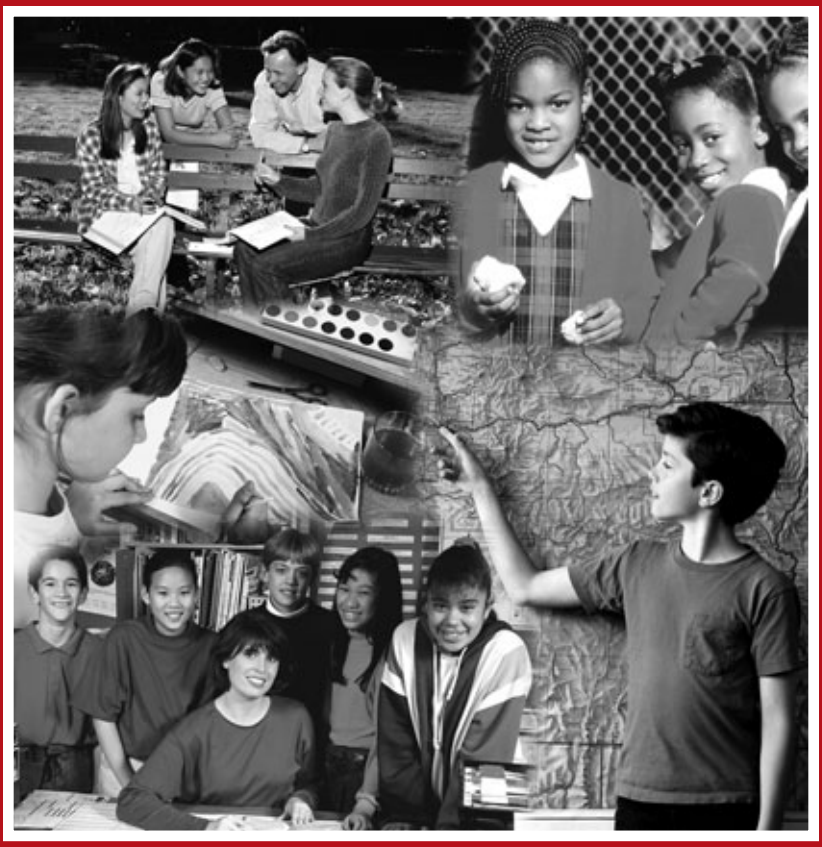


Background Information Framework for the National Assessment of Educational Progress

National Assessment Governing Board
U.S. Department of Education

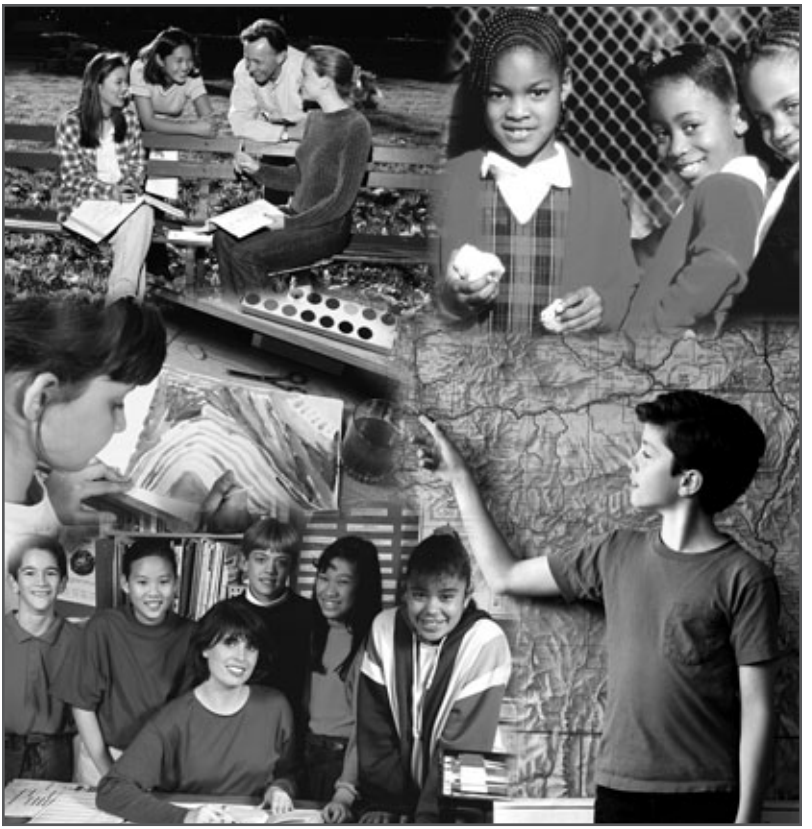
What Is NAEP?

