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

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

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

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

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.

- Tracking progress in implementing key instructional, curricular, and technological changes and educational policy initiatives, such as the Common Core standards.

- Monitoring student motivation and out-of-school learning as research-based factors affecting student achievement.

- Benchmarking high-performing states and urban districts and those with high achievement growth to identify factors that differentiate high-performers from lower-performers on NAEP. This domestic effort would parallel the extensive reporting of background variables in PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) that have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing and low-performing education systems.

The panel proposes building a strategy to make the NAEP background questions an important national resource for educators, policymakers, and the public. The panel sees the need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions though regular publications, and improve the capacity for analysis by users around the world.

We offer recommendations in four areas (see Exhibit A):

(1)   Ask Important Questions.
(2)   Improve the Accuracy of Measures.
(3)   Strengthen Sampling Efficiency.
(4)   Reinstitute Meaningful Analysis and Reporting.

Exhibit A. Expert Panel Recommendations to Strengthen NAEP Background Questions in Four Areas

1. Ask Important Questions	2. Improve the Accuracy of Measures	3. Strengthen Sampling Efficiency	4. Reinstitute Meaningful Analyses & Reporting
•Core questions •Rotated questions •Policy questions •Theoretical frameworks •Consistent questions overtime •Delete duplicative or low-priority questions	•Valid •Reliable •Coordinated (with domestic and international surveys) •Cognitive labs	•Spiral sampling •Extended questionnaire time •Alternate surveys •Pooling item responses across surveys	•Special background question reports •Online compendium of responses •Report descriptive not causal findings •Externally conducted research •Improve online tools
• 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 Particpating in the 2011 Trial Urban Development Assessment

Jurisdictions	Grade 8 All Students	Eligible for National School Lunch	Grade 8 Students Absent 5 or more days last month	Grade 8 Students in Algebra	Grade 8 Students 5 or more Hours of Math Per Week	Grade 8 Students 1 Hour or More Math Homework	Grade 8 Does Math At An 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 Avaiable to Teachers and Stundents
	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 though in-depth study. Specific examples of the use of NAEP background questions for domestic benchmarking might include examining:

- A high overall-performing state such as Massachusetts or a state like Texas that has a relatively small white-Hispanic performance gap compared with other states.
- A high-performing district such as New York City that has low-income students achieving above the national average for all low-income students in both reading and math at grades 4 and 8.
- The nearly one standard deviation growth in grade 4 math since 1990 and the instructional, curriculum and teacher changes that occurred over this period.

The panel recognizes the justifiable concern over misuse of the NAEP background variables in making causal interpretations. NAEP is not able to reduce countervailing explanations for causation like a well-designed experiment. Also, successive NAEP assessments will sample different students in the same grade, so the data are not a measure of change over time for the same students as in a true longitudinal design. However, the panel believes that a valid concern over causal interpretations has led to a serious and unjustified overreaction. NAEP's national and state representative data uniquely address many important descriptive questions. These data can track progress on variables shown by research to be important for achievement. The NAEP background questions can inform national policies by providing descriptive data about the quality of implementation. Also, because NAEP is already in the schools to administer its assessments, data can be collected at relatively low cost compared with other survey vehicles.

Yet for the past decade NAEP has stopped publishing all but the most minimal background information.

- NAEP no longer systematically reports on the responses to the background questions when publishing its assessment results, except for the congressionally required student reporting categories (e.g., race/ethnicity, low-income).[i2]
- In-depth special reports using the background questions are rare (e.g., the 2010 report on American Indian Educational Experiences was an exception).
- Data are made available almost entirely through an online database called the NAEP Data Explorer. This is a useful tool, but it is not a substitute for carefully prepared summary data tables and analyses. Most educators, policy makers and members of the public do not have the time or inclination to master use of the Data Explorer, but many would pay attention to focused reports and make use of summary tabular information.

