# National Assessment Governing Board

## Ad Hoc Committee on NAEP Background Information

### November 29, 2012
### 2:00 p.m.-4:00 p.m.

## AGENDA

| | | |
|---|---|---|
| 2:00 – 2:15 pm | Welcome and Introductions<br>Overview of Committee Work<br>*Terry Holliday, Committee Chair* | Attachment A |
| 2:15 – 2:50pm | Measuring and Reporting on Socio-Economic Status<br>*William Ward, NCES*<br>*Charles Cowan, Analytic Focus, LLC* | Attachment B |
| 2:50 – 3:20 pm | Exploratory Analyses of NAEP Data<br>*Alan Ginsburg, Consultant* | Attachment C<br>To be sent separately |
| 3:20 – 3:50 pm | Implementation of Board Policy on NAEP Background Questions and Use of Contextual Data in Reporting<br>*James Deaton, NCES* | |
| 3:50 – 4:00 pm | Next Steps<br>*Committee Members* | Attachment D |

# Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting

## INTRODUCTION

By statute, the purpose of the National Assessment of Educational Progress is to provide a "fair and accurate" measure of student achievement and achievement trends. Academic or cognitive questions are its primary focus; the American public is its primary audience. However, in addition to reporting on what American students know and can do, NAEP has collected data for more than 40 years that provide a context for reporting and interpreting achievement results. According to the statute, such factors, both in and out of school, must be "directly related to the appraisal of academic achievement."

In each assessment NAEP administers background questionnaires for students, their teachers, and schools. The questionnaires deal with educational experiences and other factors, such as teacher training or out-of-school learning activities, that are related to academic achievement. Data on several hundred background or noncognitive variables are available on the Internet through the NAEP Data Explorer. However, for more than a decade, little use has been made of this information in NAEP reports. The data have received minimal attention and had little impact despite the considerable efforts expended in developing and approving questionnaires and collecting and tabulating responses.

In October 2011 the National Assessment Governing Board convened an expert panel to recommend how to make better use of existing NAEP background questions and to propose an analytic agenda for additional topics and questions that would be useful in developing education policy and of value to the public. The panel report, entitled, *NAEP Background Questions: An Underused National Resource*, was presented to the Board in March 2012 by Marshall Smith, former U.S. Under Secretary of Education, who chaired the six-member panel.

Many of the panel recommendations build on the *Background Information Framework for the National Assessment of Educational Progress*, adopted by the Governing Board after it received final authority from Congress over non-cognitive items on the assessment. The framework was adopted in 2003, but has not been fully implemented.

The following policies are based on recommendations by the expert panel. The Board has also taken into consideration a wide range of public comment and the analysis provided by the National Center for Education Statistics.

It is important to understand that the National Assessment is not designed to show cause-and-effect relationships. Its data should not be used to "prove" what schools should do. But, as the *Background Information Framework* declares, NAEP's "descriptions of the educational circumstances of students…, considered in light of research from other sources, may provide important information for public discussion and policy action." The Board believes the National Assessment should improve upon its efforts to collect contextual information and present it clearly to the public, which will add to NAEP's value to the nation.

## POLICY PRINCIPLES

1. NAEP reporting should be enriched by greater use of contextual data derived from background or non-cognitive questions asked of students, teachers, and schools. Such data will be used both in regular Report Cards and in special focused reports.

2. Reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Care should be taken not to suggest causation.

3. Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in background questionnaires and their connection to student achievement. Such frameworks should be updated for each assessment cycle and provide the basis for new topics and questions.

4. An ad hoc committee of the Board will be established for one year to monitor implementation of this resolution, review the *NAEP Background Information Framework*, and recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP.

## IMPLEMENTATION GUIDELINES

### *For Questions and Questionnaires*

1. Clusters of questions will be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning, which could be used regularly or rotated across assessment cycles.

2. Modules will be prepared for special one-time studies to provide descriptive information on issues of current policy interest.

3. A thorough review will be conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions will be dropped.

4. NAEP will include background questions from international assessments, such as PISA and TIMSS, to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

5. Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest.

6. The development and use of improved measures of socio-economic status (SES) will be accelerated, including further exploration of an SES index for NAEP reporting.

***For Data Collection***

7. The maximum time for students to answer the background questionnaire will be increased from 10 to 15 minutes on new computer-based assessments. Consideration should be given to a similar increase in paper-and-pencil assessments.

8. Whenever feasible, assessment samples should be divided (spiral sampling) and background questions rotated in different years in order to cover more topics without increasing respondent burden. These practices will be initiated in the assessments of reading and mathematics, which are conducted frequently, and considered for other subject areas if the frequency of testing permits.

***For Reporting***

9. Special focused reports with data through the 2013 assessment will be issued on the following topics: private schools, charter schools, gender gaps, and black male students. Reports shall include significant contextual information as well as cognitive results. Advisory committees, composed of a range of knowledgeable persons, may be appointed to provide input on reporting issues.

10. Exploratory analyses will be carried out to determine if existing background questions may form the basis for additional focused reports. Such reports may be issued by the Governing Board as well as by the National Center for Education Statistics.

11. The NAEP Data Explorer should be further improved to make data more accessible to general, non-specialist users. Tables and very simple-to-construct charts will be prepared to present data on important topics of wide public interest. Additional means of disseminating information, using new technology such as simple apps that would allow parents, teachers, and others to access background and achievement data, will be explored.

# National Assessment Governing Board

## Ad Hoc Committee on NAEP Background Information

### BACKGROUND AND PURPOSE

As part of the resolution on NAEP background questions, adopted by the Governing Board in August 2012, an Ad Hoc Committee on NAEP Background Information is to be established for one year. Its purposes are as follows:

1. Monitor implementation of the Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting, adopted August 4, 2012.

2. Review the *NAEP Background Information Framework*, adopted August 1, 2003. Recommend revisions, additions, or replacement, as deemed necessary or desirable.

3. Recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP. This work is now divided between the Reporting and Dissemination and Assessment Development committees of the Board.

### SPECIFIC ISSUES

As listed in the resolution, these include:

- Making greater use of contextual data in NAEP Report Cards and focused reports.
- Using background data to describe patterns and trends, including the educational experiences of different student groups.
- Detailed frameworks to support the selection of non-cognitive topics and questions, including their connection to student achievement.
- Clusters of questions on topics of continuing interest, such as technology and out-of-school learning, to be used regularly or rotated across cycles.
- Modules on issues of current policy interest.
- Elimination of duplicative, low-priority, or unproductive topics and questions.
- Use of questions from international assessments, such as TIMSS and PISA.
- Improved measures of socio-economic status (SES), including exploration of SES index.
- Spiral sampling and rotation of background questions in different years.
- Increasing the maximum time for students to answer background questions.
- Additional focused reports with the appointment of advisory committees in some cases.
- Exploratory analyses of existing data that may form the basis for subsequent reports.
- Consistency of wording to preserve trends.
- Further improvements in the NAEP Data Explorer.

## COMMITTEE COMPOSITION AND TIMELINE

The Ad Hoc Committee will include six or seven Board members with a variety of perspectives and membership in different standing committees of the Board. The Committee will convene during each quarterly meeting of the Governing Board, and is expected to make its final report in August 2013.

# *Background Information*

# *Framework for the*

# *National Assessment*

# *of Educational Progress*

*EXCERPTS*

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

**National Assessment Governing Board**

Darvin M. Winick
*Chair*

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

Charles E. Smith
*Executive Director*

Lawrence Feinberg
*Project Officer*

**Background Framework Project**

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

David Grissmer
*RAND*

Consultants

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

# Preface

***by the National Assessment Governing Board***

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

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

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

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

The framework reflects the following key principles:

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

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

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

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

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

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

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

Because No Child Left Behind places particular emphasis on closing the persistent performance gaps between various student groups, NAEP must be able to report on changes in achievement for all groups specified by law. Through its background questions, the National Assessment might also provide useful information about the students left behind and those who are ahead of them, including the sorts of schools that high-achieving and low-achieving students attend, the courses they take, the patterns of

how they are taught, and the qualifications of their teachers. Over time, such descriptive information will allow NAEP to track changes in contextual and instructional factors related to student achievement and in the distribution of important educational resources.

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

# Executive Summary

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

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

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

This background data framework will refocus the collection of non-cognitive variables on NAEP's primary mission: providing a fair and accurate measure of student achievement and on achievement trends over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results. NAEP may contribute to research into improving education policy and practice, but its role in this respect is limited and the framework is not a comprehensive list of possible factors to explore.

Since by law NAEP may only collect information that is "directly related to the appraisal of academic achievement," it must concentrate on non-cognitive variables that are known from other research to have such a relationship. The law also specifically prohibits NAEP from asking about personal or family beliefs and attitudes. These points are emphasized in the Governing Board Policy Statement on the Collection and Reporting of Background Data by the National Assessment (adopted on May 18, 2002). That policy is incorporated into this framework. It is attached in the appendix.

## PRIORITIES

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

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

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

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

## SELECTION CRITERIA

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

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

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

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

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

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

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

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

- ***Will the information obtained be of value in understanding academic performance and taking steps to improve it?*** This is a fundamental issue to be addressed in evaluating all background questions proposed for NAEP.

## DATA COLLECTION

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

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

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

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

periodically, and the use of data from non-NAEP sources should increase. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.

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

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

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

## REPORTING

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on patterns and trends of non-cognitive variables known to have a relationship to academic achievement and may contain disaggregated data on school conditions and practices for various groups of students. Data on courses taken before NAEP assessments (either from transcripts or questionnaires) is of great public interest and can be related to academic results.

In addition, supplemental reports may be prepared that focus on particular aspects of the background data collected. In all cases, NAEP reports published by the National Center for Education Statistics must not state conclusions as to cause and effect relationships and avoid simplistic presentations that imply best practice.

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

## RESEARCH

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

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

The primary purpose of NAEP is to provide fair and accurate information on student achievement. Its primary audience is the American public. The Governing Board believes that in serving its purpose and audience well, NAEP can contribute to educational research. It welcomes the interest and efforts of researchers.

# A NEW MEASURE OF SOCIOECONOMIC STATUS (SES) FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Submitted to the National Assessment Governing Board
by the National Center for Education Statistics

As part of the *Background Information Framework for the National Assessment of Educational Progress*, the Governing Board called for improvement in the measurement of socioeconomic status (SES) and its relationship to academic achievement as measured by NAEP. To date, NCES and its contractors have conducted several research initiatives and have enlisted the contributions of external experts to assist in accomplishing this goal.

On behalf of NCES, a panel of experts was convened and met three times in 2010 to 2012 to discuss and provide recommendations for a new SES, with a focus on theoretical aspects of SES measurement. Panel members were invited on the basis of expertise in the fields of socioeconomic measurement, education, statistics, poverty, economics, and sociology. The panel developed a consensus definition of SES, and proposed both a core SES measure, reflecting parental education, parental occupation, and family income, and an expanded measure reflecting in addition, school and neighborhood factors. The panel reviewed the history of SES measurement, suggested various approaches to measuring its components, and discussed issues and methods for combining SES components into a single composite for reporting.

The panel concluded with a series of six recommendations covering core and expanded definitions of SES, the importance of conducting research on explanatory variables associated with SES and on the use of subjective SES in educational applications, the value of a composite measure of SES, and the importance of conducting further research on linking to U.S. Census Bureau data to improve the measurement of SES. The panel summarized their deliberations and recommendations in a white paper.

The background of the initiative and the white paper will be presented at the meeting.

# A NEW MEASURE OF SOCIOECONOMIC STATUS (SES) FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

## Expert Panel Members

- Charles D. Cowan, Analytic Focus, LLC
- Robert M. Hauser, University of Wisconsin-Madison; National Research Council
- Robert Kominski, U.S. Census Bureau
- Henry M. Levin, Teachers College, Columbia University
- Samuel R. Lucas, University of California-Berkeley
- Stephen L. Morgan, Cornell University
- Margaret Beale Spencer, University of Chicago
- Christopher Chapman, National Center for Education Statistics (Ex officio)

# IMPROVING THE MEASUREMENT OF SOCIOECONOMIC STATUS FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS: A THEORETICAL FOUNDATION

## Contents

## EXPERT PANEL

*Charles D. Cowan*
Partner and Chief Executive Officer
Analytic Focus, LLC

*Robert M. Hauser*
Vilas Research Professor Emeritus
Department of Sociology, University of Wisconsin-Madison

Executive Director
Division of Behavioral and Social Sciences and Education
National Research Council

*Robert A. Kominski*
Assistant Chief for Social Characteristics
Social, Economic, and Housing Statistics Division, U.S. Census Bureau

*Henry M. Levin*
Professor of Economics and Education
Teachers College, Columbia University

*Samuel R. Lucas*
Professor
Department of Sociology, University of California-Berkeley

*Stephen L. Morgan*
Jan Rock Zubrow '77 Professor in the Social Sciences
Department of Sociology, Cornell University

*Margaret Beale Spencer*
Marshall Field IV Professor of Urban Education
Social Sciences Division, Comparative Human Development
University of Chicago

*Chris Chapman*
Acting Associate Commissioner
Early Childhood, International, and Crosscutting Studies Division
National Center for Education Statistics (Ex officio)

# EXECUTIVE SUMMARY

At the request of the National Assessment Governing Board (NAGB, 2003), the National Center for Education Statistics (NCES) convened a panel of experts to provide recommendations concerning socioeconomic status (SES) as a construct, with the understanding that their recommendations might ultimately lead to a new measure of SES that could be used for the National Assessment of Educational Progress (NAEP). The current, most prominent NAEP measure of student SES, National School Lunch Program (NSLP) eligibility, has become less valid over time. The panel's main focus was on the theoretical foundations of SES.

## Objectives

Several objectives guided the panel's recommendations:

- provide a definition of SES,
- identify components of SES,
- review data collection and measurement approaches,
- create an SES composite, and
- consider implications of a new measure of SES.

## Defining SES

The panel developed the following consensus definition of SES:

> *SES can be defined broadly as one's access to financial, social, cultural, and human capital resources. Traditionally a student's SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.*

## Components and Correlates of SES

The panel concluded that the components of a core student SES measure were the "big 3" variables (family income, parental educational attainment, and parental occupational status),

but also suggested that home neighborhood and school SES could be used to construct an expanded measure of SES. Identifying such variables and including them in an expanded SES composite could help improve the explanatory power of SES in accounting for NAEP scores. In addition, some psychological process variables (e.g., coping mechanisms, emotional control, or perceptions of the environment) and some subjective measures (i.e., how one views one's SES), might be understood as useful contextual and potentially explanatory variables that could help interpret student NAEP scores. Although psychological process and subjective factors were not included as components of a core or expanded SES as developed by the panel, it is important that research be conducted to evaluate the effects of these factors on achievement.

## Approaches to Measuring SES Components

In addition to current measures of family income, additional variables, such as housing tenure (rent or own), number of moves in the past year, presence of a household member needing healthcare assistance, and others, could be studied for potential use as indirect measures of family income. Parental educational attainment is currently measured through the NAEP questionnaire, but only for 8$^{th}$- and 12$^{th}$-graders, and parental occupational status – one of the big three variables – is not collected in the Student Questionnaire, nor is it available through school records. Cognitive laboratory studies should be conducted on various question types for collecting student reports on parental occupation.

There are currently no direct measures of neighborhood components of a possible expanded SES measure, although NAEP student questionnaire items and information from school records could be aggregated to serve as neighborhood measures. American Community Survey data could be used to provide much of the information not available through NAEP questionnaires and school records. The upcoming Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a), which tests 4$^{th}$-graders in 2014, represents an ideal opportunity to inform SES measurement.

## SES Composite

There are reporting and interpretation advantages and disadvantages for treating SES as a single measured variable, as several single measured variables, or as a composite of several measured variables. The advantages of a composite variable over the use of single variables outweigh the disadvantages. There are a variety of schemes by which SES components could

be combined into a composite measure. A challenge in developing an SES composite is determining whether weights should vary depending on factors such as location or grade level. A review of the existing literature and data quality should be conducted before proposing a recommendation on a component weighting scheme.

## Implications

A new SES measure will affect NAEP reporting, including whether and how to characterize SES levels, whether a bridge study must be conducted to link new and old measures of SES, and how a new SES measure will affect NAEP's conditioning model. The research, framework, and findings associated with the development of a new SES measure could benefit other programs that measure SES, both within and outside NCES. For example, states are continually seeking better measures of SES. In addition, the development of a new SES measure is likely to incur both anticipated and unanticipated side effects, including the requirement to coordinate with other federal programs within and outside NCES, and consequences such as attention given to equity and educational resource distribution.

## Key Recommendations

The panel made four key recommendations to improve measurement and reporting of SES:

1. Family income and other indicators of home possessions and resources, parental educational attainment, and parental occupational status should be considered components of a core SES measure, and should be the subject of immediate focus for NAEP reporting.

2. Neighborhood and school SES could be used to construct an expanded SES measure, and measures of these variables could contribute to an expanded SES.

3. Composite measures have many advantages, such as being a single summary useful for reporting, greater reliability, and representing the full range of SES factors. In addition, treating SES as a composite measure does not preclude reporting on relationships between individual SES components and achievement. Therefore, attempts should be made to develop an SES composite measure.