The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what American students know and can do in various academic subjects. It is a congressionally mandated project of the U.S. Department of Education's National Center for Education Statistics. NAEP surveys have been conducted on a national sample basis since 1969 in reading, mathematics, writing, science, and other elementary and secondary school subjects. State-level assessments have been conducted since 1990.

The National Assessment Governing Board

The National Assessment Governing Board (NAGB) was created by Congress in 1988 to formulate policy for NAEP. Among the Board's responsibilities are determining the content of NAEP and designing the assessment methodology. The Board has final authority on the appropriateness of all test items.

Background Information Framework for the National Assessment of Educational Progress



National Assessment Governing Board
Adopted August 1, 2003

For further information, contact the
National Assessment Governing Board:

800 North Capitol Street, N.W.
Suite 825
Washington, DC 20002
www.nagb.org

Table of Contents

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 the 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 in the 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 information framework will refocus the collection of non-cognitive variables on NAEP's primary mission: to provide 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 when 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 noncognitive 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 for 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 sources 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 about 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.

The National Center for Education Statistics (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 burden they impose on respondents. 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 in 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) are 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. However, 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 educational research. The NCES program of secondary analysis grants for researchers to analyze NAEP data should continue. Education 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.

Chapter One

Introduction

The National Assessment of Educational Progress is the only continuous long-term measure of student achievement in U.S. elementary and secondary schools. The primary purpose of NAEP is to report to the American public on academic achievement and its change over time.

Nature and Purpose of NAEP

The NAEP survey consists of two major components: (1) academic assessments, which measure the achievement of students on a broad range of content, and (2) non-cognitive survey questions, which collect descriptive information from students, teachers, and school administrators about demographic characteristics and the educational process. Since 1969 NAEP has measured achievement in most areas of the school curriculum, including mathematics, reading, writing, science, U.S. history, world geography, civics, economics, foreign language, computer science, and the arts. The content of NAEP assessments is determined through a framework development process that articulates the content parameters for each area and recommends subject-specific non-cognitive areas for data collection and reporting.

NAEP's purpose is to report to the public on the status of academic achievement in the United States. The assessment does not report results for individual students, but only for groups with large, representative samples, for example, students from rural schools, from various ethnic groups, or from participating states, and, on a trial basis, large urban school districts. It must be able to provide data for fair and accurate comparisons among the states, districts, and subgroups on which it reports. The background data play a crucial role in ensuring the fair comparisons—over time and between student groups—that are at the heart of NAEP's mission and value.

Nature and Purpose of Background Data

The most recent NAEP reauthorization (P.L. 107-110) gives the National Assessment Governing Board "final authority" to approve "all cognitive and non-cognitive assessment items." This framework deals with the non-cognitive side of the Board's responsibility, including the items that identify students in NAEP's required reporting categories and the other information that provides a context for results and tracks factors associated with academic achievement.

The term "non-cognitive," as used in the law, seems more inclusive than "background questions," the phrase usually used by NAEP in the past for items designed to collect non-academic information. However, non-cognitive is also less readily understandable than background information, and so the two terms are used interchangeably in this document. Both will refer to all of the information beyond the academic assessment that NAEP uses to make its academic results more meaningful to the public.

When NAEP began, the collection of non-cognitive data was limited to the demographic categories of gender and race/ethnicity, and to two measures of home environment or socio-economic status (SES)—level of parents' education and literacy materials in the home. In addition, an index was constructed, based on data from the U.S. Census and a brief school questionnaire, to report achievement results for schools in three types of communities—disadvantaged urban, advantaged urban, and rural.

During the 1980s the use of non-cognitive questions was greatly expanded to accommodate several functions within NAEP (Reckase, 2002). First, they were used to define a more extensive array of subgroups of the student population for reporting purposes. For example, NAEP results are now reported by gender, race/ethnicity, parents' highest level of education, type of school, participation in Title I, and eligibility for free/reduced-price lunch.