Reporting the background questions would be a great service to the nation in identifying and tracking important national and state trends in education. Here, the panel finds that the NAEP background questionnaires severely limit their potential usefulness by not explicitly asking questions about the progress and challenges of implementing key national policies in different states and urban districts. Yet the *NAEP Background Information Framework* (2003), which sets out principles to guide background question selection and reporting, explicitly recognizes that the background questions should "focus on the most important variables related to public policy."

NAEP's de-emphasis of the background questions is in marked contrast to the significance that all the major international surveys – PISA (Program for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), and PIRLS (Progress in International Reading Literacy Study) – give to background

---

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

variables in participating countries.

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

To do so the panel has developed recommendations for improvements in four areas:

(1) Ask Important Questions.
(2) Improve the Accuracy of the Measures.
(3) Strengthen Sampling Efficiency.
(4) Reinstitute Meaningful Analysis and Reporting.

Within each area, Exhibit 1 identifies the specific individual recommendations.

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
•Core questions •Rotated questions •Policy questions •Theoretical frameworks •Consistent questions overtime •Delete duplicative or low-priority questions	•Valid •Reliable •Coordinated (with domestic and international surveys) •Cognitive labs	•Spiral sampling •Extended questionnaire time •Alternate surveys •Pooling item responses across surveys	•Special reports •Online compendium of responses •Report descriptive not causal findings •Externally conducted research •Improve online tools
• 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.

<u>Discussion</u>

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).

<u>Discussion</u>

The 2003 *Background Information Framework* for NAEP states the principle that "Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be "directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information." NAEP to its credit employs panels involving contractors and multiple external groups in its question development.

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

NCES has a good start toward building the necessary research foundation for developing such frameworks in the papers prepared by the Education Testing Service (ETS). ETS (2010) has developed three in-depth literature reviews, one each to support the topics currently or potentially addressed in the student, teacher and school questionnaires. The student and school questionnaire reviews also compare the current NAEP content items with the content measured in other large-scale national and international assessments.

The panel's proposal would build-on the current literature reviews by:

- Using the research to develop theoretical frameworks that identify for major topics the component variables around which to build clusters of questions. The current ETS literature reviews although useful, are largely a description of discrete findings. Exhibit 3 is an example of how PISA presents a research-based, theoretical framework to organize background questions around the components of student engagement in reading and reading strategies. In this example, PISA operationalizes engagement in reading in terms of five components: reading for school, enjoyment of reading, time spent reading for enjoyment, diversity of reading materials, and diversity of online reading activities. Multiple questions then ask students about their reading behaviors with respect to each component.

- Organizing literature reviews around topics, which is preferable to the current organization around three separate questionnaires. Some topics may cut across the student, teacher and school questionnaires. For example, the current ETS literature review considers family involvement only in terms of the student questionnaire and the items describing home learning activities and resources. A broader research-based theoretical framework around the issue of parental involvement would extend the construct to include how teachers and schools reach out and support families, not just what families do by themselves. Indeed, Title I longitudinal evaluations have shown that student achievement improves when schools reach out and support parental involvement. (USED, 2001).

Once developed, these research-based frameworks would form the basis for developing valid and reliable questions to measure the different aspects of a topic domestically and to coordinate measurement with major international surveys. (Section 2 below).