4. The validity of NSLP eligibility has been decreasing due to jurisdiction-wide eligibility and other factors, and that trend is likely to continue. Furthermore, there is concern over the quality

of student reports, particularly regarding parental educational attainment (for 4th-graders) and occupational status (for all grades). Due to these data quality issues, along with burden considerations, attempts should be made to explore the possibility of linking to Census data on SES components.

# 1. SETTING THE STAGE

The National Assessment of Educational Progress (NAEP) is a congressionally authorized project of the National Center for Education Statistics (NCES) within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of NCES is responsible for carrying out the NAEP project, while the National Assessment Governing Board (Governing Board) oversees and sets policy for NAEP. NAEP measures student progress over time in a variety of subject areas, including reading, writing, mathematics, science, and U.S. history. NAEP does not report individual student scores; rather, the assessment is designed to produce public-domain data about student achievement at the group level. Because NAEP results are meant to inform educators, policymakers, and the general public about the performance of students at the 4th, 8th, and 12th grade levels, reports include overall results as well as scores for student subgroups that are of interest to the target audiences, such as gender, race/ethnicity, and socioeconomic status (SES).

In response to a call by the Governing Board (NAGB, 2003) to improve the measurement and reporting of SES and its relationship to academic achievement in the context of NAEP, NCES convened a panel of experts in the fields of economics, education, statistics, human development, and sociology with substantive expertise in the effects of poverty and disadvantage on student achievement as well as methodological expertise in the measurement of socioeconomic standing. The panel was asked to provide recommendations concerning SES as a construct with the understanding that those recommendations might ultimately lead to a new SES measure that could be used in programs such as NAEP. The guidance was to focus on the theoretical aspects of SES measurement, not on operational aspects. Specifically, the panel was tasked with considering issues surrounding SES, including the creation of a composite measure of SES, how a new SES variable could be used in a reporting context, and how its derivation could be explained to both technical and general audiences. The panel met three times between 2010 and 2012.

This report reflects the discussions and recommendations of the panel and provides the context and background for those discussions. The report was prepared with key NCES stakeholders in mind, including the general public and education policymakers at both the state and national levels. The panel discussed the reporting of SES in NAEP and other large-scale assessments, such as the Program for International Student Assessment (PISA) and the Trends in

International Mathematics and Science Study (TIMSS), to learn from those assessments and also to seek to inform them.

# 2. BACKGROUND

There is a long history of SES being reported to correlate with educational achievement (Cuff, 1934; Holley, 1916; Lynd & Lynd, 1929). The Equality of Educational Opportunity Commission Report (Coleman et al., 1966) played a major role in bringing this finding to prominence in policy circles. Since then, measures of SES have been routinely included in educational research studies as background variables. Researchers and policy makers are interested in SES as a contextual variable to study educational equity and fairness issues, as a covariate with achievement to examine the effects of other variables such as class size or school governance policies, and as a matching variable to ensure the equivalence of treatment and control groups in educational intervention studies. NAEP treats SES as one of five background reporting variables (see law excerpt, below) and scores are reported separately for different SES subgroups. NAEP is mandated to report scores by SES by the No Child Left Behind Act of 2001 (P.L. 107-110, 2002), which acknowledges the importance of SES in educational achievement:

> *The Commissioner, in carrying out the measurement and reporting described in paragraph (1), shall— ''(G) include information on special groups, including, whenever feasible, information collected, cross tabulated, compared, and reported by race, ethnicity, **socioeconomic status**, gender, disability and limited English proficiency* (Sec. 411. National Assessment of Educational Progress, Paragraph (b) Purpose; State Assessments; Subparagraph (2) Measurement and Reporting). (115 STAT. 1898)

However, the mandate does not provide specific guidelines on how SES is to be measured, nor even on how it is defined. Current NAEP practice is to measure SES through a set of proxy variables, most notably eligibility for the National School Lunch Program (NSLP; 2008), but also through school Title 1 status, parental educational attainment, and reading materials in the home. For reporting purposes, all of these are treated as individual variables, rather than as a composite SES variable.

It is instructive to review how socioeconomic status is treated in NAEP score reporting. Typically no mention is made of SES per se, but NAEP scores are reported by different variables that might be interpreted as SES measures. In the recent 2009 NAEP Science report (NCES, 2011), for example, for 4$^{th}$- and 8$^{th}$-grade students, NAEP scores were reported by eligibility for NSLP in three categories (*not eligible*, *eligible for a reduced-price lunch*, and

*eligible for a free lunch*). NSLP eligibility was reported to be "an indicator of low income" (p. 60). The *Technical Notes* section of the report states that scores were not reported by NSLP eligibility for 12th-grade students "[b]ecause students' eligibility for free or reduced-price school lunch may be underreported at grade 12." (p. 60) (See discussion in the *Measuring SES* section, below.) These data are obtainable from the NAEP Data Explorer, however. For 8th-grade and 12th-grade students, NAEP scores were reported by parental educational attainment, which is widely regarded as a central component of SES.

A broad and widely accepted definition of SES in the scientific literature emphasizes its role in reflecting access to resources. Furthermore, students' SES is traditionally defined as a combination of family income, parental occupational status, and parental educational attainment. Although the proxy variables currently used in NAEP reflect these factors to some extent, questions have been raised about the quality of the data, the narrowness of the measure, and the lack of a composite SES measure.

Consequently, there have been calls to explore alternative SES measures. Among the suggestions have been to create a composite measure rather than relying on single proxy measures (Barton, 2003), and to use data linked from other sources, such as the U.S. Census Bureau's American Community Survey, which provides data on income, parental occupation, and parental educational attainment (Hauser & Andrew, 2007). The problems identified with current NAEP practice in measuring SES, along with conceptual and empirical developments in understanding SES, suggest that the time is right to consider alternatives in developing a new SES measure for NAEP.

## Measuring SES

The history of SES measurement and the identification of possible explanatory correlates show that SES is defined as a broad construct, ideally measured with several diverse indicators. However, there are some advantages to using NSLP eligibility as an operational SES measure for NAEP. First, NSLP eligibility is available through school records for almost every student in the U.S., making data collection inexpensive and minimizing problems with missing data. In addition, NSLP status is a three-level categorical variable, which is convenient for reporting purposes and easily understood by a variety of audiences. Finally, NSLP eligibility status is also tied to federal definitions of poverty, which means that maintenance or updating is handled automatically through updating of federal poverty guidelines.

On the other hand, there are problems with using NSLP eligibility as the main measure of SES in NAEP reporting. These problems can be summarized as follows:

1. *NSLP eligibility measures only one SES component, family income* (adjusted for household composition). NSLP eligibility does not reflect parental educational attainment or occupational status.
2. *Due to the process of eligibility certification, NSLP eligibility may not be the most reliable measure of family income* (Harwell & LeBeau, 2010). Approximately 20 percent of students either are not eligible but are deemed eligible or are eligible but are not recognized as such (Food and Nutrition Service, 1990; Harwell & LeBeau, 2010; Hauser, 1994). The problem of eligible students failing to apply (whether due to social stigma or some other cause) increases with grade level, and is particularly prevalent for 12$^{th}$-graders (Office of Research, Nutrition, and Analysis, 1994). Failure to apply when eligible is also thought to correlate with immigration status and to be more prevalent among students who speak English as a second language.
3. *Because there are only three levels of NSLP eligibility, there are large within-category SES differences, particularly in the non-eligible category.* Furthermore, the categories contain uneven shares of the distribution; there is approximately an 8:1 ratio of students in the free vs. reduced-price lunch categories.
4. *School-level and jurisdiction-level eligibility threatens the validity of NSLP eligibility as a measure of an individual student's family income*. All students in a school with greater than 80 percent eligibility are categorized as NSLP eligible, regardless of their family income. Likewise, all students in some jurisdictions, such as Puerto Rico, and many of the urban districts are declared eligible regardless of family income levels.[1]

The remainder of this paper is organized into seven additional chapters. *Objectives* reviews project goals, which are to articulate a definition of SES; identify SES components; address

[1] Regarding this last point, the most significant problem with the NSLP eligibility measure for the future is the introduction of Community Eligibility (Provision 4) through The Healthy, Hunger-Free Kids Act of 2010 (PL 111-296). Community Eligibility means that schools will no longer be required to keep individual student eligibility information once they have determined a baseline percentage of eligible students, which may result in missing or inaccurate individual student eligibility information. This change in eligibility certification is expected to be phased in, but would nevertheless affect the validity of the NSLP eligibility measure.

data collection issues, which should lead to a new SES composite; and consider implications of a new SES measure. *Defining SES* emphasizes a broad definition of SES as a student's access to resources and reviews its expected relationship to achievement. *Components and Correlates* presents the idea of SES as a composite of "the big 3" variables—family income, parental educational attainment, and parental occupational status—and additional variables, most notably neighborhood and school SES. Also included here is a discussion of variables that could be considered as either components or correlates of SES, including subjective (perceived) SES, cultural capital, and other factors. The chapter also reviews variables that correlate with SES and variables (moderators) that interact with SES in its relationship with achievement. *Approaches to Measuring SES Components* reviews how the proposed components of SES can be measured. *SES Composite* reviews various ways in which SES components can be weighted and combined, and discusses issues with missing data. *Implications* focuses on the consequences of a new measure of SES for reporting, including anticipated and unanticipated side effects, and discusses possible uses of the new SES measure by other units, departments, and agencies. Finally, *Discussion* provides a general summary of the paper and concludes with key recommendations.

## 3. OBJECTIVES

The panel was to provide recommendations for a new SES measure for NAEP that would continue to meet the requirements of reporting SES and also improve the measurement and reporting of SES through the collection of higher quality data.

The primary purpose of proposing a new measure of SES was to meet the requirements of the No Child Left Behind Act of 2001 (P.L. 107-110, 2002) in the best feasible way. The law mandates the reporting of scores by SES in acknowledgement of the importance of SES in educational achievement. The law was not specific about how SES was to be measured. The primary objective was to have a panel provide expert guidance and interpretation on how the law's mandates could be realized.

Specifically, the panel was charged with the following study objectives:

## Provide a Definition of SES

Although it is possible to develop and use a measure of socioeconomic status without a clear definition of what it is—for example, basing it on measures that have been used or are currently used in different projects—there are many advantages to communicating a clear definition of SES.

## Identify Components of SES

Historically, SES has been defined as some combination of family income, parental educational attainment, and parental occupational status. Other variables have also been considered as part of SES, including various school factors, community or neighborhood factors, and subjective measures of socioeconomic status, such as where individuals see themselves on a status ladder. An objective for this study was to identify which of the various components should be included as part of SES for NAEP reporting.

## Review Data Collection and Measurement Approaches

Some SES measures, such as eligibility for the National School Lunch Program, have been obtained from school records. Others, such as parental educational attainment, have been obtained from the student questionnaire. A methodology that has been discussed, and experimented with, involves obtaining geographically aggregated Census data (aggregated over ZIP code tabulation areas, Census blocks, or Census tracts) to impute student family data, such as family income, household status (e.g., single vs. dual head of household), and parental occupations. Census data can be obtained either from the United States Census 2000 long form (to analyze previously-collected data for research purposes only), or from the ongoing American Community Survey 5-year estimates (which could be analyzed for both research purposes and operational use). An objective of this study was to review data collection and measurement pertinent to these various approaches.

## Create an SES Composite (or Justify Use of Multiple Single Variables)

An initial panel objective was to consider alternatives and make recommendations on how an SES composite could be formed. However, the scope was widened to include the possibility of using multiple single variables to measure SES rather than a composite. Thus an objective for

the study was to consider the pros and cons of an SES composite vs. multiple single-variable measures of SES. The charge was also to consider various issues in forming a composite, such as how to weight the components of a composite, and whether to vary or keep weights constant across grades, whether to adjust weights (such as income) for locality, how to deal with the issue of missing data, and so forth.

### Consider Implications of a New Measure of SES

A new measure of SES would have implications for the reporting of NAEP scores. For example, a new measure of SES might show greater achievement differences between low and high SES groups, compared to free lunch vs. non-subsidized lunch groups. A sudden change in how SES was defined might therefore disrupt trends in the relationship between SES and NAEP achievement scores, which would create significant challenges to interpreting SES estimates over time. Eligibility for a free or reduced-price lunch is a variable with three categories, which is convenient for reporting. A new measure of SES could be a continuous variable, in which case a decision would have to be made about whether to transform it into a categorical variable, or treat it in some other fashion. An objective for this study was to consider these and other implications of a new measure of SES.

## 4. DEFINING SES

SES is measured by different variables in different studies (e.g., Sirin, 2005), which makes it difficult to appreciate exactly what it is, or what researchers or policy makers mean by SES. However, studies on the relationship between SES and educational achievement cover more than nine decades of research (e.g., Bryant, Glazer, Hansen, & Kursch, 1974; Coleman et al., 1966; Cowan & Sellman, 2008; Harwell & LeBeau, 2010; Holley, 1916; Lynd & Lynd, 1929; White, 1982). It is useful to consider this history in developing a definition of SES.

SES emerged as a concept because of observations that students of parents with low educational attainment, working in low-status jobs, or with low income performed more poorly in school and on tests that reflected school achievement. One of the earliest SES conceptualizations was Taussig's (1920) classification, which was based solely on father's occupational status, classified into seven categories. In a later study by Cuff (1934), the Sims (1927) Score Card was employed as a measure of SES. The Sims Score Card contained a survey with 23 questions about home possessions (books, telephones), rooms in the home,

extracurricular and cultural activities, parents' educational attainment, and father's occupation. The Chapin (1933) scale was a rating scale based on the idea that socioeconomic status reflected cultural and material possessions (furniture, accessories), income, and participation in community activities, and which were reflected in and therefore could be measured by home possessions in and the condition of one's living room.

The development of instruments such as the Sims Score Card and the Chapin scale led to increased measurement sophistication of SES. An example was Sewell's (1940) classic study of the measurement of SES in farm families, one of the earliest sociological applications of factor analysis. Ganzeboom, De Graaf, and Treiman (1992) developed a model-based approach in which they proposed that occupational status mediated the relationship between education and income. They then computed occupational status accordingly. The Ganzeboom et al (1992) measure is currently used in The Program for International Student Assessment (PISA) to measure occupational status. Hauser and Warren (1997) similarly took into account educational levels in measuring occupational status.

In the present day, large-scale international assessments routinely include measures of SES. PISA, for example, includes items administered to fifteen-year-old students that form an SES composite called the PISA index of economic, social, and cultural status (ESCS) (OECD, 2010a; see pp. 131). The ESCS is a weighted composite (based on a principal component analysis) of three variables:

- occupational status of the parent with the higher occupational status (based on the Ganzeboom et al. [1992] model, described above),
- educational attainment of the parent with the higher educational attainment, and
- an index of home possessions.

The index of home possessions is itself a composite of three variables (derived from 16 survey questions related to home possessions) and a categorical measure of total number of books in the home:

- wealth (room of their own, Internet link, dishwasher, DVD player, and 3 country-specific measures),
- cultural possessions (classic literature, books of poetry, classic art), and

- home educational resources (desk and quiet place to study, a computer available for school work, educational software, books to help with school work, technical reference books, and a dictionary)
- number of books in the home (four categories: 0-10; 11-100; 101-500; over 500).

Numerous studies over the years have attempted to provide an explanation for why SES correlates with academic outcomes. The Wisconsin Model developed by William H. Sewell and colleagues (Sewell, Haller, & Portes, 1969), based on the Blau and Duncan model (1967), was one of the first attempts to account for educational and occupational attainment by proposing a recursive model including personal aspirations, the influence of peers, educational achievement, parents' SES, and cognitive ability. Along these lines, SES is related to the kind of school and the kind of classroom a student attends (Reynolds & Walberg, 1992), with schools differing characteristically in the kind of instruction offered, materials provided, teacher experience, and access to teachers (Wenglinsky, 1998), as well as the kind of relationship that exists between school staff members and parents (Watkins, 1997).

It may not be family income or poverty per se that drives the relationship between SES and achievement, and life success (Mayer 1997). Spaeth (1976) suggested that SES might indicate the complexity of a child's cognitive environment and that exposure to cognitively challenging home environments prepares students better for the challenges of school. Levin & Belfield (2002) suggested several "pathways" or home environment variables through which SES might affect student achievement. These include the learning environment, language and literacy, parent-child interactions, and daily routine. Low SES children are less likely to have a "school-like" home and follow a daily routine; they have weaker language interaction with parents, weaker literacy engagement, and more conflicting interactions. Walpole (2003) noted that low SES students also tend to have less access to cultural capital (specialized or insider knowledge not taught in schools) and social capital (contacts in networks that can lead to personal or professional gains; Coleman, 1988), which have been argued to be key components of a student's educational success. Recent research in genetics suggests that SES may limit opportunities for children to pursue and benefit from educational experiences congruent with genetically-influenced intellectual interests (Tucker-Drob & Harden, 2012). There also is research linking family socioeconomic resources, including a consideration of family size and structure, to student test scores (Duncan & Magnuson, 2005).

Together, these studies suggest that SES may broadly be seen as a general variable that indexes resources available to the student, including economic, social, and cultural resources. Furthermore, the "big 3" variables discussed earlier can be thought to capture different aspects of resources available to students.

Recently, the American Psychological Association (2007a) created an Office on Socioeconomic Status and issued a report from a specially commissioned American Psychological Association Task Force on Socioeconomic Status. The commission provided a framework for defining and developing SES measures. They characterized three models for understanding SES and social class-related inequalities, across three domains: education, health, and human welfare. One model, reflecting most of the SES literature as reviewed here, was what they referred to as the traditional materialist model. Another model emphasized social gradients and individuals' positions relative to others', which motivates the use of subjective SES measures. A third model focused on social capital, but seemed not to have resulted in specific SES measurement approaches.