A second reason for collecting non-cognitive information is to inform educational policy by describing the contexts for learning, sometimes called opportunities to learn (Mullis, 2002). Broadly, this involves the content specified in the curriculum, whether and how that content actually is taught, students' propensity to learn, as well as home and school factors that can enhance learning.

In conjunction with the descriptions of students, background information about educational settings and experiences can reveal striking differences in how important aspects of education and educational resources are distributed among different groups. For example, do disadvantaged minority students have less access to science laboratory equipment than more advantaged groups? Do girls take less rigorous mathematics courses than boys? The data on course taking has been used widely to discuss the patterns and trends in mathematics achievement. Having this information as part of NAEP has added to the public impact of assessment results.

A third function of the non-cognitive questions has been to support research into factors that may be related to student achievement. The background questions serving this function have sought information not only on curriculum, teaching methods, and discipline in the school, but also on educational activities at home. For example, *The 1998 NAEP Reading Report Card* (Donahue, Voelkl, Campbell, and Mazzeo, 1999) reports on television viewing, daily reading habits, classroom reading and writing assignments, and discussion of schoolwork at home. While secondary researchers have used NAEP to investigate relationships to student achievement, the basic design of the assessment as a cross-sectional survey without longitudinal data limits its usefulness. Research has been most productive when NAEP is combined with other data sources and in descriptive studies that track changes over time.

Non-cognitive data are also necessary to support certain technical functions of NAEP. For example, some non-cognitive information is used to evaluate the potential for bias resulting from non-participation. That is, did the students absent or refusing to participate in the assessment differ in such significant ways from those who did take part that results were changed? Non-cognitive variables also play an important role in NAEP's sampling and weighting procedures, and sometimes in checking the validity of results. Many of these variables are taken from other data sources, such as the Common Core of Data (CCD), but some come from the administration roster collected from schools prior to testing, the records kept by test administrators, and student questionnaires.

Finally, NAEP non-cognitive questions have been used in the technical process for preparing estimates of student proficiency distributions on the cognitive component of the assessment. But their role in this process is limited to facilitating data analysis. Only the

student responses to cognitive questions are used to determine achievement results. Background variables are used to define the groups for which cognitive data are reported.

Once test results for a group are determined, the NAEP analytic process makes use of background data available to prepare a second data set—identical in its group scores to the first—that can be handled by much simpler computer programs to prepare other analyses and reports. However, only the background factors to be reported on are needed for this analytical work, called conditioning. The precision of NAEP results is not reduced if background items not used for reporting are eliminated.

This background information framework will focus the collection of non-cognitive information on NAEP's primary mission: providing, as the law stipulates, "a fair and accurate measurement of student academic achievement and reporting trends in such achievement" over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results.

Development of NAEP Background Information Framework

In the Policy Statement on Redesigning the National Assessment of Educational Progress (adopted in August 1996), the Governing Board sought to improve the validity of background information on NAEP, increase the efficiency with which it is collected, and reduce the number of background questions in the assessment itself. The statement was based on the report of a Design/Feasibility Team (Forsyth et al., 1996), headed by Robert Forsyth, which recommended a design that would rotate the collection of non-cognitive data into distinct modules administered over several assessment cycles. NAGB endorsed implementing that recommendation through a system of standard and comprehensive NAEP assessments that would be administered on a cyclical basis (NAGB, 1996).

Standard assessments would ask a short, essential core of background questions associated with a content area. Periodically, a *comprehensive* assessment would employ a much fuller complement of such questions to probe that area more extensively. Although some efforts have been made to reduce the background questionnaires and

streamline data collection, the full impact of the NAGB policy has not yet been realized.

In early 2002, the No Child Left Behind Act transferred final authority over the non-cognitive questions from the National Center for Education Statistics to the National Assessment Governing Board. The Board adopted a new policy governing the development and selection of non-cognitive questions in May 2002, and initiated a process to prepare a general framework for non-cognitive data (NAGB, 2002). This framework would define the scope of NAEP background questionnaires, the priorities for collecting non-cognitive information, and the criteria for reporting non-cognitive data in NAEP. (See Appendix for full text of the policy.)

The Board created an Ad Hoc Committee on Background Questions and conducted an all-day workshop on the NAEP non-cognitive questions on September 24, 2002. Six consultants prepared and presented papers at the meeting that was attended by Board members, academic researchers, representatives of the national teacher organizations and other education groups, and NAEP contractors and staff. The six consultants are identified on the title page as contributors to this document.