Exhibit 3. PISA Analytic Framework for Student Engagement in Reading and Learning Strategies to Inform Decisions about Improving Reading
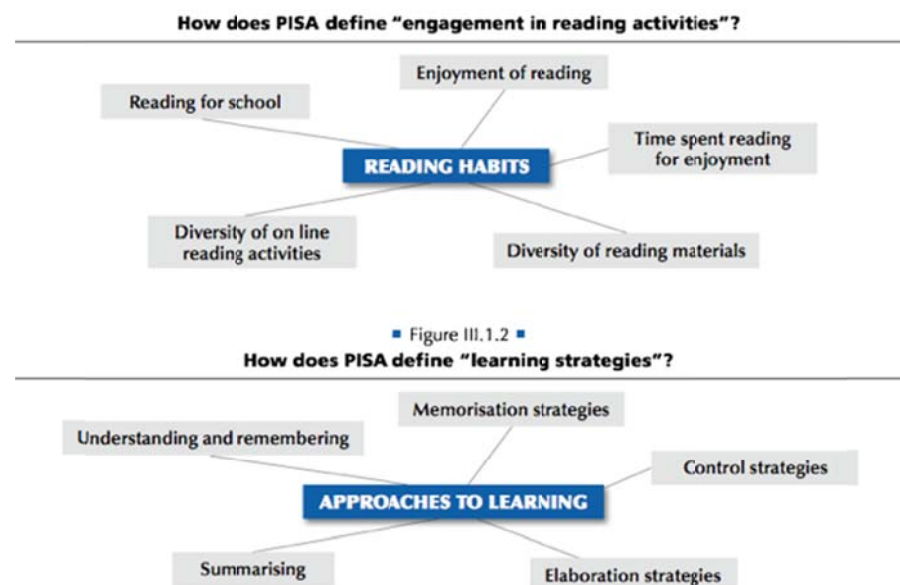
Source. OECD, PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices

Recommendation 1d. Use consistency over time as a criterion to consider for question selection and wording.

- NAEP's inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.
- Recognizing that NAEP needs to periodically refresh its question set, nonetheless NAEP question selection seems haphazard – important questions may not be asked for two or more assessments and then they may reappear with changed wording that disrupt the time series reporting.
- Rather than total eliminating some potentially important survey questions on a topic, NAEP should consider rotating questions so that a question may be asked only once every 4-6 years.
- When rewording is necessary, NAEP should do *bridge studies* to link the new question responses with prior ones to form an unbroken time series of responses.

<u>**Discussion**</u>

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

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.

<u>Discussion</u>

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					
	Not useful at all					Very useful
	(1)	(2)	(3)	(4)	(5)	(6)
a) I concentrate on the parts of the text that are easy to understand.	O_1	O_2	O_3	O_4	O_5	O_6
b) I quickly read through the text twice.	O_1	O_2	O_3	O_4	O_5	O_6
c) After reading the text, I discuss its content with other people.	O_1	O_2	O_3	O_4	O_5	O_6
d) I underline important parts of the text.	O_1	O_2	O_3	O_4	O_5	O_6
e) I summarize the text in my own words.	O_1	O_2	O_3	O_4	O_5	O_6
f) I read the text aloud to another person.	O_1	O_2	O_3	O_4	O_5	O_6

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".

<u>Discussion</u>

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.

<u>Discussion</u>

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	528	21	561	58	576	16
U.S. 429	5	430	23	465	20	511	38	480	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.

<u>Discussion</u>

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.

<u>**Discussion**</u>

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.

<u>Discussion</u>

These sample reforms could expand the number of background items surveyed over a multi-year period, while maintaining accurate State-by-state reporting of background questions. However, each involves its own tradeoffs in terms of questionnaire time and the availability of items on any one survey. The panel requests that NCES examine and report to NAGB the comparative strengths and weaknesses of different approaches to expanding questionnaire items.

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

Rich responses to relevant background questions would mean little if NAEP continues its present practice of including very few findings from the background questionnaires in its reports. The main exception is the reporting of achievement by the congressionally required student subgroups. For other background information, the only recourse for a potential user to these data is to conduct one's own analyses using the NAEP Data Explorer. As a practical matter, this is an option that only professional researchers (and few others) will have the time and skills to undertake.

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

---

[6] *TIMSS 2011 Assessment Design* (p126) describes expected student testing time at grade 4 of 72 minutes for the student achievement booklet and 30 minutes for the student questionnaire. The grade-8 times are 90 minutes for the student achievement booklet and 30 minutes for the student questionnaire

Recommendation 4a. Prepare special reports highlighting the background question findings.