Several studies have investigated what kinds of variables have been used in studies of educational achievement to measure SES. White (1982) conducted a meta-analysis on studies before 1980, and Sirin (2005) conducted a meta-analysis of studies published between 1990 and 2000. Both studies indicated a medium to strong relationship between SES and achievement with some measures showing stronger relationships than others. Sirin (2005) found that measures could be placed into the SES categories of parental educational attainment (30 studies), parental occupational status (15 studies), family income (14 studies), free or reduced-price lunch (10 studies), neighborhood (6 studies), and home resources (4 studies). Parental educational attainment was also the most commonly used measure in the studies White (1982) reviewed, and parental occupational status and family income were frequently used as SES measures.

Based on both the history of SES and the measures used to assess SES in studies of educational achievement, the primary measurement of SES over the years has been the "big 3" variables: (a) family income, (b) occupational status of heads of household, and (c) educational attainment of heads of household, consistent with what Hauser (1994) pointed out. However, school and neighborhood variables have also been included in SES measurement for some time (Hauser, 1969). For example, Fertig (2003) examined student peer group achievement heterogeneity on student achievement using PISA data. Van Ewijk and Sleegers (2010)

conducted a meta-regression analysis of the effects of peer socioeconomic status on student achievement, and showed effects at both the individual and class levels.

### Panel Recommendation: A Definition of SES

A consensus definition of SES is as follows:

> *SES can be defined broadly as one's access to financial, social, cultural, and human capital resources. Traditionally a student's SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.*

## 5. COMPONENTS AND CORRELATES OF SES

In thinking about how SES should be defined for NAEP, it is useful to draw a distinction between components and correlates of SES. An SES component is a variable that should be included as part of SES—that is, as part of the measurement of SES. An SES correlate is simply a variable that correlates with SES, but should not be considered part of SES. It is a high priority in future NAEP data collections to include SES components, while collecting data on SES correlates is a lower priority.

This chapter focuses on determining the components and correlates of SES without regard to the practical aspects of measurement. The next chapter focuses on possibilities for measuring the components of SES.

### The "Big 3"

Given the history of SES and how it has been defined over the years, and given the common ways it has been measured in research (see chapter 4), the "big 3"—family income, parental educational attainment, and parental occupational status—should be considered components of SES. Home possessions could be used to measure family income, but there are several drawbacks to using possessions as such a measure: possessions are not typically measured in surveys, do not necessarily represent an accurate picture of family income, and vary over the life cycle in uneven ways with respect to income. Still, possession measures are widely used as

SES measures in student educational surveys because they are less intrusive than income measures.

There are additional factors that could be considered components of SES. Specifically, insofar as SES is defined as access to financial, social, and human capital resources, particularly as these factors relate to schooling, they could also be considered components of SES.

## Neighborhood SES

The argument for including neighborhood SES information in an expanded measure of student SES is that not all financial, social, and human capital resources available to the individual student come from the family. Some resources come from the neighborhood or community in which the student resides. The resources shape the home environment, broadly conceived, and have been shown to be associated with school achievement.

Traditional indicators of neighborhood SES include the percentages of families below the poverty line, unemployed adults in a neighborhood, and the adults in the neighborhood with a low education level (e.g., percentage without a high school credential). Additional indicators could include the percentage of single parent homes and the percentage of homes where English is not spoken well. In addition, there are social and physical resources associated with neighborhoods, both negative (e.g., the presence of abandoned buildings and roads and walkways in poor condition), and positive (e.g., the availability of parks, recreational areas, and public libraries), that could also be considered part of a neighborhood SES construct. There also are family and household characteristics of a neighborhood, such as aggregated family possessions (e.g., number of rooms in residents' homes, books in their homes, and backyard facilities) which may indicate social and cultural status of a neighborhood.

Empirically, it is not necessarily the case that neighborhood SES data adds information not already available from individual level data. For example, some previous analyses of 8$^{th}$-graders found that neighborhood data from Census added very little to the relationship between student and parent SES reports (Rivas & Hauser, 2008). However, there is a conceptual distinction between individual family and neighborhood measures of SES, and neighborhood SES should be considered an additional SES component.

There are advantages in including neighborhood SES as part of an individual-level measure of student SES. For one, neighborhood SES can be critical to understanding how student

psychological processes (see description in the *Psychological Process Variables* section, below) interact with the context "in real time," and these processes may be influenced by, for example, the creation or expansion of libraries and parks or the diminishment of features such as abandoned buildings and unsafe walkways. Additionally, Census variables that might not be linked at the individual level could be used at the neighborhood level.

Defining what is meant by "neighborhood" (e.g., ZIP code, tract, block group) is difficult, however, and should be considered an operational decision to be decided later. There is also an important distinction to be made between school neighborhood and living neighborhood, as the neighborhood where students live may not have the same characteristics as the neighborhood of the school the student attends, even if they are located in the same ZIP code. For these reasons and others, the prospects for creating components of neighborhood SES that are specific enough to increase the prediction of individual-level NAEP tests scores are uncertain. The odds are sufficiently high that additional work is warranted.

## School SES

Many students attend school in the neighborhood in which they live, but some students attend schools outside of their neighborhood due to school choice initiatives and other factors. School choice is a major movement that may lead to more disconnect between neighborhood SES and the SES composition of the schools that students attend. Therefore, both school and neighborhood SES information could be included as distinct components in an expanded measure of SES. School SES can be defined as the aggregate of the individual students' SES. Currently, school SES is commonly measured by Title 1 status and percentage of students eligible for NSLP.

There are other characteristics of schools (e.g., school safety, physical surroundings) that are relevant for student achievement. However, they should not be considered direct components of an expanded SES measure.

## Psychological Process Variables

Research has shown that students at different SES levels have varied levels of exposure to experience with events such as frequent moving or having contact with law enforcement in different ways. Research has also suggested that that low SES is associated with significant

risk exposure and low protection factors, and these are likely to influence achievement. Student perceptions of parental involvement and parental monitoring may affect NAEP outcomes. In addition, certain neighborhoods may lead students to adopt coping mechanisms that may not function well in a school environment, or inhibit the development of noncognitive skills such as emotional control. While these are important variables for understanding how students make sense of their environments, psychological process variables, such as coping mechanisms, perceptions, and emotional control, are variables best understood as consequences or correlates of SES rather than as necessary components of SES.

## Subjective SES

Research on subjective SES suggests that how one thinks of one's status subjectively can be as important as objective SES measures in relating to outcomes. For example, subjective SES has been shown to predict physical and mental health outcomes after controlling for objective SES (Demakakos, Nazroo, Breeze, & Marmot, 2008). That is, believing you are high status might compensate for lower objective status.

Measurement of subjective SES has relied extensively on the SES ladder technique (e.g., Demakakos, et al., 2008), in which respondents are shown a picture of a ten-rung ladder designed to reflect SES and asked to indicate where they think they (or their family) would stand on the ladder. Other methods for measuring subjective SES include a simple "get along" measure, asking whether the student or student's family has enough money to get along, which has the advantage of being a relative measure that is adjusted over time. Gallup has administered a "get along" question for several decades in various adult surveys as a means to obtain a subjective estimate of poverty level (e.g., Citro & Michael, 1995).

A number of measurement challenges could hinder development of a valid measure of the subjective SES of students, particularly for 4$^{th}$-graders. For example, the meaning of subjective SES may vary based on geographic location. A subjective SES measure also might not capture distinctions between high earners with modest educational backgrounds and highly educated middle-level earners. A subjective SES measure could be susceptible to reference group effects (Crede, Bashshur, & Niehorster, 2010), that is, differences in how students see themselves due to the reference group to whom they are comparing themselves. For example, students from homogeneous neighborhoods might interpret objectively small neighbor-to-neighbor differences as large because their reference group is the neighborhood in which they live.

(Effects of school heterogeneity on self evaluations has been studied in international surveys [Lafontaine & Monseur, 2007], but it seems that comparable studies have not been conducted with neighborhood heterogeneity.) However, if a valid measure of subjective SES could be developed, it might prove useful as a way to capture whether the child perceives that they have the resources to succeed. This would not, however, be consistent with a measure of SES that indexes access to actual resources of various types.

### Panel Recommendations: Identifying Components and Correlates of SES

1. The primary components of SES are the "big 3" variables—family income, parental educational attainment, and parental occupational status.

2. Additional components of an expanded SES measure could include neighborhood and school SES.

3. Psychological variables and some subjective measures of SES may be useful contextual and potentially explanatory variables that could help interpret NAEP scores.

## 6. APPROACHES TO MEASURING SES COMPONENTS

The purpose of this section is to review ways of measuring the SES components identified in the previous section. The focus is on measuring the "big 3" and neighborhood and school SES. This section, reviews existing measures of each of the SES components,, including school records, the NAEP student background questionnaire, the NAEP 2012 pilot student background questionnaire, and the American Community Survey (ACS). The ACS measures are included because they provide alternative socioeconomic measures, and they may be useful in characterizing geographic areas.

### Family income

As reviewed previously, NSLP eligibility, obtained through school records, is a measure of income (adjusted for family composition), and is featured prominently in NAEP reporting. The NAEP student background questionnaire also includes items yielding data that could be understood as reflecting family income:

- Books in the home
- Encyclopedia in the home
- Magazines in the home
- Computer in the home

The 2012 NAEP pilot student background questionnaire includes additional items that may yield data pertaining to family income:

- Home possessions (internet access, clothes dryer, dishwasher, more than one bathroom, your own bedroom)

The ACS includes items pertaining to income:

- Income (9 questions, total) (for each member of the household)
- Home possessions (8 items)
- Rooms in the home (2 items)

### *Other indirect measures of family income*

Several other variables could be considered indirect measures of family income, but are not currently measured in NAEP background questionnaires. These include:

- Housing tenure (rent or own)
- Number of moves in the past year
- Presence of household member needing healthcare assistance
- Immigration status (recency of immigration)
- School resources
- Student's perceived level of support (home, school, neighborhood)

*Housing tenure* (owning as opposed to renting one's place of residence) is an indicator of income and wealth and of residential stability. In addition, there is considerable evidence regarding its relationship to age-grade retardation and high school dropout (Frederick & Hauser, 2008; Hauser, Frederick, & Andrew, 2007; Hauser, Pager, & Simmons, 2004; Hauser, Simmons, & Pager, 2004), *Number of moves in the past year* serves as an indirect measure of housing tenure, and also as a measure of instability and high risk status. *Presence of household*

*member needing healthcare assistance* can drain family financial resources. *Immigration status* is an indirect indicator of English language proficiency, social capital, and wealth. *School resources* is not typically thought of as measuring family income, but could be considered indirect measures of family resources, and school resources reflect resources available to the student. *Student's perceived level of support* at home, at school, and in the neighborhood also reflects the availability of resources to the student. Many of these measures could be collected through the student (and teacher and school) questionnaires, and some might be obtainable through school and Census (ACS) records.

The 2012 NAEP pilot student background questionnaire includes an item that may yield data indicating students' English language proficiency, social capital, and wealth:

- How long have you lived in the United States?

### *Household composition*

Household composition—number of parents and siblings—should be included when measuring family income. Partly this is due to the fact that family income has to be distributed across the members of the household, and so financial resources available to the individual student will be a function of both family income and the number of individuals that income is spread across. One or two parents in the household will have an opposite effect, as two parents may provide more social and emotional support than one. NSLP eligibility itself implicitly includes household composition, as its Income Eligibility Guidelines (based on the federal income poverty guidelines) are stated by household size. There are no additional questions on household composition in the NAEP student questionnaire.

The 2012 NAEP pilot student background questionnaire includes the following household structure questions:

- Size of household (total, number of adults)
- Household structure (single- vs. dual-parent, and other relatives)

There may be some ACS variables that could be added to this list, such as number of workers in the household and number of earners in the family.

## Parental educational attainment

The NAEP student questionnaire includes two parental educational attainment questions:

- Mother's educational attainment ($8^{th}$ and $12^{th}$ grade only)
- Father's educational attainment ($8^{th}$ and $12^{th}$ grade only)

The ACS includes educational attainment questions for each member of the household:

- Whether currently attending school (level and type)
- Educational attainment
- Major (for bachelor's degree holders)

## Parental Occupational Status and Employment Status

The NAEP student questionnaire does not include any questions about parental occupation and employment status, nor is such information available from school records. Therefore this SES component has been absent from NAEP reporting.

The 2012 NAEP pilot student background questionnaire includes the following question about parental and household resident employment status:

- How many adults living in your home have a job?

The ACS includes the following employment status and occupation questions for each member of the household:

- Employment status (working for pay or not, part-time vs. full-time, etc.; 22 questions total)
- Occupation (6 questions total)

The National Education Longitudinal Study (NELS:88) (Ingels, 1990) asked 8th-graders for their mother's and father's occupation in an open-ended question. But it also included a closed (multiple-choice) question: "what kind of work do you expect to be doing when you are 30 years old?" The response choices included categories such as craftsperson or operator, farmer or farm manager, professional business or managerial, and so on.

Cognitive laboratory studies must be conducted on various question types for collecting student reports on parental occupation. If questions could be developed to provide reliable information on parental occupation, then it would be useful to use these data in creating a better measure of SES, even if such information does not reach the same reliability and validity level as other questionnaire responses.

There are upcoming opportunities to collect data on new SES component measures. For example, the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a) wave will test 4th-graders in 2014. That study will collect data from both parents and students. Certain questions concerning parental educational attainment, occupation, home possessions, or any other SES-related questions could be inserted into the parent and the student ECLS-K:2011 questionnaires for the 2014 study. A comparison could be made between the responses to evaluate the validity of 4th-grade student data.

## Neighborhood SES

There are currently no direct measures of neighborhood SES from either the NAEP student questionnaire or school records. However, the 2012 NAEP pilot student background questionnaire includes the self-reported ZIP code item, from which neighborhood information could be obtained.

There are a variety of ways to measure neighborhood SES. In addition to measures such as percentage below poverty, unemployed, and with low educational attainment, other variables include the availability of parks and libraries in the neighborhood, the absence of abandoned buildings, proportion of single-parent households, and the proportion of households in which English is not spoken well.

Some neighborhood SES information could be obtained through the extended school questionnaire. Some items from the student questionnaire and from school records could also be aggregated so as to serve as neighborhood SES measures.

Additionally, the ACS includes a number of items that could be treated as measures of neighborhood SES, including unemployment, education, and income levels, household overcrowding, poverty, home ownership, and perhaps some indicators of vulnerability. ACS data would be suitable for neighborhood measures of SES, though there are some challenges in using ACS data, such as heterogeneity in neighborhoods at the tract level and above. Information obtained from the ACS 5-year estimates (with the least sampling error and provided at the smallest geographical unit) would not reflect rapid changes in a neighborhood, but neighborhoods tend to be very stable, and this is not likely to be a problem. Another challenge is that the size and boundaries of a neighborhood for the purposes of creating a neighborhood SES variable are undefined. Perhaps Census blocks (the smallest geographic area for which data are collected and tabulated), or block groups (optimal size of 1,500 people), or possibly even Census tracts (optimal size of 4,000 people), or ZIP code tabulation areas could serve as neighborhood boundaries for this purpose. Using Census blocks or tracts would require obtaining more precise location information for student households than student ZIP codes and would require special arrangements with the Census Bureau.

## School SES

As noted above, school SES is most commonly measured by Title 1 status and percentage of students who are eligible for NSLP. However, additional school SES variables could be formed as aggregations of student-level variables obtained through school records or student questionnaires, such as percentage of English language learners, average level of parental educational attainment, average home possessions, and so forth. In addition, characteristics of the schools and school climate could be obtained through teacher and school questionnaires, and these *could* be treated as part of a school SES variable. Because most students attend neighborhood schools, it is likely that school and neighborhood SES measures would correlate highly, but it also may be that there is unique information in the school and neighborhood SES measures.

## Panel Recommendation: Review Data Collection and Measurement Approaches

1. Additional variables could be studied as indirect measures associated with family income, including housing tenure (rent or own), number of moves in the past year, presence of a household member needing healthcare assistance, immigration status (and recency of

immigration), school resources, and student's perceived level of support (home, school, neighborhood).

2. Family/household composition and structure—size of household and whether single or dual-parent—are also important variables to consider both because single parenthood is generally considered a disadvantage and because household resources are diluted in large households.

3. Parental education is currently measured through the NAEP questionnaire, but only for 8th- and 12th-graders. The ACS includes parental educational attainment questions, which could be used to obtain this measure for 4th-graders. However, a strategy would have to be developed to determine how to link ACS data to NAEP.

4. Cognitive laboratory studies should be conducted on various question types for collecting student reports on parental occupation. If a proper format could be identified for collecting reliable information on parental occupation, then it might be useful to include such questions in future questionnaires even if the reliability and validity level were not as high as is expected for other questionnaire responses.

5. There are currently no direct measures of neighborhood SES from either the NAEP student questionnaire or school records. However, if student ZIP code could be obtained, it may be possible for NAEP data to be linked to ACS data in order to compute neighborhood SES measures for the students' residential neighborhoods (such as unemployment, education, and income levels, household overcrowding, poverty, home ownership, and perhaps some indicators of vulnerability). Research is needed to determine whether ZIP-code defined "neighborhoods" will yield useful additional components for an expanded SES measure. Data from student questionnaires and information from school records also could be aggregated to serve as neighborhood SES measures, although only for neighborhood schools.