In the months after the workshop, a draft framework was prepared. It was refined at several meetings of the Ad Hoc Committee, posted for public comment on the Internet, and was the subject of a public forum in Washington, D.C., on May 1, 2003. Altogether, oral comment and written testimony were received from 22 persons and organizations, many with differing perspectives and views. The Ad Hoc Committee and the Board carefully considered these comments, and the draft framework was revised at a Committee meeting on June 25. The Committee heard additional comment and made final revisions on July 31. The background information framework was reviewed by the full Governing Board several times during the course of its development. The Board adopted it unanimously on August 1, 2003.

Although this framework is not a consensus document, it does encompass the thinking of a wide range of researchers, policy analysts, and users of NAEP data. It is the product of discussion and deliberation by the Governing Board, and incorporates Board decisions on the nature and focus of the background information to be included in NAEP. The framework will become operative in the 2006 National Assessment.

Requirements of NAEP Statute

The No Child Left Behind Act of 2001 (P.L. 107-110) requires NAEP to collect information on gender, race/ethnicity, socioeconomic status, disability, and limited English proficiency. It must report test data on these groups, whenever feasible, that is cross-tabulated, compared, and reported according to the categories required.

The law also requires NAEP to collect only information that is directly related to academic achievement and to the presentation of such information in a fair and accurate manner. This means that NAEP needs to concentrate on variables that are known to be related to achievement rather than on theoretical constructs. The statute requires the Governing Board to ensure that all NAEP questions are "free from racial, cultural, gender, or regional bias"—a provision from previous law. But it adds new language that questions must be "secular, neutral, and non-ideological" and must not "evaluate or assess personal or family beliefs and attitudes."

In their report on the bill, the House-Senate conference committee that negotiated its final form says the law "does not preclude the use of non-intrusive, non-cognitive questions, approved by the National Assessment Governing Board, whose direct relationship to academic achievement has been demonstrated and is being studied as part of [NAEP] for the purposes of improving such achievement." The report language is not binding, but is intended to guide implementation of the law. ***This framework emphasizes that the legal prohibitions must be followed in preparing background questions and collecting any other non-cognitive data for NAEP.***

In addition, the law makes it clear that NAEP may not disclose any personally identifiable information or maintain any system of records that contains such data. These restrictions are not new. They have dictated careful procedures in the past, which must be continued.

Purpose and Rationale of Background Information Framework

The purpose of the framework on background information is similar to that of NAEP's content area frameworks: to guide the development of the assessment. The content frameworks have described the topics to be tested by NAEP and provided an outline of the assessment for each subject area. Purposefully, the frameworks

attempt to be independent of a particular pedagogy. They do not specify what educational resources or processes should be used, but rather describe important achievement results. They provide states, schools, policymakers, and the public with a logical outline of the approach used in constructing the assessment.

The framework for NAEP background data will specify the parameters of the assessment from a reporting perspective. The background information that NAEP uses in its reports helps to give context and meaning to the cognitive results. It must be collected in a systematic way from the NAEP testing samples either through questionnaires or from other reliable sources, such as school records and other federal surveys. Collecting descriptive information from a variety of sources can improve the quality of the data obtained and increase efficiency while reducing the burden on respondents.

The Governing Board adopted a Policy Statement on the Collection of Reporting of Background Data on May 18, 2002 (NAGB, 2002). The statement is incorporated into this framework and attached in the Appendix.

Chapter Two

Priorities and Criteria for Collecting and Reporting Non-Cognitive Data on NAEP

This chapter presents priorities for collecting and reporting non-cognitive information on NAEP. It also includes the criteria for selecting particular topics and questions, and for determining the frequency with which various data elements are reported. A final section presents criteria for identifying and selecting background data sources.

Priorities for Non-Cognitive Information

The following priorities for collecting and reporting non-cognitive information are based on legal requirements, the purposes of NAEP, and the strengths and limitations of the assessment. They should be followed in planning 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 and course requirements, student mobility, school safety and discipline, teacher-related factors such as teacher demographics, preparation, credentials, and experience, and other factors related to

students, schools, and educationally relevant variables beyond the school. Although these types of non-cognitive variables are of interest, they must be limited so that they meet the needs of NAEP reporting. In all cases, they must be clearly related to academic achievement or to the fair presentation of achievement results.