- The special reports would provide interested readers with key findings derived from the background questions. These special reports could be prepared and released either with the achievement report or during the interval between assessment administrations. The Panel recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. A third report that explores benchmarking to find correlates of high-performing states and districts should also be considered.
- These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

<u>Discussion</u>

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 Equation [7]

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

Jurisdictions	Grade 8 All Students (Scale Score)	Eligible for National School Lunch (Percentages)	Grade 8 Students Absent 5 or more days last month (Percentages)	Grade 8 Students in Algebra (Percentages)	Grade 8 Students 5 or more Hours of Math Per Week (Percentages)	Grade 8 Students 1 Hour or More Math Homework (Percentages)	Grade 8 Does Math At An Afterschool or Tutoring Program (Percentages)	Grade 8 Entered Math Through Alternative Certification (Percentages)	Grade 8 Teacher Has Math Major/Minor/Special Emphasis (Percentages)	Grade 8 Full-time Math Specialist At School (Percentages)	Grade 8 Assigned To Math By Ability (Percentages)	Grade 8 26+ Students in Math Class (Percentages)	Grade 8 Computers Available to Teachers and Students (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	68	78	39	30	13	89	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	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	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	48	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	29	89	28	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	68	48	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.

<u>**Discussion**</u>

Other statistical agencies routinely support in-depth analyses of their statistical data. For example, the Bureau of Labor Statistics (BLS) has its own employment research and

program development staff to conduct original research using BLS data. The ASA/NSF/Research Fellow program is jointly supported by American Statistical Association and The National Science Foundation with participation of the U.S. Census Bureau, and the Bureau of Economic Analysis. This program jointly supports a Federal Statistics Fellowship program bringing academic researchers to work with statisticians and social scientists in the three federal agencies for up to one year.

NAEP should consider launching a similar program through small grants ($10,000-$50,000) competitively given to independently conduct research using NAEP data including the background questions. The focus of this research would be primarily on measurement and other statistical issues to improve the election and quality of the background variables.

The Panel also suggests that NAEP consider various strategies for encouraging and supporting outside researchers to conduct analyses of the NAEP data. NCES may want to work cooperatively with other organizations and foundations in these efforts. For example, NCES partially supported with foundations the widely cited research by Grissmer (2000) to analyze the state-level NAEP repeated time series achievement and background questions to examine the impact of systemic reform on improved achievement.

Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.

- NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that may be analyzed though software such as Excel.
- Extend the Data Explorer to facilitate the manipulation and analyses of the background questions by themselves without the achievement results. Extending software to build-in multivariate analyses should be considered.

Discussion

NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that is analyzable though software such as Excel. Each NAEP table and chart contains useful breakouts of the overall assessment and background data, which have been extracted and organized to focus on particular topics. Analysts and researchers may want to build off these tables to add more data series, conduct descriptive statistical analyses or pull apart and regroup the data to emphasize different points. Currently, NAEP offers no direct means to work off of the tables and charts in the reports other than to reenter the data by hand or to try and recreate them using the NAEP Data Explorer.

The Panel urges NAEP reporting to follow the lead of PISA by attaching a "statlink" to a downloadable excel file of the data in the table so that the user is able to access directly the data content without burdensome data reentry. Exhibit 9 shows how statlink was used to highlight the U.S. score compared with Singapore. The published PISA chart was