6. School SES can be measured using Title 1 status and percentage of students eligible for NSLP. Additional school SES variables could be computed as aggregations of student-level variables, obtained through school records or through student questionnaires, such as percentage of English language learners, average level of parental educational attainment, and average home possessions. School characteristic and climate variables could be obtained through teacher and school questionnaires and these *could* be part of a school SES variable.

7. An ideal opportunity to inform SES measurement is available through participation in the upcoming Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a), which tests 4th-graders in 2014. The study will collect data from both parents and students, enabling a comparison of parent and 4th-grade student reports to test the validity of the student reports.

## 7. SES COMPOSITE

In the literature and in official reports SES is sometimes reported as a single variable, such as parental educational attainment level or NSLP eligibility, and sometimes as a composite variable with several component variables summed together. Initially, the panel was to consider alternatives and make recommendations on how an SES composite could be formed. However, during panel deliberations the scope was widened to include the possibility of the use of a single variable (or multiple single variables) rather than a composite to measure SES. Thus an objective for the panel was to consider the pros and cons of an SES composite vs. single-variable measure of SES. The charge was also to consider various issues in how to form a composite, such as how to weight the components of a composite, and whether to vary or keep weights constant across grades, whether to adjust weights (such as income) for locality, whether to change weights every year, or periodically, and so forth.

An advantage of treating SES as a single variable is that the meaning of a single variable is typically clear and easy to communicate. For example, audiences know what it means to have a parent who completed college, or to be eligible for a free lunch through the NSLP. A related advantage is that the meaning of different SES levels when defined as a single variable can be transparent—it is clear what the differences are between groups of students who are eligible for a free lunch, eligible for a reduced price lunch, or not eligible. It is also clear what the differences are for students whose parents completed high school versus completed college.

A disadvantage to treating SES as a single variable is that SES is typically understood as having multiple components, including family income, parental occupational status, and parental educational attainment. Treating SES as only one of these components is at odds with the conventional definition of SES. There also is more measurement error in a single variable compared to a composite variable. Some of these problems could be overcome by treating SES as multiple separate variables. However, doing so complicates reporting and interpretation. The separate variables constitute separate operational definitions of SES, which could lead to

potentially conflicting data about the relationship between achievement and SES, defined in different ways. For example, achievement for levels of parental educational attainment might show greater variability than achievement for different levels of income.

A composite variable combines information from all the components in a single variable, maintaining simplicity in reporting and avoiding conflicting stories about relationships to achievement. This could mask differences between components in their relationship to achievement, of course, and that could be a potential disadvantage to a composite variable. Nevertheless, the advantages of a composite variable generally outweigh the disadvantages. The remainder of this chapter focuses on a composite measure of SES.

## General model

There are several ways to think about what a composite SES variable might look like and how it could be formed. In the psychometrics literature there are two kinds of latent variable models, formative and reflective (Bollen, 2002; Edwards & Bagozzi, 2000). A *reflective* measurement model is one in which the latent variable is assumed to be the cause of the measures or indicators (i.e., which are commonly called reflective or effect indicators, Blalock, 1964). For example, cognitive ability and personality are commonly assumed to cause responses to particular tests or test items. Changes in the latent variable cause changes in the indicator variables. Cronbach's (1959) alpha, factor analysis, and classical test theory are all reflective measurement models—covariation among indicators is assumed to be caused by an underlying latent variable. A *formative* measurement model is one in which the latent variable is assumed to be caused by the indicators (i.e., which are commonly called formative or cause indicators). That is, changes in the indicators cause changes in the latent variable. SES is commonly understood as a latent variable in a formative measurement sense because SES does not cause income, educational attainment, or occupational status; rather, income, educational attainment, and occupational status cause (or determine) SES. A fundamental difference between reflective and formative measurement is that latent variables in reflective measurement are defined by the degree to which indicators covary, and in fact the pattern of covariances can be used to compute latent factor scores. The situation is different with SES and other latent variables in formative measurement. Here, indicator variables have no necessary relationship with one another and can be uncorrelated or negatively correlated with each other. An example of formative measurement could be variables such as the stress scale (Holmes & Rahe, 1967), which is simply a count of the number of stress-inducing events experienced by

an individual within a relatively short period of time (e.g., a year). Indicators are life events, such as death of a spouse, imprisonment, personal injury, or pregnancy, none of which have any necessary relationship to each other (i.e., they can correlate positively, negatively, or be uncorrelated). However, the stress scale predicts future events, such as subsequent illness. In the same way, SES is useful for its ability to predict present and future academic achievement and other life outcomes.

To produce a composite index or score in formative measurement, some scheme must be used to weight the components in some fashion. Below is a review of several ways to do this.

## Arbitrary weighting

An infinite number of arbitrary weighting schemes are possible for forming an SES composite. For example, the number of years of parental educational attainment could be added to annual family income and a rating of job status to form an SES composite. However, the weight of these measures would be related to the variance of the components, and somewhat arbitrary re-scalings of components (e.g., changing from income in dollars to income in cents or to a three-level value, such as 1 = high; 2 = medium; 3 = low) could have dramatic effects on the composite. Putting components on the same scale (e.g., through the use of standard scores, or *z* scores) would be a way to avoid this problem. An analyst or policy maker might believe that parental educational attainment is the most important component of SES in an educational application such as NAEP, and so parental educational attainment could be given more weight (e.g., twice the weight) than the other components in forming an SES composite.

An advantage of arbitrary weighting is that it is easy to communicate the rules by which components are combined to form a composite. For example, the "misery index" is the sum of the employment rate plus the inflation rate; the United Nations Development Program's "human development index" is a more complicated geometric mean of normalized indices, but is nevertheless arbitrary. However, the arbitrariness of these indexes is a visible feature, in that their makeup is clear and transparent. A disadvantage of arbitrary weighting is that it is arbitrary. There is no reason to prefer one set of weights to another, and different weights might give different answers to substantive questions (such as, what is the relationship between SES and achievement?).

## Empirical weighting

If SES were treated as a latent variable with reflective indicators then component weights could be developed using factor analysis or principal component analysis of the indicators. A rationale for treating SES as a latent variable with reflective indicators is that its components correlate. Treating SES as a latent variable with reflective indicators implies that changing SES would result in a change in income, parental occupational status, and parental education, which seems implausible. Still, the reflective indicators assumption and approach to identifying and weighting an SES composite is used in PISA (see e.g., OECD, 2010a). In PISA, SES (PISA's Economic, Social, and Cultural Status index, or ESCS index) is computed from highest parental educational attainment (in number of years of education), highest parental occupation (converted to a status index), and number of home possessions (summing over 20 items), including books in the home. Weighting these three components is determined by a principal component analysis (conducted separately for each participating country) based on the covariances among the three components, and it has typically yielded approximately equal weights for the three components (although with job status given the most weight, education the second most, and home possessions the smallest). Thus PISA treats SES, at least partly, as a latent variable with reflective indicators.

However, as discussed previously, SES is more commonly thought of as a latent variable with formative indicators, because it is assumed that SES is caused by its indicators rather than the other way around. If SES is treated as a latent variable with formative indicators, then weights cannot be assigned by a covariance-based approach (e.g., principal component analysis) that only considers the components of SES. Such a system weights components according to their centrality (similarity or correlation) with respect to each other, but components do not have to be correlated with each other in a formative measurement model. Instead, under a formative variable assumption, an approach to forming a composite with non-arbitrary weights would be to compute weights through multiple regression analysis using an outcome variable. An outcome variable, such as NAEP Mathematics scores, could be regressed on the SES component variables, and the estimated weights could be used to form an SES composite optimized for predicting NAEP Mathematics scores for that particular grade and year. There are potential drawbacks to producing composite scores this way. One perspective is that SES emerged as a construct because of its predictive relationship with educational outcomes. It is therefore fair and reasonable to weight SES components according to the regression weights of those components when predicting educational outcomes, such as NAEP scores. A complexity

related to this perspective is that regression weights will, in general, change depending on which NAEP scores are being predicted (e.g., 4th-grade reading, 8th-grade mathematics, 12th-grade civics, etc.), and in what year they are being predicted (e.g., 2011, 2013). This issue is revisited in the next section of this chapter.

An alternative viewpoint is that SES should have an identity separate from its ability to predict particular achievement outcomes. This viewpoint reflects the perspective that little is learned by studying the relationship between SES and achievement if SES becomes little more than a set of variables optimally weighted to predict achievement. Instead, the relationship between SES and achievement should be a finding rather than an optimization exercise.

There are two ways out of this impasse between a tailored (i.e., regression-weighted) and independent (weights determined without regard to the composite's prediction of achievement) SES composite. If a composite predicts achievement equally well under a range of composite weights (e.g., unit weights, weights determined by regression with 4th-grade mathematics, or with 12th-grade reading), then the distinction between the two composite weighting approaches is of little practical importance. There is some evidence for this perspective (Noel-Miller & Hauser, 2011; Wilks, 1938).

Another way to define an SES composite empirically without tuning component weights to maximize prediction with NAEP scores is to consider additional outcome measures. This was an approach originally suggested by Hauser and Goldberger (1971) as the multiple indicator multiple cause (MIMIC) model (for a recent discussion regarding how this strategy helps identify formative latent variable models, see MacCallum & Browne, 1993). For example, SES is a widely used construct in the health literature, and a health outcome (e.g., absenteeism due to illness) could be used as an additional outcome variable that could be regressed on the SES components. Such a model could be estimated using a structural equation modeling (SEM) approach. Doing so would lead to an SES composite that was not being tuned specifically to the prediction of NAEP achievement. (Multiple NAEP population groups, for example, with various achievement scores at different grade levels, could also be used for this purpose, but the generalization would be to NAEP achievement, not to life outcomes in general.) At the same time, component weights would not be arbitrary, but would be based on the predictiveness of the SES composite across diverse outcomes.

## Measurement invariance goals

In developing an SES composite, regardless of whether arbitrary or empirical weights are used and whether SES is a latent variable with formative or reflective indicators, there is an issue concerning the degree to which the composite should be defined in a consistent (i.e., invariant) way (i.e., given the same component weights) across situations; that is, across grades, across NAEP subject areas, across time, across locations, and so forth. One (extreme) option would be to have a specific SES composite for every measurement occasion. That is, there could be a 4th-grade mathematics SES composite for New York in 2013, and a separate SES composite for New Jersey, and separate SES composites for each grade, for each subject, and for each testing year. There could be other categories by which SES composites could be separately formed, such as urban, suburban, and rural, or by cost-of-living areas (a given family income, say $40,000/year, might indicate different socioeconomic status depending on whether the family resided in Manhattan or the rural south). A National Research Council report on a new poverty measure (Citro & Michael, 1995) recommended adjustment for geographic differences in the cost of housing and insurance.

However, there is a benefit of having an SES composite that maintains the same component weights across all measurement occasions (i.e., across grades, subjects, locations, and years), namely, consistent SES measurement can simplify reporting and interpretation. In current reporting, SES indicators, such as parental educational attainment, NSLP eligibility, and home possessions, are measured the same across all contexts. That is, these indicators are invariant in the raw (manifest) metric across grades, subjects, locations, and years.

The issue of how and the degree to which the SES composite can be kept invariant is affected by the form of the SES composite. For example, with arbitrary weights for the components (sum of unit weighted parental educational attainment, parental occupational status, family income), educational attainment, occupational status, and income could be standardized separately within 4th, 8th, and 12th grade, or they could be standardized across grades (e.g., parental educational attainment could be placed on a common scale across all three grades, or on separate scales, one for each grade). With regression-based weighting for SES components, weights could be obtained from a regression analysis for one grade and one subject, and applied to other grades and subjects, or separate regression analyses for each grade and subject could be conducted, or weights could be averaged across subjects or grades (c.f., Noel-Miller & Hauser, 2011).

Weights could be identical or similar across subjects and years, which might make interpretation and reporting simpler. However, this might not be possible across grades due to the differences in information that can be collected from 4th-graders versus 8th- or 12th-graders. Based on prior research (e.g., Dawes, 1979; Noel-Warren & Hauser, 2011; Wilks, 1938), the weighting scheme might not have much impact on the identity of SES (i.e., applying two sets of component weights to construct an SES composite would likely result in two versions of SES that were highly correlated). Therefore a simpler approach, such as using unit weights or average weights (where average weights involve averaging the component weights obtained in one context with comparable weights obtained in another context), might be advisable. Examining the existing literature (Cohen, 1990) and closely reviewing the quality of data (once it was determined what data would be collected and from what sources) is the most appropriate course of action in determining weighting.

## Missing data issues

Dealing with the issue of missing data may be more critical in the case of composite variables compared to single variables such as parental educational attainment (or NSLP), simply because there are more opportunities for data to be missing (e.g., through skips by the respondent). If casewise deletion were invoked any time any of the component items for an SES composite were missing, that could result in both a relatively high number of missing values, and the introduction of bias if data were not missing completely at random (using the standard terminology from Little & Rubin, 1987).

However, there are probably no special problems associated with imputing missing data in the case of computing the SES composite. For example, a standard practice (e.g., used in PISA) is to impute missing values for students with missing data for one of the SES components using data from the other two components. In general, either a maximum likelihood approach for handling missing data in the context of modeling the data, or a multiple imputation approach similar to that used for handling missing achievement data, could be used and would be worth exploring for this purpose (Enders, 2010).

## Panel Recommendation: Create an SES Composite

1. The advantages of treating SES as a composite of several variables rather than as a single variable or multiple single variables outweigh the disadvantages.

2. The formative-reflective measurement model distinction was important in considering how to combine SES components into a composite measure. The literature and data quality should be examined before proposing a recommendation on a component weighting scheme.

3. Further study is necessary to address missing data issues in SES measurement.

# 8. IMPLICATIONS

Adopting a new measure of SES would have various implications on the reporting of NAEP scores. To begin with, a new measure would have to be clearly explained and communicated, because a new measure of SES might show greater achievement differences between low and high SES groups, compared to free lunch versus non-subsidized lunch groups. A sudden change in how SES was defined might therefore disrupt trends in the relationship between SES and NAEP achievement scores, which would create significant challenges to interpreting SES estimates over time.

## Reporting and Implications for Trend

As reviewed in chapter 2, achievement scores are disaggregated in NAEP reports by individual SES proxy variables, most notably eligibility for NSLP (*not eligible*, *eligible for a reduced-price lunch*, and *eligible for a free lunch*). Eligibility for a free or reduced-price lunch is a variable with three categories, which is convenient for reporting. A new measure of SES could, and likely would, be a continuous variable. In that case, a decision would have to be made about whether to transform the continuous variable into a categorical variable, or treat it in some other fashion. If it were transformed into a categorical variable, a decision would have to be made about how many categories it could be reported by (e.g., three, more?) and how these categories would be labeled (e.g., low, medium, and high SES).

A new measure of SES would not have to be treated as categorical, however. In PISA (OECD 2010a, Figure II.1.3, p. 32), for example, SES data are reported on a continuous scale, with scatter plots of achievement scores and the PISA index of economic, social, and cultural status (ESCS), and a regression line of achievement on ESCS. With ESCS presented as a continuous variable, PISA reporting makes considerable use of presentations (e.g., tables and scatter plots of ESCS against a variety of variables), and the use of ESCS as a control variable in examining factors such as single-parent families and the like. PISA also computes “socio-economic

gradients" that characterize the within-country relationships between ESCS and achievement, facilitating country-to-country comparisons on that measure.

A continuous SES variable could be used in NAEP reporting, but it would not have to be limited to presentations in scatter plots, or as a gradient index. For example, expected SES achievement at, say, the 20th, 50th, and 80th percentiles of SES could be presented, or at the mean SES and at a level one standard deviation above and below the mean of SES. These displays would take a form similar to that taken by NSLP eligibility.

For understanding trends in variables undergoing changes, as SES would be if a new measure were adopted, it is useful to conduct bridge studies, such as those conducted as a result of new race/ethnicity classifications introduced in Census 2000 (Parker, Schenker, Ingram, Weed, Heck, & Madans, 2004). For SES, a carefully constructed study enabling bridging to NSLP eligibility could be useful for understanding trends. For example, for a reporting cycle or two, both SES and NSLP eligibility could be reported as the audience became familiar with the new scale. This would allow readers to compare SES effect sizes (on achievement) with SES measured by NSLP eligibility versus SES measured by a new composite.

## Data conditioning

NAEP uses a balanced incomplete block design for administering only subsets of the item pool to particular students (i.e., each student only takes 2 of 11 blocks of items). Background information, including SES, along with data from the items actually administered, is used to estimate scores on the items that are not administered to a particular student, a process referred to as conditioning (Mislevy, 1991). Changing the measure of SES would likely lead to changes in the conditioning model, and changes in the posterior distributions of student responses from which plausible values that secondary analysts use are drawn. A question is how severe a difference in the conditioning model would likely result from a change in the makeup of SES (e.g., from NSLP to a new SES index). There is literature suggesting that the demographic variables are the most important background variables affecting the conditioning model (Thomas, 2002). And SES is likely to be among the more important demographic variables. That same literature, however, suggests that background variables are not as important to the conditioning model as the cognitive variables themselves (i.e., the responses to the cognitive items that are administered). As with the proposed bridge study, current variables, such as

NSLP, could be retained, and differences in the conditioning model due to the inclusion of a new SES measure could be studied.