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

With regard to the points above, Walberg (2002) makes a suggestion that might be a workable solution to consider. Just as students in the NAEP samples do not respond to all the questions, say, in reading, but only to a portion of those for any one grade-level, so too, the non-cognitive questions could be rotated through different (smaller) NAEP samples. These non-cognitive "testlets" could be rotated through the NAEP samples by class or school, with students receiving different, expanded "testlets" in addition to a core set of background questions.

Criteria for Selecting Non-Cognitive Topics and Questions

The Advisory Council on Education Statistics (ACES), a technical panel that used to advise the National Center for Education Statistics, spent a considerable amount of effort on the issue of NAEP non-cognitive questions. Its guidelines, adopted in May 1997, include a set of key questions that should be utilized in selecting topics and questions for NAEP background data collection. The questions with commentary are summarized below:

- ***Does the current or proposed non-cognitive variable relate to the primary purpose of NAEP and how?*** The primary purpose

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

- ***Do the current or proposed non-cognitive variables meet professional standards for reliability and validity?*** The NAEP legislation requires that the assessment "use widely accepted professional testing standards (P.L.107-110, Sec. 411 (b) (5)." This requirement applies equally to non-cognitive and academic variables. It is already known that some non-cognitive variables in NAEP have weak reliability (e.g., data from 4th graders on their parents' highest level of education and the self-reports of teachers on classroom practice). If more reliable sources of such data cannot be found, these variables should be deleted from the assessment.
- ***How stable is the non-cognitive variable from period to period?*** If a variable shows little change from year to year, it should be reviewed to determine whether it should be deleted or used on a periodic basis rather than in every assessment.
- ***Is the proposed or current non-cognitive variable of timely interest?*** The educational environment changes from time to time, and consequently public interest in particular variables will change as well. It would serve NAEP well to review the set of non-cognitive variables periodically with this criterion in mind, deleting those that do not meet the test of timeliness and substituting others of current interest.
- ***If new questions are added, have others been deleted in order to limit the burden and expense of NAEP's 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?*** For example, a question that asks teachers whether they adhere to national standards in mathematics or another subject is conclusionary and hard to interpret, since many teachers are

apt to say yes, regardless of what they do. It would be better to ask about specific behaviors, such as homework assignments or computer use. Caution is advisable in this area too because self-reports are often unreliable.

The Board believes three other important criteria must also be considered:

- ***Will the topic or question meet the test of broad public acceptability and not be viewed as intrusive or prying?*** NAEP's non-cognitive questions are not kept secure and must readily be available to anyone requesting a copy. Under Board policy, all questions asked are to be posted on the Internet. Possible objections should be considered in deciding whether or not to ask them.
- ***Does the topic or question deal with a factor in which trends over time are of importance?*** If trends are deemed important and the factor is related to achievement, the topic or question should be included periodically on a four-year or eight-year cycle, rather than being part of the background questionnaire each year. For example, measuring television watching in every NAEP assessment is not necessary. But it can be valuable to measure TV-watching every four or eight years to find out whether or not it is increasing.
- ***Will the information obtained be of value in understanding academic performance and taking steps to improve it?*** This is a fundamental issue to be addressed in evaluating all background questions proposed for NAEP.

Criteria for Selecting Data Sources

NAEP has collected non-cognitive information from students, teachers, and schools, using NAEP background questionnaires. There are also administration rosters, completed by test administrators at the school level in advance of testing to determine characteristics of the testing samples. The Common Core of Data (CCD) is used to identify characteristics of schools (e.g., Title I funding), and schools also complete a questionnaire on special needs students (e.g., students with disabilities and limited English proficiency).