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
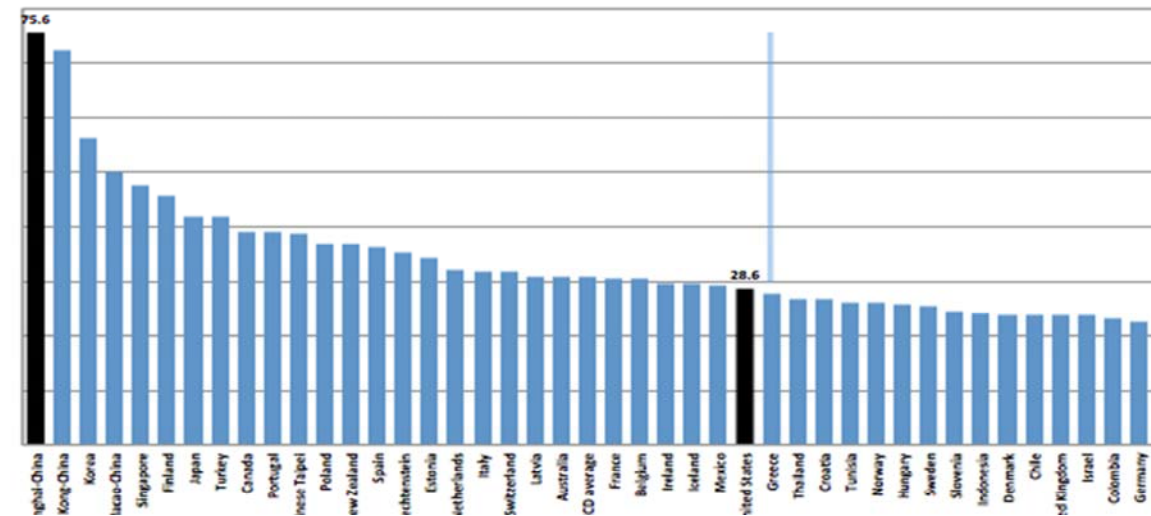
modified to highlight the gap between the U.S. compared with top performing Singapore in the performance of the bottom quarter of the most disadvantaged students (low SES) within each country who achieve in the top quarter on PISA.

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

		Less than 3 hours		3-4.9 hours		5-6.9 hours		7 hours or more	
Year	Jurisdiction	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.

<u>Discussion</u>

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.

<u>Discussion</u>

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	Recom-mendation	Activities/Products
1.. *Select core and rotated clusters of questions around research-based theoretical frameworks* 1a, 1c		• Identify 1st tier core clusters (student sub-groups student learning opportunities, student motivation) • Identify 2nd tier rotated questions • Publish background questions with research-based justifications for question clusters
2. *Extend NAEP Background Questionnaires to monitor topics of current policy interest* 1b		• Identify current and future policy issues that are suited for NAEP Background Question (Common Core, Teacher evaluation, online instruction. • Propose rotating cycle of 3 major policy areas beginning with 2013 assessment.
3. *Launch a process for the continual examination of the validity, reliability, efficiency, and consistency of measures* 2a,2b,2c 1d, 2f		• Report on validity & reliability of SES & responses at different age levels • Implement quality review procedures for reliability and consistency of questions. • Launch a cognitive laboratory capability with possibly an available small standing supplementary panel.
4. *Report on item sampling reforms to incorporate extended question sets and topics including eliminating duplicative and low-priority items* 3a, 3b		• Report on a strategy to add questions for cluster analyses and policy issues through questionnaire spiraling, alternating questions across assessment administrations, adding extra questionnaire time and eliminating low-priority items,

Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.

Discussion

While the Data Explorer is an excellent tool for online access of NAEP achievement data, addressing weaknesses in the analyses and display of the background data in the Data Explorer and publications would extend the usefulness of NAEP's current online tools.

- Simplify and clarify how to use the Data Explorer to analyze the distribution of responses on background questions.
- Explore the potential for conducting multivariate analyses through the Data Explorer
- Build links that allow the data in tables and charts in NAEP publications to transfer to excel spreadsheets for further analyses.

Recommendation 5d. Promote implementation by creating a single Governing Board committee responsible for all background questions; provide adequate resource support, while ensuring efficient resource use; and publicize background question products and findings.

<u>Discussion</u>

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 Dcember 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