## Use by other units, departments, agencies

The focus of the present effort is developing a new SES measure for NAEP. A new SES measure could have direct effects in reporting NAEP scores, such as providing a more valid estimate of the relationship between SES and achievement. In addition, SES is used for the conditioning model in NAEP to assist in the estimation of proficiency scores, and a better measure of SES could be more predictive of proficiency scores and thereby more useful for data conditioning. The quality of NAEP data reported could therefore improve as a result of a better SES measure. In addition to these specific benefits for NAEP reporting, there would be additional benefits based on secondary analysis of NAEP. SES, or proxy measures such as NSLP, is widely used in secondary analysis of NAEP data (e.g., Harwell & LeBeau, 2010; Sirin, 2005).

NCES programs beyond NAEP might benefit from the work conducted in defining and developing a new SES measure. NCES Fast Facts (2012b) provides a list of NCES surveys, many of which use SES measures of some kind. These include adult literacy surveys (National Assessments of Adult Literacy [NAAL], the Program for International Assessment of Adult Competencies [PIAAC]), international comparative surveys (Trends in International Mathematics and Science Study [TIMSS], Progress in International Reading Literacy [PIRLS], the Program for International Student Assessment [PISA]), longitudinal surveys (the Early Childhood Longitudinal Study [ECLS], Baccalaureate and Beyond [B&B], Beginning Postsecondary Students Longitudinal Study [BPS]), and so forth. For some of these studies there could be a fairly direct transfer of findings on improving SES measurement. For other studies some of the research, methods, and lessons learned in developing an improved SES measure could be incorporated into future study designs.

Investigating new methods for measuring SES could produce benefits that extend beyond NCES and the U.S. Department of Education. For example, in the health sector, there is an extensive literature that relates SES to women's health, public health, and psychological health (APA, 2007b); to specific conditions, such as cancer (Singh, Miller, Hankey, Edwards, 2003) and cardiovascular disease (Winkleby, Jatulis, Frank, Fortmann, 1992); and to other health and wellness issues. Agencies such as the National Institutes of Health (NIH) and the Center for

Disease Control (CDC) may benefit from research conducted for NAEP by NCES in developing improved measures of SES.

## Anticipated effects and unanticipated side effects

Developing a new SES measure is likely to involve both anticipated effects and unanticipated side effects. It is reasonable to assume that developing the new measure will involve an interagency agreement and collaboration between NCES and the Census Bureau. Such interagency collaborations are beneficial, but often introduce scheduling complications, new costs, and other challenges that require flexibility, patience, and a willingness to consider a variety of approaches to solving potential problems. Another relatively minor change will be a requirement to collect ZIP code information from respondents, perhaps on the NAEP questionnaire. Privacy issues are also likely to be important to resolve.

It is always difficult to anticipate the unanticipated side effects of measurement changes, but as with any assessment, new measurement is often accompanied by the element of consequential validity (Messick, 1995). Consequential validity refers to the changes in practice or culture that accompany changes in assessment. For example, introducing writing assessments can lead to an increased emphasis on writing instruction in the schools; introducing a high-stakes noncognitive skills assessment can lead to more emphasis on developing noncognitive skills. It is not entirely predictable what changes might accompany the introduction of a new SES measure, but if such a measure proves to be more valid than current measures, it is possible that more attention could be given to the importance of the SES-achievement relationship and to a more equitable distribution of educational resources.

## Panel Recommendation: Consider Implications of a New Measure of SES

1. There are reporting and psychometric implications that should be considered before implementation of a new SES measure. They include whether and how to characterize SES levels, whether to conduct a bridge study linking new and old measures of SES, and studying the implications of a new SES measure on the conditioning model used by NAEP to generate plausible values.

# 9. DISCUSSION

The goal of this panel was to provide recommendations for a new measure of SES that could be used in NAEP. The role of the white paper was to serve as technical documentation of the panel deliberations and to bring this work to the attention of stakeholders and the research community to engage discussion about SES and its measurement.

NAEP is required by law to report scores by SES. Current SES measures, such as NSLP eligibility and parental educational attainment, are single proxy variables, which are limited in several ways. Historically, SES has been defined as a composite measure reflecting resources available to the individual, as expressed in family income, parental educational attainment, parental occupational status, and sometimes neighborhood resources. A common view, as reflected in other large-scale educational assessments such as PISA, is that composite measures that include all of the SES components may be more informative than single measures.

A second limitation of current SES measurement concerns the quality of the data. Student reports of some SES components (such as parental educational attainment) may be unreliable and biased, and reports on variables like these by 4$^{th}$-graders are likely to be particularly unreliable. This is not to say that they are unusable. Attempts to collect data from 4$^{th}$-graders on parental educational attainment and perhaps even parental occupational status should be revisited. However, additional data sources such as NCES and state assessment databases and private data sources should also be considered to help bolster the quality of an SES measure.

Perhaps the most critical data quality issue in current SES measurement concerns NSLP eligibility. Measures of NSLP eligibility have several problems, including large errors in eligibility certification and jurisdiction-wide eligibility which fails to differentiate poverty levels within schools or jurisdictions where everyone is declared NSLP eligible (Harwell & LeBeau, 2010; Hauser, 1994). Most importantly, that trend is likely to continue and even get worse.

Given the current limitations of how NAEP measures SES, a major contribution of the panel was to devise a consensus definition of SES, based on a review of various perspectives on SES:

> *SES can be defined broadly as one's access to financial, social, cultural, and human capital resources. Traditionally a student's SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate*

*adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.*

Note that this definition outlines and provides a justification for both a core SES measure, which should be the subject of immediate focus for operational reporting, and a more expanded measure, which could be treated as a research project intended to illuminate some of the more contextual and explanatory aspects of SES.

There are other potential components of SES, such as subjective SES and psychological factors. These are best understood as contextual and explanatory variables that could help in the interpretation of SES-achievement relationships, but these contextual factors should not be considered part of a core SES meeting the charge of a congressionally mandated reporting variable. A research program studying these variables, however, could be critical for understanding the importance of measuring SES in the context of an educational achievement survey.

The panel reviewed existing and proposed new measures of SES components from sources including school records, the student questionnaire, additional potential NAEP questionnaire items that were pilot-tested in 2009, 2011, and 2012, and questions from the American Community Survey (ACS). Measures reviewed included ones pertaining to family income and home possessions, parental educational attainment, parental occupational status, and neighborhood wealth and resource indicators. Additional measures that might be related to family income and resources, such as housing tenure, number of residence moves, household members' healthcare needs, immigration status, and household composition measures were also considered. Some of these can be obtained from ACS data. Although NAEP 4$^{th}$-grade questionnaires do not ask students to indicate parental educational attainment in the questionnaire due to low data quality, such information can be obtained from ACS data. Occupational information is not asked about in the NAEP questionnaires, again due to concerns with low data quality, but such information can be obtained from ACS data. An extensive amount of neighborhood SES data could be obtained from the ACS, including neighborhood poverty levels, unemployment, educational attainment, presence of parks and libraries, abandoned buildings, single-parent households, and non-English speaking households. However, there are challenges in obtaining these kinds of data from ACS and for linking ACS data to NAEP data, such as determining how best to aggregate data in linking datasets.

There are a wide variety of ways to combine all the information on components of SES. A composite can be assembled by summing variables reflecting family income, parental educational attainment, parental occupational status, and neighborhood SES indicators. The primary distinction is in whether the summing would occur by arbitrarily weighting the components (e.g., unit weighted), or by allowing the components to be weighted to best predict some outcome, such as student achievement. There are advantages and disadvantages to both approaches. Another important consideration would be how to maintain the meaning of SES across grades, across locations (e.g., varying cost-of-living regions), and across time. However, with respect to the issue of component weighting, there is some evidence that this may be merely academic and that practically how variables are weighted might not make much difference in what SES is (Noel-Miller & Hauser, 2011). That is, an SES composite with a set of weights determined from one context is likely to be highly correlated with an SES composite based on a set of weights determined from a different context, given that the components themselves tend to be highly correlated, and neither will be much different from a unit-weighted composite, as has been long known (Wilks, 1938).

Developing a new SES measure for NAEP has implications for reporting and elsewhere. If a new measure were developed, it might be useful to report achievement results disaggregated by SES, measured both by the current measures (e.g., NSLP eligibility, parental educational attainment) and by the new composite measure. While it may be valuable to treat SES as a continuous variable, it could also be treated as a categorical variable (e.g., low, medium, and high SES). The research and findings resulting from developing the new SES measure for NAEP would benefit other federal programs both within and outside NCES.

## Key Recommendations

Summarized below are the panel's key recommendations for improving the measurement and reporting of SES.

*Recommendation 1.* Family income and other indicators of home possessions and resources, parental educational attainment, and parental occupational status (the "big 3"), should all be considered components of a core SES measure; that is, part of the measurement of a core SES variable. The core SES measure should be the subject of immediate focus for operational reporting. This recommendation reflects the academic literature on SES.

*Recommendation 2.* An expanded SES measure could include additional variable components besides family income, parental educational attainment, and parental occupational status. These additional components could include resources available in the student's neighborhood or community and resources available at school. Consideration should be given to the development of an expanded SES measure in addition to the core SES measure.

*Recommendation 3.* The advantages of treating SES as a composite—e.g., a single summary for reporting, greater reliability, and representation of the full range of SES factors—outweigh the disadvantages, especially because the use of the composite would not preclude using and reporting on single measures. Therefore, attempts should be made to develop an SES composite measure.

*Recommendation 4.* The validity of the most widely used measure of SES—NSLP eligibility—has been decreasing due to jurisdiction-wide eligibility and other factors, and that trend is likely to continue. There will be growing pressure to replace NSLP eligibility with a new, more valid measure. Burden issues prohibit a longer questionnaire, and there is concern about the reliability of student reports on SES components, particularly educational attainment (for 4$^{th}$-graders) and occupation (for all grades). Because of data quality issues, along with burden considerations, attempts should be made to explore the possibility of linking to Census data on SES components. Studies should be conducted with the U.S. Census Bureau to determine the feasibility of linking Census data to NAEP and to evaluate the quality of the data that would result from various linking strategies.

## REFERENCES

American Psychological Association. (2007a). *Office on socioeconomic status*. Retrieved from http://www.apa.org/pi/ses/homepage.html.

American Psychological Association. (2007b). *Report of the APA Task Force on socioeconomic status*. Washington, DC: Author. Retrieved from http://www2.apa.org/pi/SES_task_force report.pdf.

Barton, P. E. (2003). *Parsing the achievement gap: Baselines for tracking progress.* Princeton, NJ: Policy Information Center, Educational Testing Service.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley and Sons.

Blalock, H.M. (1964) *Causal Inferences in Nonexperimental Research*, Chapel Hill, NC: University of North Carolina Press.

Bollen, K.A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53(1),* 605-634.

Bryant, E. C., Glazer, E., Hansen, M. A., & Kursch, A. (1974). *Associations between educational outcomes and background variables [Monograph]*. Denver, CO: National Assessment of Educational Progress.

Burton, L., Price-Spratlen, T., & Spencer, M. B. (1997). On ways of thinking about measuring neighborhoods: Implications for studying context and developmental outcomes for children. In J. Brooks-Gunn, G. Duncan, & J. Lawrence Aber (Eds.), *Neighborhood poverty: Context and consequences for children* (Vol. 2, pp. 132–144). New York: Russell Sage Foundation Press.

Chapin, F. S. (1933). *The measurement of social status, 3,* University of Minnesota Press.

Citro, C.F. & Michael, R.T. (1995). *Measuring poverty: A new approach. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council.* Washington, DC: National Academy Press.

Cohen, Jacob. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312. doi:10.1037/0003-066X.45.12.1304

Coleman, J. S. (1988). Social capital in the creation of human capital. American Journal of Sociology, 94, S95–S120.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., et al. (1966). *Equality of educational opportunity (2 vols.).* Washington, DC: U.S. Government Printing Office.

Cowan C.D., Sellman S.W. (2008). *Improving the quality of NAEP socioeconomic status information: Report on research activities.* Alexandria, VA: Human Resources Research Organization (HumRRO). (Retrieved from http://www.humrro.org/corpsite/publication/improving-quality-naep-socioeconomic-status-information-report-research-activities, February, 2012.)

Credé, M., Bashshur, M. & Niehorster, S. (2010) Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment, 92,* 390 – 399.

Cuff, N. B. (1934). The vectors of socio-economic status. Pe*abody Journal of Education, 12 (3),* 114-117.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist 34,* 571-582. doi:10.1037/0003-066X.34.7.571.

Demakakos, P., Nazroo, J., Breeze, E., & Marmot, M. (2008). Socioeconomic status and health: The role of subjective socioeconomic status. *Social Science and Medicine, 67,* 330-340.

Duncan, G.J., & Magnuson, K. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, *15,* 35-54.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of the relationship between constructs and measures. *Psychological Methods, 5,* 155-174.

Enders, Craig K. (2010). *Applied Missing Data Analysis*. New York: Guildford Press.

Fertig, M. (2003). Educational Production, Endogenous Peer Group Formation and Class Composition – Evidence from the PISA 2000 Study, IZA DP No. 714, Bonn, Germany: Institute for the Study of Labor.

Food and Nutrition Service. (1990). Study of Income Verification in the National School Lunch Program, Final report (Vol. 1, Rep. No. FNS- 53–3198-6–44, prepared by R. St. Pierre, M. Puma, M. Battaglia, & J. Layzer). Alexandria, VA: U.S. Department of Agriculture. Retrieved from http://www.fns.usda.gov/oane/MENU/Published/CNP/cnp-archive.htm.

Frederick, C., & Hauser, R. M. (2008). Have We Put an End to Social Promotion? Changes in School Progress among Children Aged 6 to 17 from 1972 to 2005, *Demography*, 719-740.

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research 21(1),* 1-56.

Harwell, M. R., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES Measure in education research*. Educational Researcher, 39,* 120-131.

Hauser, Robert M. (1969) Schools and the Stratification Process. *American Journal of Sociology, 74 (6),* 587-611.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, *65*, 1541–1545.

Hauser, R. M., & Andrew, M. (2007). *Reliability of student and parent reports of socioeconomic status in NELS-88*. Working paper presented at ITP seminar at the University of Wisconsin–Madison. Retrieved from http://www.wcer.wisc.edu/itp/Spring%20 08%20seminar/HauserNELS-SES%20measurement_070607a.pdf.

Hauser, R. M., Frederick, C., & M. Andrew (2007). Grade Retention in the Age of Standards-Based Reform. In A. Gamoran (Ed.) *No Child Left Behind and Poverty* (Pp. 120-53). Washington, DC: Brookings Institution.

Hauser, R. M., & Goldberger, A.S. (1971). The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological Methodology (pp. 81-117).* San Francisco: Jossey-Bass.

Hauser, R. M., Pager, D. I., & Simmons, S. J. (2004). Race-Ethnicity, Social Background, and Grade Retention. In H. Walberg, A. J. Reynolds, & M. C. Wang (Eds), *Can Unlike Students Learn Together? Grade Retention, Tracking, and Grouping.* (pp. 97-114).Greenwich, CT: Information Age Publishing.

Hauser, R. M., Simmons, S. J., & Pager, D. I. (2004). High School Dropout, Race-Ethnicity, and Social Background from the 1970s to the 1990s. In G. Orfield (Ed.), *Dropouts in America:*

*Confronting the Graduation Rate Crisis* (Pp. 85-106). Cambridge, MA: Harvard Education Press, 2004.

Hauser, R. M., & Warren, J. R. (1997). Sociological indexes for occupations: A review, update, and critique. *Sociological Methodology, 27*, 177–298.

Holley, C. E. (1916). *The relationship between persistence in school and home conditions*. Chicago, IL: University of Chicago Press.

Holmes, T.H., & Rahe, R.H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research, 11(2),* 213–218.

Ingels, S. J. (1990). *National education longitudinal study of 1988: base year: student component data file user's manual.* Washington, D.C.: U.S. Dept. of Education, Office of Educational Research and Improvement. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002323.

Kurki, A., Boyle, A., & Aladjem, D.K. (2005*). Beyond free lunch: Alternative poverty measures in educational research and program evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Lafontaine, D., Monseur, C. (August, 2007). Why do non-cognitive variables better predict mathematics achievement in some countries than in others? A methodological study of PISA 2003. Earli Conference, Budapest

Levin, H. M. & Belfield, C. R. (2002). Families as contractual partners in education. *UCLA Law Review, 49*, 1799-1824.

Little, R.J.A., & Rubin, D.B. (1987) *Statistical analysis with missing data.* New York: Wiley.

Lynd, R. S., & Lynd, H. M. (1929). *Middletown: A study in American culture*. New York: Harcourt Brace.

MacArthur (2011). *Research Network on SES & Health: Research, Psychosocial Notebook.* http://www.macses.ucsf.edu/research/psychosocial/usladder.php (retrieved 4 August 2011).

MacCallum, R.C., & Browne, M.W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114(3),* 533-541.

Mayer, Susan E. 1997. *What money can't buy: Family income and children's life chances.* Cambridge, MA: Harvard University Press

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56,* 177-196.

National Assessment Governing Board. (2003). *Background information framework for the National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education, National Assessment Governing Board.

