the nation and states.