However, the collection of non-cognitive data may be shifted among these sources or to new sources in order to improve reliability, increase efficiency, or reduce burden. State management information systems and data collected for school report cards, as required by the No Child Left Behind Act, may become very useful for NAEP. ***Whenever possible, NAEP should use information from school records and other reliable data collections about students and schools in order to improve the validity of the information collected and limit the background questionnaires in NAEP itself.***

In exploring the utility of different data sources, the following criteria should be considered:

- ***Validity—***Is the data obtained from the new source a valid indicator of what it purports to measure?
- ***Reliability—***Is the data from the new source at least as reliable and consistent as that from the source previously used?
- ***Universality—***Can the required data be collected by this method for all (or almost all) of the students and schools participating in NAEP and will it support valid comparisons over time?
- ***Currency—***Will data obtained from a new data source be current enough to relate clearly to the assessment being conducted? If data from the census or some other source is several years old it may not accurately describe school or neighborhood conditions at the time of testing.
- ***Respondent burden—***Will the new source(s) reduce the burden on students, teachers, and schools in filling out NAEP questionnaires? Will the total amount of respondent burden be decreased?
- ***Logistics—***Will the alternative source(s) be logistically possible, or will there be more logistical problems than with the previous data source? Logistics includes such considerations as cost, time, administrative personnel resources, and steps needed to ensure accurate coding and data analysis.
- ***Efficiency and cost-effectiveness—***How efficient will the new data source be in comparison to the previous one? For example, it may be more efficient to collect data from a state management information system about the state's schools, teachers,

or students, rather than obtaining it from the test samples directly, but efficiency and cost-effectiveness should be determined before a change is made.

- ***Timeliness of NAEP reporting*—**How will a change in data sources affect the speed with which NAEP can be reported? Some changes will speed operations, but those that slow down NAEP reporting are not desirable.

Chapter Three

Topics and Types of 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.

Chapter Four

Non-Cognitive Data Sources and Collection

This chapter discusses the sources of non-cognitive information for NAEP and the reporting categories that the information describes. It includes a NAEP Background Information Matrix, organized by priorities, which summarizes the types of descriptive information NAEP collects, reporting units, and data sources.

NAEP Student, Teacher, and School Samples

The NAEP student samples vary in size and purpose. Their overall total has become very large. Starting in 2003, national NAEP samples are specified at the state and jurisdictional levels, with approximately 3,000 students per subject and grade (4 and 8 only) for each of the 50 states, plus the District of Columbia, and Department of Defense domestic and overseas schools. Puerto Rico (in mathematics only) has a sample of about 3,000. In addition, the ten Trial Urban District Assessment (TUDA) districts have sample sizes of the order of 3,000 to 5,000 each. There also are a nationally representative sample of charter schools, totaling about 3,000 students, and national private school samples totaling about 12,000 in each grade.

At grade 4, therefore, the total NAEP sample approximates 436,000 students. The grade 8 sample is about the same at 432,000 (excepting charter schools). The grade 12 sample is for a pilot test and includes only about 6,000 students (Rust, 2002). In most future years the 12th grade samples are expected to have about 30,000–40,000 students assessed in national samples only for three subjects.

In addition to the nearly one million students tested, about 80,000 teachers of those students complete teacher questionnaires and some 13,000 schools complete school questionnaires. Several thousand school districts also supply data for the assessment. The sampling and weighting procedures in NAEP use data from the CCD files as well as census data and school-level achievement data from the

states for improving NAEP stratification procedures. The NAEP non-cognitive data collection effort is enormous and challenging.

Other Data Sources

The Governing Board is strongly committed to improving the quality of background information while reducing respondent burden and the complexity of data collection and analysis. The self-report questionnaires given to students, teachers, and schools are sometimes burdensome to fill out, labor-intensive to collate and analyze, and subject to concerns about reliability. All questionnaires should be scrutinized to replace as many items as possible with data from centralized records, gathered by test administrators, or, ideally, from computerized data files.

The data available from federal, state, district, and school records should be carefully explored. With implementation of the school report card requirements of the No Child Left Behind Act, much more information should be available soon in standardized computer formats. Barton (2002) has suggested some specific sources of data collected outside of NAEP that should be considered to improve NAEP reporting. These include the U.S. Census, Quality Education Data, Inc. (QED), and the Common Core of Data (CCD) and School and Staffing Survey (SASS), both compiled by the National Center for Education Statistics.

This approach of utilizing more data from outside specific NAEP data collections has been elaborated on extensively in the most recent evaluation of NAEP by the National Academy of Sciences (Pellegrino, Jones, and Mitchell, 1999). The panel proposed "a coordinated system of indicators for assessing educational progress, housed within NCES and including NAEP and other currently discrete, large-scale data collections" (p. 34). Figure 1 is reprinted from the NAS report to show the extent of these data collections on students, teachers, and schools, and to indicate what might be obtained from these other sources. To use them for NAEP would greatly lessen the burden on the assessment itself. Merged data sets could be made available, some to the general public, and more to researchers in restricted data files.