National School Lunch Program. (2008). *Fact sheet*. Retrieved from http://www.fns.usda.gov/cnd/governance/notices/naps/NAPs.htm.

NCES (2011). The Nation's Report Card: Science 2009 (NCES 2011–451). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

NCES (2012a). *Early Childhood Longitudinal Program*. Retrieved from http://nces.ed.gov/ecls/Kindergarten.asp.

NCES (2012b). Fast Facts. http://nces.ed.gov/fastfacts/display.asp?id=11

No Child Left Behind Act of 2001. Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Noel-Miller, C. & Hauser, R. M. (2011). Unpublished manuscript.

Office of Research, Nutrition, and Analysis. (1994). *School lunch eligible non-participants: Final report* (Rep. No. 53–3198–018, prepared by F. Glantz, R. Berg, D. Porcari, E. Sackoff, & S. Pazer). Alexandria, VA: U.S. Department of Agriculture. Retrieved from http:// www.fns.usda.gov/oane/MENU/Published/CNP/cnp-archive.htm

OECD (2010a). *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II).* Retrieved from http://dx.doi.org/10.1787/9789264091504-en, December 6, 2011.

OECD (2010b). *Education at a glance 2010: OECD indicators*. Retrieved from http://www.oecd.org/document/52/0,3746,en_2649_39263238_45897844_1_1_1_1,00.html.

Parker, J.D., Schenker, N., Ingram, D.D., Weed, J.A., Heck, K.E., Madans, J.H. (2004). Bridging between two standards for collecting information on race and ethnicity: An application to Census 2000 and vital rates. *Public Health Reports (119),* 192-205.

Reynolds, A. J., & Walberg, H. J. (1992). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology, 84,* 371-382.

Rivas, S, & Hauser, R. M., (2008). *Progress Report: Toward Improved Socioeconomic Measurement in NAEP.* University of Wisconsin. Unpublished manuscript.

Sewell, W.H. (1940). The construction and standardization of a scale for the measurement of socio-economic status of Oklahoma farm families. *Oklahoma Agricultural and Mechanical College Experimental Station, Technical Bulletin No. 9*.

Sewell, W. H., Haller, A., & Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review,* 34, 82-92.

Sims V. M. (1927).*The measurement of socioeconomic status.* Bloomington, IL: Public School Printing Co.

Singh, G.K., Miller, B.A., Hankey, B.F., Edwards, B.K. (2003). *Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975–1999.* NCI Cancer Surveillance Monograph Series (4), NIH Publication No. 03-5417. Bethesda, MD: National Cancer Institute.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453.

Spaeth, J. L. (1976). Cognitive complexity: A dimension underlying the socioeconomic achievement process. In W. H. Sewall, R. M. Hauser, & D. L. Featherman (Eds.), *Schooling and achievement in American society* (pp. 103-131). New York: Academic Press.

Taussig, F. W. (1920). *Principles of economics.* Newcastle: Cambridge Scholars Publishing.

Thomas, N. (2002). The role of secondary covariates when estimating latent trait population distributions. *Psychometrika, 67(1),* 33-48.

Trends in International Mathematics and Science Study (2003). *Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Retrieved from http://pirls.bc.edu/timss2003i/PDF/t03_af_book.pdf

Tucker-Drob, E. M., & Harden, K. P. (2012). Intellectual interest mediates gene-by-socioeconomic status interaction on adolescent academic achievement. *Child Development, 83,* 743-757.

van Ewijk, R. & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5 (2)*, 134-150.

Watkins, T.J. (1997). Teacher communications, child achievement, and parent traits in parent involvement models. *Journal of Educational Research, 91(1),* 3-14.

Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *Review of Higher Education*, *27*, 45–73.

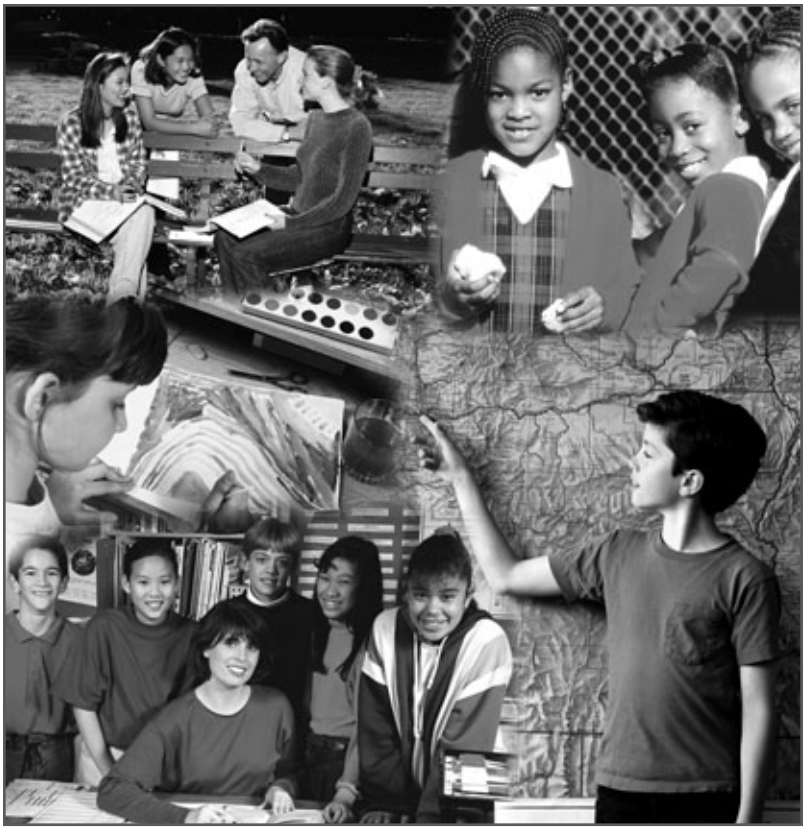
Wenglinsky, H. (1998). Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement. *Education Evaluation and Policy Analysis, 20(4),* 269-283.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481.

Wilks, S. S. (1938). "Weighting systems for linear functions of correlated variables when there is no dependent variable". *Psychometrika*, *3*, 23–40.

Winkleby, M.A., Jatulis, D.E., Frank, E., Fortmann, S.P. (1992). Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health, 82(6),* 816-820.

# Background Information Framework for the National Assessment of Educational Progress



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

# Chapter Three

## Topics and Types of Background Data

This chapter will cover the non-cognitive topics that are required for reporting under the No Child Left Behind Act of 2001 (P.L. 107-110), as well as those that should be considered for inclusion in NAEP on a cyclical basis. It discusses socio-economic status (SES), contextual factors of interest to public policy, and subject-specific variables.

### Demographic Reporting Categories

The demographic variables currently collected by NAEP are gender, age, race/ethnicity, and two elements of socio-economic status (SES)—participation in Title I, and eligibility for free or reduced-price lunch, which is based on family income. In addition, information is obtained on disability status and on students who are classified as limited English proficient. All of this information is collected on an administration roster, completed from school records in advance of testing. In addition, data on race/ethnicity is also collected on the NAEP student questionnaire, and students are asked to report on two other SES variables: the number of reading materials at home and the highest level of each parent's education.

A more extensive questionnaire is completed by school staff on each student selected for NAEP who is classified as either disabled or limited English proficient (LEP). For students with disabilities (SD), the questionnaire collects data on the specific disability and its severity, the student's Individualized Education Plan (IEP), type of curriculum, whether the student participates in standardized testing (with or without accommodations), and the accommodations allowed on state and district standardized tests in presentation, response, setting, and timing. For LEP students, the questionnaire covers native language, number of years of academic instruction in English, percent of instruction in English and/or native language, and the testing accommodations provided under district or state policy. In the future, NAEP might also identify students who recently exited from LEP programs and track their achievement.

NAEP is required to collect information on all of these categories (except age), but has some discretion in determining definitions and aggregating responses. These data will continue to be collected in a uniform manner in every NAEP assessment, although, for socio-economic status, as explained in the section below, there may be some variation, with a uniform core and more extensive data-gathering in some cases.

## Socio-Economic Status (SES)

Under current law, NAEP is required to collect information on socio-economic status. SES also is clearly a factor that has been shown to be related to academic achievement in many research studies, beginning with the Equality of Educational Opportunity Commission Report (Coleman et al., 1966). The research community's consensus over the past four decades has been to deal with the influence of SES on other achievement-related variables by holding SES constant while examining the other effects, for example, adjusting for SES while looking at effects of class size or teacher training. NAEP does not adjust for SES, but it does report on the relationship between student achievement and SES proxy variables like parents' education or Title I participation.

NAEP has not been able to measure SES directly, using its present set of questions and data sources, i.e., the student, teacher, and school questionnaires. The assessment has used "proxy variables" for SES, including students' eligibility for the National School Lunch program, participation in Title I, parents' education, and the number of reading materials in the home (newspapers, magazines, books, etc.)—information on the latter two factors being reported by students in the assessment samples. In addition, NAEP uses census data to classify schools by type of location, based on Census Bureau definitions, such as central city, suburban/large town, and rural/small town.

Strictly speaking, these are individual proxy variables and are not combined into a composite variable. However, both the questions on parent education and home environment are coded in a pseudo-composite manner. For example, the parent education related to the student is the higher of either the mother's or father's education level. On the four home environment questions, student responses are coded differently for a "yes" answer to two questions or fewer,

"yes" to three questions, and "yes" to four questions, as well as omitted responses (Allen, Carlson, and Zelenak, 1999).

At the lower grade levels, students' reports of their parents' education are questionable at best, while the National School Lunch program sorts students only into three categories (Yes, No, and Unknown) and Title I into two categories (Yes or No). For many years, NAEP used a reporting category of disadvantaged urban schools, which was constructed from information provided by school principals. This was discontinued in the mid-1990s because the category lacked a consistent definition from year to year and between different state samples. There also were serious doubts about the reliability of the information on which it was based. In short, there has been considerable concern over many years about the quality of the SES measures in NAEP, both for reporting to the public and for analysis by researchers.

Barton (2002) suggests two alternative approaches for improvement: (1) a composite index for SES, or (2) a parent questionnaire. A composite index is viable using the same information that is currently collected in NAEP, or perhaps augmented with a few targeted questions or census data, possibly the zip code of student home addresses. ***The necessary analytical work should be initiated through small research studies using extant NAEP data sets in order to check systematically the validity of a composite index as a better measure of SES in NAEP samples. The results could vary by grade level, in which case, adjustments might be needed in the way the data are collected, augmented, and/or confirmed. NAEP may never be able to produce a full composite of income, education, and occupation, but efforts ought to be made to find an approximation that is more reliable than the current set of individual proxy variables.***

The argument in favor of this approach is that it advances the goals of the current law without impacting data collection in unforeseen ways. Barton suggests that such an index would enable NAEP to report results in terms of SES quartiles (much the same way that the National Educational Longitudinal Survey, NELS, does). Further, it would allow the assessment to report cross-tabulations on distributions of students in the NAEP achievement level categories by SES. A good measure of SES would improve the monitoring of achievement gaps among various racial/ethnic groups, although

sample sizes may not be large enough within all ethnic groups or types of schools. Finally, a composite SES index may be beneficial to states and districts in the Trial District Assessment, enabling NAEP to compare the performance of groups of students with the same socio-economic status, which is a factor of high public and policy interest.

The argument against such an approach is that SES would continue to be measured indirectly, i.e., by using proxy variables, albeit through a composite index. There would also be disagreements about precisely which variables to include in the index and how to weight different factors. For example, Armor (D. J. Armor, personal communication, December 18, 2002) has suggested that two variables recently deleted from the NAEP student questionnaire be reinstated, namely, the number of siblings in the home and family status (student lives with both parents, mother or father, neither). These variables were dropped because of concerns about intrusiveness, but they may be of considerable importance in constructing an SES index. The Board will have to weigh the considerations involved, and may decide there is value in using them periodically or in limited samples.

A parent questionnaire has been proposed as a more reliable means of collecting SES data than relying on student reports, school records, or census data. Other National Center for Education Statistics surveys, for example, NELS and the Early Childhood Longitudinal Study, have employed parent questionnaires that ask direct questions regarding occupation and income.

However, the National Assessment of Educational Progress involves far more students than any of these research surveys. Accordingly, a parent questionnaire on NAEP would entail far more respondent burden and might arouse more controversy, making it more difficult to accomplish the primary mission of the assessment to measure student achievement. A parent questionnaire has been considered by NAGB in the past, but rejected as too burdensome and intrusive. Because these considerations are still persuasive, particularly as the scope of NAEP has expanded, no work should be undertaken on developing a parent questionnaire.

***In sum, because of its importance and the requirements of law, information on socio-economic status must be collected in all***

***NAEP samples, although there may be some variation in the number of factors on which data are obtained. Research should be conducted into creating a composite index of SES.***

***A core of SES information should be collected in every assessment, such as type of community (e.g., central city, rural, etc.), poverty status (e.g., eligibility for free or reduced-price lunch and Title I participation), reading materials in the home, and level of parent education—though steps must be taken to ensure that such data are reliable. An expanded set of SES variables may be included periodically and administered to limited samples, including such factors as number of siblings and parents at home, possession of computers, and parent occupation.***

***NAEP should explore the use of an SES index derived from proxy variables currently in either the administration roster or student questionnaire. To the extent that an index can be sharpened by additional information from readily available sources, such as zip codes and/or census data, this option should be considered as well.***

## Public Policy Contextual Factors

For the past two decades NAEP has collected information on student, teacher, school, and beyond-school factors that are of interest to policymakers and the public. For students, some of these factors have included course-taking patterns, television watching, homework, and use of computers. For teachers, the contextual factors have included educational background, credentials, years of experience, and participation in professional organizations, to name a few.

The lists of factors have been long. They have become burdensome both to respondents and to the efficient scoring, analysis, and reporting of the NAEP survey. The way they have been reported—through simple one-way tabulations—has encouraged unwarranted conclusions about cause-and-effect relationships.

***We propose a careful review of the contextual factors on which information is collected by NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. Data should be obtained periodically, on a rotating basis, over several NAEP cycles, although a limited set of factors may be included***

***in every assessment. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.***

Two documents are helpful in surveying the research base and presenting alternatives for NAGB to consider. The first is *Monitoring School Quality: An Indicators Report* (Mayer, Mullens, and Moore, 2001), prepared by Mathematica Policy Research, Inc., for NCES. This report presents a research synthesis, indicating factors for which there is a research base showing a strong relationship to academic achievement. The synthesis, involving a review panel as well as statistical analyses, identifies the following as factors related to student achievement results: the academic skills of teachers, teacher assignments (such as out-of-field teaching), course content, student discipline and school safety, class size, and focus on academic achievement. Other sources of information are available on all of these factors, but only through NAEP can they be related to the achievement of broad groups of students over time.

The second document, *Making Connections* (Greenberg, Stancavage, Farr, and Bohrnstedt, 2001), was prepared for NCES by the American Institutes for Research and presents an elaborate typology of non-cognitive variables that could be measured by NAEP. It is organized into seven broad categories of non-cognitive information related to students, instructional content and practice, teachers, schools, school community factors, beyond school factors, and federal, state, and district policy. The listing goes beyond what NAEP can and should handle, but its discussion is thoughtful and the document is useful for planning.

## Subject-Specific Background Data

For each subject assessed by NAEP, additional subject-specific background information has been collected from students, teachers, and schools. These data fall into the broad category of instructional content and practice. Under that umbrella come such topics as the curriculum taught, course offerings, class management and style, ability grouping, and modes of instruction. Subject-specific data collection has expanded enormously over the past two decades, and in recent years has included five to ten minutes of questions for students, about 30 minutes of questions for teachers, and 30 to 45 minutes of questions for school administrators.

Now is the time for these questions to be focused, limited, and prioritized. Future subject-matter frameworks adopted by the Governing Board should spell out clearly what these priorities will be.

A design for doing this was presented to the Board in the 1996 report of the Design/Feasibility Team of prominent researchers (Forsyth et al., 1996). The group recommended that a core set of non-cognitive questions should be administered to students each time a subject is assessed by NAEP. In addition, a more comprehensive questionnaire would be given whenever a new framework is introduced and repeated every eight to ten years. For example, an extensive set of background questions in reading and mathematics (grades 4 and 8) was administered in 2003, the baseline year for the No Child Left Behind legislation. Another complete set should be administered in mathematics in 2005 and in reading in 2009, the years in which revised frameworks are first used, and then should be repeated every eight years. In the intervening years, only the more limited core modules will be administered. Similar patterns should be established for the school and teacher questionnaires.

The NAEP assessments in other subjects, such as writing, science, history, geography, and civics, should have a core set of non-cognitive questions administered to the full sample, with longer, more extensive questionnaires being administered to smaller subsamples. With states now required to participate in NAEP every two years, the total number of students tested has expanded substantially. This makes even more compelling the case for limiting the NAEP background questionnaires and rotating the background questions.

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

Table 1 presents a model schedule for comprehensive and core sets of subject-related variables through 2013. It is based on the schedule of assessments approved by the Board in May 2003.

**Table 1. Model Data Collection Schedule for Comprehensive and Core Sets of Non-Cognitive Variables by Subject Area**

| Subject Area | Data Collection Year for Comprehensive Set of Variables | Data Collection Year for Core Variables Only |
|---|---|---|
| Reading | 2003, 2009, | 2005, 2007, 2011, 2013 |
| Mathematics | 2003, 2005, 2013 | 2007, 2009, 2011 |
| Foreign Language (12) | 2004, 2012 | |
| World History (12) | 2010 | TBD |
| Economics (12) | 2006 | TBD |
| Civics | 1998, 2012 | 2006 |
| Writing | 2002, 2011 | 2007 |
| Arts (8) | 1997, 2008 | |
| Science | 2000, 2009 | 2005 |
| US History | 2001, 2006 | |
| Geography | 2001, 2010 | |

NOTE: Based on schedule approved by NAGB on May 17, 2003.

# NAEP BACKGROUND QUESTIONS: AN UNDERUSED NATIONAL RESOURCE

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

## February 22, 2012

*Chair: Marshall S. Smith*

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

# Table of Contents

# 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<br>•Rotated questions<br>•Policy questions<br>•Theoretical frameworks<br>•Consistent questions overtime<br>•Delete duplicative or low-priority questions | •Valid<br>•Reliable<br>•Coordinated (with domestic and international surveys)<br>•Cognitive labs | •Spiral sampling<br>•Extended questionnaire time<br>•Alternate surveys<br>•Pooling item responses across surveys | •Special background question reports<br>•Online compendium of responses<br>•Report descriptive not causal findings<br>•Externally conducted research<br>•Improve online tools |
| • Establish a single NAGB committee overseeing background questions<br>• 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 Avaialble 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<br>•Rotated questions<br>•Policy questions<br>•Theoretical frameworks<br>•Consistent questions overtime<br>•Delete duplicative or low-priority questions | •Valid<br>•Reliable<br>•Coordinated (with domestic and international surveys)<br>•Cognitive labs | •Spiral sampling<br>•Extended questionnaire time<br>•Alternate surveys<br>•Pooling item responses across surveys | •Special reports<br>•Online compendium of responses<br>•Report descriptive not causal findings<br>•Externally conducted research<br>•Improve online tools |
| • Establish a single NAGB committee overseeing background questions<br>• Review budget include need for staff to implement recomendations | | | |

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**<br><br>Questions:<br>- Student & family background and out-of-school learning<br>- Subject specific: self-perception and school courses content | | | **Teachers 30 Min**<br><br>Questions (subject specific):<br>- Teachers Background: education and training;<br>- Classroom Organization and Instructional practices | | | **Schools 30 Min**<br><br>Questions:<br>- School Characteristics (including a special charter school survey)<br>- Subject specific: course, student placement, staff composition, training, technology | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gr. 4 (2011) | Gr. 8 (2011) | Gr. 12 (2009) | Gr. 4 (2011) | Gr. 8 (2011) | Gr. 12 (2009) | Gr. 4 (2011) | Gr. 8* (2011) | Gr. 12* (2009) |
| Math: 2011 | 31 | 30 | 40 | 48 | 31 | | 39 | 49 | 48 |
| Reading: 2011 | 32 | 26 | 34 | | 30 | | | | |

*School questionnaire for grades 8 and 12 covers reading, math and science. Teacher questionnaire is not administered at grade 12.
Source: NAEP Background Questionnaires. Available Feb 2012: http://nces.ed.gov/nationsreportcard/bgquest.asp

A 10-minute student questionnaire consisting of approximately 30 questions asks about family background, school and home experiences, and out-of-school learning activities.

- Since NAEP does not administer a questionnaire to survey parents, the student questionnaire is the primary source of information on students' home characteristics and out-of-school learning activities. (School records do provide an alternative source for race, ethnicity and school lunch eligibility data).
- With respect to socio-economic status, grade 4 students are only asked about household items (computers in the home, numbers of books). Students in grades 8 and 12 are also queried about their mother's and father's highest level of education.
- A few questions are asked about students' out-of-school learning-related activities -- talk about things studied in school, read for fun on your own time, or studying and reading at an after-school program.
- A few items are included about student self-perception and enjoyment of a specific subject, for example whether reading and math are favorite subjects.
- Students are asked a number of questions about their classes in the subject assessed – for example, the frequency of reading aloud and discussing what they read in class, and in math many questions about using technology (calculators, graphing programs and spreadsheets).

A 30-minute *teacher* questionnaire of 30-40 questions is filled-out by the teacher in grade 4 or 8 in the subject assessed, usually the classroom teacher at grade 4 and the English or mathematics teacher at grade 8. This questionnaire covers:

- Teacher background information on race/ethnicity, education, certification and experience and professional development.
- Classroom organization items about class size, hours of instruction and ability grouping.
- Instructional items about topic emphasis, instructional approach, homework, evaluating student progress and access to resources and technology. The math questionnaire includes extensive questions about calculators of all types, computers, the Internet and CD-ROMs.

A 30-minute *school* questionnaire of about 40 questions covers:

- Overall school characteristics including grades, status as a charter, student composition and turnover, teacher absenteeism, volunteerism, and Title I federal program participation.
- Subject-specific items about specialist staff, structuring of content with standards and assessments, resource availability with emphasis on technology,
- Special charter school questionnaire about legal status and focus of charter.

Looking across the surveys, several issues of questionnaire coverage emerge:

- The student questionnaire includes items obtainable elsewhere and may be duplicative. For example, student-reported information on classroom instructional approaches overlaps with information on the teacher questionnaire.

- Although the three surveys collectively cover a broad range of important background topics, the surveys omit a few topics with a strong base supporting their relationship to achievement. Two examples are the degree to which schools reach out to parents, and school discipline and the climate for learning.
- The questionnaires largely ignore major national policy issues prominent over the last decade involving the response to federal mandates for state-based student testing and high-stakes accountability.

The panel believes there is a need to address these and other issues of questionnaire content through a systematic process for identifying topics and questions that best relate to understanding NAEP student achievement results without being excessively burdensome or invasive.

### Recommendation 1a. Continually review and refine a core and second-tier set of background topics and questions that are common across NAEP surveys.

- NAEP should build on its current process for specifying a *common core* set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.
- NAEP should develop a *second tier* of priority background question clusters that could be rotated across assessment cycles. Important topics that might be explored include school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time.
- NAEP should prioritize core and second tier items in terms of information value and respondent time, select high-priority items, and eliminate current low-priority items.
- NAEP should regularly publish its background questionnaires and provide justifications for all questions asked in terms of research and policy. Core and second-tier background questions should be identified.

**Discussion**

This recommendation would expand NAEP's current set of *core* background questions focused primarily on the congressionally required *student subgroups*. The panel recommends including as an additional part of the core, a second cluster for *instructional and other school learning opportunities*. This cluster would allow examination of student learning environments by describing the curriculum, instructional approaches, and teacher qualifications. Many of these types of questions are now included in the teacher questionnaire and would be folded into this category.

A third core cluster of core questions is recommended to cover the area of *student motivation and control over the environment*. Measures such as whether students believe that success depends more on ability than effort or students' locus of control have been

documented over several decades as strongly related to academic performance (Coleman, 1966; Chen & Stevenson, 1995). Also, students' educational expectations predict their educational achievement and occupational expectations predict occupational attainment (ETS, 2010). When good teachers and a positive school environment influence student motivation and expectations this in turn will lead to improved achievement.

A *second tier* set of question clusters is proposed to focus on items for which there is strong research backing of their relation to achievement, but for which rotated items across alternate assessments (e.g., every four years) would be acceptable. As noted above, these second tier clusters could describe school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time. Specific clusters should vary across time as achievement levels and educational practices and policies change.

Together these clusters of items would view gains in school achievement as driven by a simple theory that sees gains in learning as a function of the curriculum, learning time, quality of instruction and student motivation These core and second-tier clusters meet the principle in the Board's Background Information Framework that "The information obtained be of value in understanding academic performance and taking steps to improve it" (2003 Background Information Framework).

The Panel recognizes that in defining these clusters NAEP will have to establish tradeoffs in terms of meeting the constraints of questionnaire length and cost. These decisions should be based on the priority of a question or question cluster in terms of information value balanced against respondent burden and costs. To make room for new high-priority items NAEP should consider eliminating or reducing low-value or duplicative questions as noted below. Time constraints may also be addressed by rotating questions on alternate survey administrations (i.e., four-year intervals) NAEP also constrains the student questionnaire length to ten minutes when TIMSS even at grade 4 is 30 minutes.

### Recommendation 1b. Extend NAEP background questions to inform topics of current policy interest.

- Implementation of this recommendation could focus on three rotating sets of policy questions each extended over a six-year period. For example, the initial set might start with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations would be added, and two years after that a module on project-based or online learning. After six-years, questions on a new policy issue would be introduced to replace the first. Using this approach each of the question sets would have three observations over the six-year time.

- The panel concurs with the 2003 Background Report caution to include only policy-relevant questions that are answered on the basis of fact rather than opinion. That is, the responses to policy-relevant questions should be objective and not reflect personal beliefs. Questions should ask about policy responses, such as training received to understand new standards or the extent to which new

standards have changed instructional content or approaches. Questions should not elicit judgments about personal policy preferences.

- The policy information collected should not duplicate what can be obtained from other sources, such as description of the law or state implementation plans. Instead, NAEP is uniquely positioned to obtain ground- level information by surveying teachers and principals about policy implementation and challenges. This would not be designed nor suited to address legal compliance with federal policy, which is the role of program monitoring. Instead, it would provide information to improve the quality of policy and practice.

- Indeed, many national policies such as the Common Core are not federal at all. In this example, NAEP would track the implementation of standards in the Common-Core states, identifying changes in instructional content and emphasis compared with non-Common cores states. NAEP teacher surveys could further address the extent of staff training and understanding of the new standards and instructional challenges.

**Discussion**

The panel's review of the current background questionnaires concluded that they insufficiently incorporate questions about school and teacher responses to policies that could strengthen policy implementation and promote student achievement. Examples of policy-relevant issues that NAEP could but currently does not report on include characteristics of instruction in schools that made adequate yearly progress, the degree to which teacher evaluations incorporate student outcomes, or the nature and extent of coordination between school and after-school instruction.

This recommendation would reinforce NAGB (2003) guidance that identifies "informing educational policy" as a reason for collecting non-cognitive information. It would also support NCES commitments to convening "a policy/contextual issues panel when needed to identify policy/contextual issues that NAEP might address in the future, and to outline the relevant constructs and identify data needed to address these issues."

The panel recognizes that policy issues should be regularly refreshed as new policies emerge that build on or replace prior strategies. Our proposal aims for roughly a six-year issue cycle to give policies sufficient time to be implemented and effect improvements. The three policies suggested in the recommendations reflect the likely timeframe of implementation. The initial focus is on Common Core implementation, which is already underway in many states. Next a question set would be added on how schools evaluate their teachers. This would include questions on how evaluations of teachers take into consideration the outcomes of a teacher's students, as this relatively new policy takes

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

hold. The third suggestion of project-based and online learning reflects expectations that the role of technology in providing instruction will substantially increase.

### Recommendation 1c. Select clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics.

- Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement. NAEP unlike TIMSS or PISA currently fails to publish clearly defined, research-based theoretical frameworks that guide question selection. Accordingly, NAEP should make explicit and publically available the underlying theoretical frameworks for question selection. The Panel recognizes that the research basis for the theoretical justifications may be less than perfect and are sometimes subject to post-hoc rationalizations. Nonetheless, the objective syntheses of research across a variety of settings to form theoretical frameworks for clusters of variables significantly enhances the odds of collecting survey information that will accurately and usefully inform practice and policy.
- Background questions should fit together to portray different important aspects of a topic (e.g., the different dimensions of SES).

**Discussion**

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

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

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

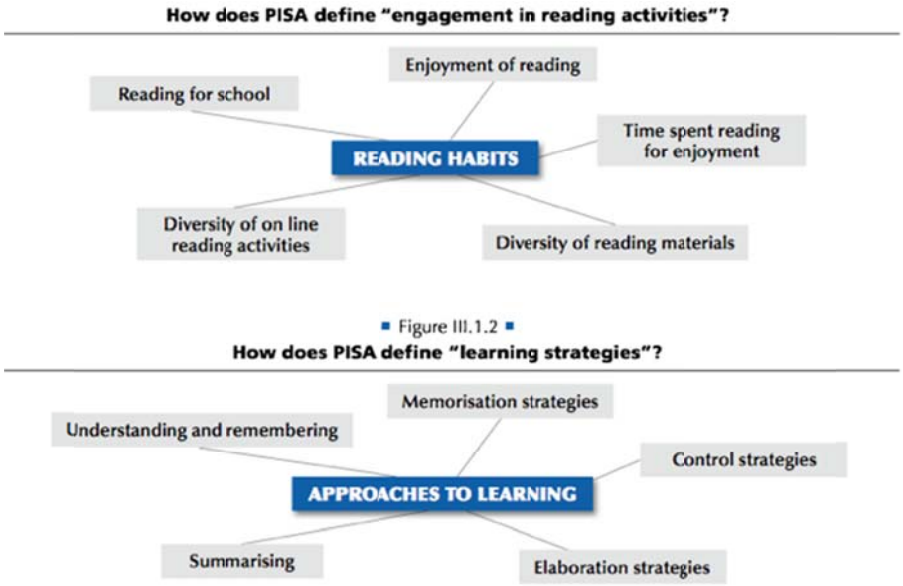
The panel's proposal would build-on the current literature reviews by:

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

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

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

**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 ai Practices

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

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

**Discussion**

The opportunity to assess progress on a background indicator over time is lost when NAEP no longer asks a prior question or disrupts the time series by asking essentially the same question in a somewhat different way. Because NAEP is the only major regular state-by-state assessment, question disruption results in a loss of important information to understand changes in a state educational context.

The panel examined the extent to which time series are available on the background question items for a sample of five broad questionnaire categories (Exhibit 4). The examination computed the percentage of questions asked under each category on the 2011 questionnaire for which there was also information for the same question for 2005 or earlier (at least a six-year trend).

- Between 70%-80% of the 2011 items about student characteristics or school demographics could be traced back to 2005 or earlier years.
- The three remaining categories that dealt with more judgmental measurement had much weaker time series availability. Only one-third of the 2011 questions asking about course offerings yielded at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

Some question categories become confusing to the user because of the considerable number of questions no longer asked. A case in point under the group of teacher factor questions is the "Preparation, Credentials and Experiences" category that contains over *400 questions of which more than 300 are no longer used,* with many replaced by just slightly different wording. . Moreover, what appears to be the exact same question maybe listed a number of times and in different places. Each instance of this all too common occurrence requires the user to search through and find all similar items and try and identify the one, if any, that is available and relevant.

Recognizing that at times changes in question wording may be necessary, the Panel recommends conducting *bridge studies* that would compare responses in the same year for prior and newly revised questions on a topic. NAEP's 2004 assessments in math and reading conducted a bridge study to compare results from students randomly assigned to the original and revised versions of the assessment (NCES, 2004). Bridge studies were also conducted for the new frameworks in reading and 12th grade math that were introduced in 2009. A similar process could be developed to bridge question changes in important areas of the background questionnaires.

Strategies for holding down the added expense of bridge studies should be carefully explored. Recognize that in conducting a bridge study on background questions, smaller representative samples of the kind used for polling may be adequate and preferable in minimizing error to having no bridge study at all. Also, it may be feasible to add background questions to other bridge studies such as those employed for the assessment.

**Exhibit 4. Percent of Background Questions Asked in 2011 Which Were Also Asked in 2005 or Earlier For a Sample of Question Categories**

| 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.

**Discussion**

*The panel supports the current NAGB and NCES reviews to identify the best way to measure SES variables within the confines of the NAEP questionnaire structure.*

This review is critically important given the well-documented limitations of the current school lunch proxy and that the first three State systems are piloting free school lunches for all students in very high-poverty school systems.

Limitations of the current school lunch measure include:

- The current measure divides the population only into two groups of free and reduced price school-lunch eligibles and ineligibles and is therefore insensitive to income differences above and below the income eligibility thresholds. SES is more accurately reflected by continuous measures. For example, this is consistent with studies showing student achievement results are sensitive to income levels over a broad income range.[4]

- School lunch eligibility is known to be underreported in secondary schools. Secondary students may not want the stigma of making known their families low-income and secondary students may not eat lunch at school. In fact, the grade 12 NAEP did not include school lunch for its 2009 report because of the problems of underreporting.

- The lengthy research literature on measuring SES consistently recommends multidimensional SES indices (Hauser & Warren, 1997) involving family resources, education and occupation. However, NAEP only reports the single student school lunch eligibility measure. NAEP's SES Project Progress Report (Noel-Miller and Hauser, August 2011) shows that a simple weighted average of indicators of home possessions and parental educational attainment does quite as well as independently estimated regression estimates in predicting math and reading achievement across grade-levels and race-ethnic subgroups.

- The 2010 *Healthy, Hunger-Free Kids* Act includes a "community eligibility" option, which would permit schools in high-poverty areas to provide free breakfast and lunch to all students without sending home individual paper applications for parents to submit income data. Three states have been selected for 2011-12 pilot eligibility (Illinois, Kentucky and Tennessee) and more states are scheduled to participate in successive years. Moreover, one urban school system Cleveland already counts 100 percent of its students as eligible for school lunch.