For many years state-level NAEP reports have included appropriate collateral data that provide a context for interpreting NAEP results; see for example the *NAEP 1996 Mathematics: Report Card for the Nation and the States* (Reese et al., 1997). These state

Figure 1. Overview of Current NCES Data Collections

Data and Design Elements	NAEP	NELS	ELS	ECLS	TIMSS	CCD	PSUS	SASS	NHES
Data Elements									
Student achievement	x	x	x	x	x				
Student background characteristics	x	x	x	x	x	x	x	x	x
Home and community support for learning	x	x		x	x				x
Standards and curricula					x				
Instructional practices and learning resources	x	x		x	x			x	
School organization/governance					x			x	
Teacher education and professional development	x				x	x	x	x	
Financial resources					x	x	x	x	
School climate	x	x		x	x			x	x
Design Elements									
Type of design (CS=cross-sectional; L=longitudinal)	CS, L	L	L	L	CS	L	L	CS,L	CS
Periodicity (TBD=to be determined)	2, 4, or 6 yrs	2–6 yrs	TBD	TBD	TBD	Annual	Biennial	2–5 yrs	2–3 yrs
Unit of observation (S=student; T=teacher; A=administrator; P=parent; SC=schools; D=district; ST=states; H=households)	S,T,A	S,T,A	S,A,P	S,T,A,P	S,T,A,P	SC,D,ST	SC	T,A,SC	H
Data collection method (S=survey; R=record analysis; I=interview; V=video; C=case study; O=other)	S	S,R	S,O	S,O	S,R,V,C	S,R	S	S	I
Population of inference (N=national; S=state; G=demographic group)	N,S,G	N,G	N,G	N,G	N				

NELS: National Education Longitudinal Study of 1988
ELS: Educational Longitudinal Study of 2002
ECLS: Early Childhood Longitudinal Study
TIMSS: Third International Mathematics and Science Study
CCD: Common Core of Data
PSUS: Private School Universe Survey
SASS: Schools and Staffing Survey
NHES: National Household Education Survey

NOTE: From *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress* (pp. 36–37), by J.A. Pellegrino, L.R. Jones, and K.J. Mitchell, 1999, Washington, DC: National Academy Press. Copyright 1999 by the National Academy of Sciences. Reprinted with permission.

contextual variables have included enrollment in elementary and secondary schools, poverty status of children from 5 to 17 years old, number of children receiving disability services, per-pupil expenditures, pupil-teacher ratios, and average teacher salaries. To the extent that these data are readily available and are helpful in setting a context for interpretation of NAEP results the practice ought to be continued. However, more effort should be made to ensure that such data are as up-to-date as possible.

NAEP Background Information Matrix

The types of descriptive information NAEP collects, reporting units, and data sources are summarized in the NAEP Background Information Matrix, which is displayed as Figure 2. The matrix is intended to assist in conceptualizing NAEP background information collections. It is organized by priorities—both for types of information and for how data should be obtained. Note that in each case information is to be obtained from reliable official records before it is sought through questionnaires.

The entries in the cells are illustrative, showing the kinds of information that are currently collected by NAEP and the various data sources (records and questionnaires) that are used. As the principles of this framework are implemented, more information will come from records, less from questionnaires. The sources with higher reliability and less respondent burden should be utilized in priority order.

The Ad Hoc Committee on NAEP Background Questions considered a proposal by Paul Barton (2002) to permit states or groups of states to add customized sets of questions to the background questionnaires. Although these might track progress on topics of particular interest and increase support for NAEP, the Committee felt strongly that the proposal should not be pursued because any customization of NAEP questionnaires would create serious logistical and quality control problems.

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. The average individual response time to answer 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.