Consistent with the research literature, PISA incorporates questions for age 15 respondents to support an international multidimensional, socio-economic index. PISA's SES index elements consist of: occupational status of the father or mother, whichever is higher; the level of education of the father or mother, whichever is higher, converted into

[4]" In data from the Early Childhood Longitudinal Study (ECLS) measuring kindergarten students achievement on the ECLS reading achievement assessment, low-income students scored at about the 30th percentile, middle- income students scored at about the 45th percentile, and upper-income students scores at about the 70th percentile." (Lacour & Tissington, 2011)

years of schooling; and the index of home possessions, obtained by asking students whether they had a desk at which they studied at home, a room of their own, a quiet place to study, educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (e.g. paintings), books to help them with their school work, a dictionary, a dishwasher, a DVD player or VCR, three other country-specific items and the number of cellular phones, televisions, computers, cars and books at home.

The panel recommends that NAEP also move toward a multidimensional index for SES using current background questions. The panel further supports a long-run direction along the lines NCES is exploring of a two-pronged approach: (1) Creating an enhanced student background questionnaire with items that probe resources in the home, parents' education level, and parents' employment status; and (2) Using geocoding software to link students' home addresses to aggregate SES data available from the United States Bureau of the Census. The geocoding would reflect neighborhood and community factors that influence student performance.

In this context, the panel strongly supports the current NCES pilot to "generate SES information from the Census American Community Survey (ACS) data using school catchment zones, and which would make the collection of students' home address unnecessary for any assigned (non-choice) school."[5]

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

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

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

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

**Discussion**

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

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

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

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

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

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

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

**Discussion**

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

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

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

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

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

| Possible strategy | Score | | | | | |
|---|---|---|---|---|---|---|
| | *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 | O | O | O | O | O |
| b) I quickly read through the text twice. | O | O | O | O | O | O |
| c) After reading the text, I discuss its content with other people. | O | O | O | O | O | O |
| d) I underline important parts of the text. | O | O | O | O | O | O |
| e) I summarize the text in my own words. | O | O | O | O | O | O |
| f) I read the text aloud to another person. | O | O | O | O | O | O |

Source: OECD PISA 2009 Student Questionnaire

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

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

**Discussion**

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

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

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

Question a): In this class, about how much time do you spend on physical science?
Answers: None = 4%, Little = 9%, Some = 27%, A lot = 60%

Question b): About how much time in total do you spend with this class on science instruction in a typical week
Answers: Less than 1 hour = 1%, 1-2.9 hours = 4%, 3-4.9 hours = 60%, 5-6.9 hours = 25% , 7 hours or more = 9%

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

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

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

**Discussion**

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

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

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

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

| Regular In-School Lessons: Mathematics (Age 15, 2006) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No Time | | Less than 2 hr | | 2-4 hr | | 4-6 hr | | 6+ hr | |
| | Math Score | % | Math Score | % | Math Score | % | Math Score | % | Math Score | % |
| Japan | | 0 | 444 | 8 | 491 | 35 | 551 | 42 | 572 | 14 |
| Korea | 416 | 1 | 451 | 3 | 529 | 21 | 561 | 58 | 576 | 16 |
| U.S. | 429 | 5 | 430 | 23 | 465 | 20 | 511 | 38 | 490 | 15 |

| Out-of-School: Mathematics (Age 15, 2006) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No Time | | Less than 2 hr | | 2-4 hr | | 4-6 hr | | 6+ hr | |
| | Math Score | % | Math Score | % | Math Score | % | Math Score | % | Math Score | % |
| Japan | 480 | 24 | 517 | 32 | 551 | 23 | 575 | 13 | 593 | 8 |
| Korea | 520 | 23 | 541 | 14 | 573 | 33 | 579 | 17 | 594 | 13 |
| U.S. | 512 | 79 | 478 | 11 | 464 | 5 | 456 | 3 | 433 | 2 |

Source: NAEP Data Explorer

- Almost 30 % of U.S. age-15 students spend less than 2 hr. in a math class per week compared with less than 10% of Japanese students and 5 percent of Korean students. Moreover, those students with the lowest scores receive the least math instructional help in-school.
- Eighty percent of U.S. age-15 students spend no time learning math in formal afterschool instruction compared with only a quarter of Japanese or Korean students.

It would be valuable for individual states to be able to compare their students' math instructional time in-school and out-of-school with those of the Asian performers, but NAEP collects very little information about learning time. For example, it asks only about number of days a week in math instruction and not about number of hours and there is no information about time spent in math or other subjects after school. Had NAEP spelled out a basic theoretical framework identifying clusters of questions about time measurement (recommendation 1c) NAEP might have been more likely to align its questions to compare states with the interesting PISA national results.

### Recommendation 2f. Build on current NCES cognitive interview techniques by using cognitive laboratories, such as small standing panels, to field test questions to establish their validity and reliability.

**Discussion**

NCES conducted cognitive laboratory investigations of the responses of students and teachers to questions from the 1996 and 1998 background questionnaires (Levine, Huberman, and Buckner, 2002). Cognitive interviews are an approach "to assess how respondents comprehend survey items and what strategies they use to devise answers."

The 1990's studies identified a number of general types of item problems:

- Behavioral frequency discrepancies. These items ask about how frequently a student or teacher engages in specific activities or practices. The average level of agreement between fourth grade students and their teachers on items that used a four-point rating scale was only 38 percent; for eighth grade students and their teachers, the level of agreement was still only 51 percent. Guessing would yield agreement of 25 percent.

- Time frame discrepancies. Differences between teachers and students in the period over which behavior is estimated were common. Teachers would generally think about the current year and students about a very immediate near-term period. Also, when teachers were asked about the frequency of a behavior such as how often a particular science topic was taught, teacher's responses applied to only when science is taught. Thus the response option, "Almost every day," was explicitly interpreted as "Almost every day that science is taught."

- Comprehension discrepancies. Different respondents may interpret items differently. When teachers responded to a question about frequency of a behavior with "students in your class," some teachers would answer about the typical student and others would respond if any one student exhibited that behavior.

- List format discrepancies: Loss of context. On a long list of items, students or teachers might forget the context in which the question was asked. A student might interpret a question about school behavior such as reading and respond with their general reading behavior in or out of school.

NAEP also conducted a cognitive laboratory analysis of the Responses of fourth and eighth graders to questionnaire items and parental assessment (Levine, et.al. 2001).

The Panel believes that cognitive lab interviews are able to detect and prevent many survey design problems. Hence, it recommends that NAEP use cognitive labs more extensively with an accompanying small panel of adult (teacher/principal) and child respondents to validate and improve background questions. In addition, small-scale pilot studies should be used to assess the feasibility, reliability, and external validity of survey items. We recognize that this may increase costs but it would help make the overall NAEP a better source of information.

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

Limitations of time and concerns over data burden severely constrain the depth of the student background questions. As a result, NAEP often lacks the richness in its background questions that would enable it to replicate the constructs such as those PISA creates from lengthy multiple items around different aspects of research-based

frameworks. To further extend the richness of its data sets, PISA also enhances its basic student and principal questionnaires with optional supplemental questionnaires. NAEP should consider expanding the depth of its questions through a variety of strategies including spiral sampling (currently already under consideration by NAEP), expanded questionnaire time and rotating background questions across samples.

### Recommendation 3a. Support NCES's exploration of a spiral sample methodology to expand the scope of background questions, while assessing the possible loss in the representativeness of disaggregated data.

- Spiraling questions so that no student takes the full set of background questions would allow NAEP to expand the scope of its background items. The current 10-minute limit for the student questionnaire severely constrains the current scope and depth of the student questionnaires. By contrast PISA is able to support richer construct development with its 30-minute student questionnaire.
- In assessing questionnaire spiraling, it is important to consider how it would reduce NAEP's ability to provide statistically-accurate state-by-state or urban district information, especially if broken out for different student sub-groups.

**Discussion**

The Panel supports exploring the proposed spiral sampling of questionnaire items in order to implement improvements in student questionnaire scope and depth. As noted, one such improvement would be to enable greater in-depth questioning through clusters of items that measure different aspects of research-based topic frameworks.

However, the Panel urges NCES to quantify how item spiraling will reduce NAEP's ability to disaggregate state or urban district responses for specific population groups. For example, will background questions be available in sufficient sample size for all population groups for which cognitive student achievement data are reported?

Illustrating this point is an analysis of whether a state has changed its grade-8 access of students to a course in algebra during the two-year interval between successive NAEP assessments. It turns out that Alabama raised the percentage of its students in schools offering grade-8 algebra by 6 percentage points during the two years and Arizona decreased it by 5-percentage points. These changes are sizeable for two years, yet neither change was statistically significant. A spiral sampling approach would further reduce the odds of obtaining statistical significance.

### Recommendation 3b. Consider other item-sampling reforms to obtain the needed questionnaire time including lengthening the student survey; establishing a 4-year interval between administration of some background questions; and pooling item responses across survey administrations.

- The ten-minute target length for responses to the student questionnaire does not seem grounded in empirical experience and NAEP would do well to consider the

merits and feasibility of a lengthier questionnaire. TIMSS grade 4 and 8 student questionnaires are targeted for 30 minutes at each grade and do not appear to suffer from high non-response rates.[6]

- Some background questions with slow-moving trends may be adequately monitored through repeating survey questions at four-year intervals.
- Pooling item responses across successive surveys may also be a permissible strategy to expand the sample provided that response changes are sufficiently slow moving.

**Discussion**

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

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

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

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

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

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

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

**Discussion**

Special reports would provide access to the background questions in manageable-size documents that don't overwhelm the reader. An example of a NAEP special report is *The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8,* which is Part II of the National Indian Education Study of 2009. Part II complements the Part I report on NAEP assessment results for American Indian students by providing information about students, their families and communities, and their school experiences.

More generally TIMSS and PISA illustrate two approaches to developing topics for the special reports. TIMSS includes individual chapters organized around different questionnaire topics:

- Students' Backgrounds and Attitudes Towards Science
- The Science Curriculum
- Teachers of Science
- Classroom Characteristics and Instruction
- School Contexts for Science Learning and Instruction

The 2009 PISA has published a series of special reports, synthesizing lessons learned to improve academic achievement:

- *Overcoming Social Background*: *Equity in Learning Opportunities and Outcomes* looks at how successful education systems moderate the impact of social background and immigrant status on student and school performance.
- *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 | Eligible for National School Lunch | Grade 8 Students Absent 5 or more days last month | Grade 8 Students in Algebra | Grade 8 Students 5 or more Hours of Math Per Week | Grade 8 Students 1 Hour or More Math Homework | Grade 8 Does Math At An Afterschool or Tutoring Program | Grade 8 Entered Math Through Alternative Certification | Grade 8 Teacher Has Math Major/ Minor/ Special Emphasis | Grade 8 Full-time Math Specialist At School | Grade 8 Assigned To Math By Ability | Grade 8 26+ Students in Math Class | Grade 8 Computers Available to Teachers and 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 | 29 | 61 | 27 | 30 | 42 | 57 | 58 | 53 | 52 | 89 |
| Baltimore City | 261 | 85 | 9 | 46 | 93 | 41 | 38 | 38 | 79 | 53 | 85 | 37 | 71 |
| 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 | 59 | 65 | 29 | 39 | 57 | 66 | 40 | 53 | 20 | 86 |
| Fresno | 256 | 88 | 10 | 51 | 32 | 11 | 26 | 6 | 37 | 23 | 91 | 75 | 59 |
| Hillsborough County (FL) | 282 | 54 | 9 | 67 | 20 | 13 | 22 | 40 | 35 | 29 | 95 | 3 | 86 |
| Houston | 279 | 76 | 6 | 29 | 63 | 26 | 37 | 56 | 63 | 25 | 84 | 58 | 68 |
| Jefferson County (KY) | 274 | 60 | 7 | 48 | 66 | 14 | 20 | 21 | 54 | 36 | 77 | 80 | 80 |
| Los Angeles | 261 | 82 | 6 | 67 | 44 | 40 | 27 | 39 | 67 | 37 | 75 | 52 | 74 |
| Miami-Dade | 272 | 72 | 5 | 35 | 43 | 47 | 25 | 38 | 72 | 25 | 90 | 13 | 88 |
| Milwaukee | 254 | 81 | 13 | 30 | 78 | 43 | 31 | 37 | 74 | 82 | 28 | 86 | 78 |
| New York City | 272 | 87 | 10 | 28 | 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.

## Discussion

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

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

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

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

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

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

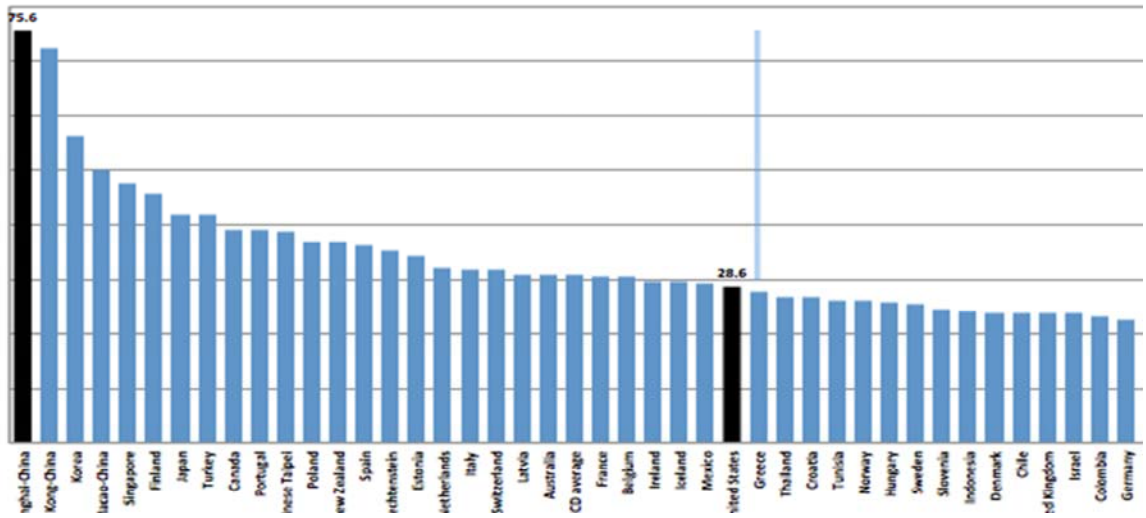
**Discussion**

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

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

**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

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

**Discussion**

The proposed special reports in 5a are designed to mine the unexploited investment in the largely unanalyzed background questions. These reports might be modeled on the special publication of background data from the National Indian Education Study of 2009, *Part II: The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*, cited in Recommendation 4a.

The special publications would describe:

- In-school learning opportunities and other educational experiences focusing on data already collected on curriculum, instruction, teachers and other school resources including technology.
- Out-of-school learning opportunities and other educational experiences including after-school and at home.
- The background characteristics of high performing states and school systems contrasted with low-performers. This benchmarking study would be purely descriptive, serving to guide follow-on research to improve understanding of the factors differentiating high and low performing states and districts.

These would be three synthesis reports, drawing on data from NAEP assessments across the curriculum and, where possible, trends over time.

## Recommendation 5b. Initiate a set of activities to build clusters of core and second-tier questions around high-priority topics for the 2015 NAEP administration.

**Discussion**

Given the long lead times for questionnaire development, this effort needs to begin immediately in order to affect the 2015 NAEP reading and mathematics administration. The revised questionnaires would refocus the background questions to identify an expanded first-tier core and second -tier set of rotated question clusters, including a rotated set of policy issues (Strategies 1 and 2, Exhibit 12). As NAEP redefines its question sets, NAEP would improve measures through published evaluations of their validity, reliability and consistency with each major assessment (Strategy 3, Exhibit 12). To find the questionnaire time to develop in-depth question sets, Strategy 4 prepares a NAEP analysis and report on a combination of sampling reforms addressing spiraling questions and extra question time.

| Exhibit 12. Longer-term Background Question Activities / Products | | |
|---|---|---|
| Strategy | 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)<br>• Identify 2nd tier rotated questions<br>• 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.<br>• 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<br>• Implement quality review procedures for reliability and consistency of questions.<br>• Launch a cognitive laboratory capability with possibly an available small standing supplementary panel. |
| *4. Report on item sampling reforms to incorporate extended question sets and topics including eliminating duplicative and low-priority items* | *3a, 3b* | • Report on a strategy to add questions for cluster analyses and policy issues through questionnaire spiraling, alternating questions across assessment administrations, adding extra questionnaire time and eliminating low-priority items, |

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

**Discussion**

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

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

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

**Discussion**

To promote implementation of the background question recommendations and make sure change occurs, the panel suggests that NAGB establish a separate standing committee to review all background questions and oversee a multi-year development plan to improve the questions and their use. Currently, the Board's responsibilities for the background questions are divided between the Assessment Development and the Reporting and Dissemination Committees. A unified standing committee should regularly monitor and report on implementation of the panel's recommendations by NCES and Governing Board staff.

The panel further recommends that a review be conducted of the resources needed in terms of time, money and personnel to implement the recommendations in this report. One approach to the problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate lower-priority activities, such as the duplicative collection of racial data and the disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

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

# REFERENCES


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

EDUCATION WEEK
Published Online: March 13, 2012
Published in Print: March 14, 2012, as **NAEP Board Considering Gathering Additional Data**

# Policymakers Weigh Gathering More Data for NAEP

## Goal is to improve understanding of performance

**By Erik W. Robelen**

Back to Story

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

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

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

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

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

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

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

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

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

### Causation vs. Correlation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

But some observers see reason for concern.

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

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

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



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

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

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

Vol. 31, Issue 24, Pages 20-21