Figure 2. NAEP Background Information Framework

Reporting Unit and Data Sources	Type of Information			
	Student Reporting Categories	**Socio-Economic Status—Core or Expanded**	**Other Contextual Information**	**Subject-Specific Information**
STUDENT				
School Records	Gender Race/ethnicity SD/LEP	Free/RP lunch participation Title I	New enrollee Type/degree of disability	Course taking in mathematics
Questionnaire	Race/ethnicity	Parent education Reading materials in home	Daily reading Discuss school work TV-watching Absenteeism Language in home	Time spent on math homework Good in math?
SCHOOL				
Dist/State Recds **School Records**	School type (public, private, charter, etc.) School ach. data	% Free/RP lunch participation Title I funding	Grade structure Days of instruction Enrollment % LEP	Graduation requirements in math/science Higher level math courses Graduation testing
CCD/Census **Questionnaire**	Community type		% Students absent % Teachers absent Enrollment mobility Grade retention Teacher retention Graduation rates Post-secondary ed rates	Extracurricular options in math and English Availability of computers for writing
TEACHER				
School Records **Dist/State Recds**			Race and gender Experience Credentials	
Questionnaire			Undergrad/grad content training Professional devel	Correct for spelling and grammar? Frequency of lab work
STATE				
CCD/Census **State Records** **Questionnaire**	Region		Non-NAEP contextual variables	
DISTRICT				
CCD/Census **State Records** **District Records** **Questionnaire**		Community type (urban, rural, etc.)		

NOTE: Information type and data sources are arranged in priority order.

Chapter Five

Using Background Data to Report NAEP Results

This chapter discusses the descriptive information that NAEP should provide, the levels of disaggregation now possible with merged national and state samples, and the importance of minimizing causal interpretations.

Use of Descriptive Information in NAEP

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on the patterns and trends of non-cognitive variables known to have a relationship to academic achievement.

In addition, 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 made available on the Internet at the time of the initial release of the principal academic results or soon afterwards so the public may be able to consider them in discussing results. Complete data files should be available to researchers for further analysis.

Implementing No Child Left Behind

The intent of the No Child Left Behind Act of 2001 (P.L.107-110) is to hold public schools accountable for closing the achievement gaps between different groups of students. NAEP is asked to contribute to this end by providing an accurate measure of the current levels of student achievement and to monitor change over time.

Descriptive information about all students, but particularly on low-performing groups, would contribute powerfully to the dialogue on the challenges before American education. For example, the NAEP achievement levels focus on the segments of the performance distribution that are at or above *Basic, Proficient,* and *Advanced.* Information should also be provided about those *Below Basic,* who clearly have been "left behind": e.g., the proportion having qualified teachers, receiving free or reduced-price lunch, or moving to different schools frequently, as measured by attending the same school for less than two years.

Such profiles of low-performing or high-performing students would not attempt to ascribe causation, but they would provide information on the distribution of practices and resources that are of concern to the public and policy-makers. Periodic collections of such background data could be used to track change in the distribution of these factors over time. Do the trends seem favorable or adverse to educational progress?

Disaggregation of NAEP Data

For more than three decades NAEP has provided data disaggregated by race/ethnicity, gender, school type (e.g., public/private), and community type (e.g., urban/rural). The No Child Left Behind law calls for disaggregation by major subgroups (when feasible) of race, ethnicity, and gender, and also by socio-economic status, disability, and limited English proficiency.

Because of the large size of the recently combined national and state NAEP samples, NAEP reports should be able to provide information disaggregated at a much greater level of detail than was previously possible. Pooling the data from all states, which now are required to provide NAEP samples in 4th and 8th grade reading and mathematics, will produce a much-enlarged national sample that will sharply reduce the number of empty cells in any cross-tabulations. Such disaggregation might add to the richness of NAEP reporting even with only a limited set of non-cognitive questions. Disaggregation is also very important for reporting on the distribution of student characteristics within the different achievement levels, as described above.

Minimizing Causal Interpretations

NAEP has often reported on the average performance of students by particular non-cognitive variables. One example, presented in many NAEP reports, is the average scale score of students who watch different amounts of television each day, cf. *The Nation's Report Card: Reading, 2000* (Donahue et al., 2001). Another example is the average scale scores for 12th graders who report different amounts of time working at a part-time job, cf. *The Nation's Report Card: Mathematics, 2000* (Braswell et al., 2001).

While there may be a correlation between television watching and reading performance, or between hours working outside school and math results, NAEP is not designed to prove cause-and-effect relationships. As a cross-sectional survey, nearly all of its data are on current activities and practices—not on the complex chain of experience in school and outside, of prior learning and achievement that all contribute heavily to current academic performance. Yet, NAEP has encouraged simple causal inferences by reporting average scores for varying amounts of time spent on current activities.

There is one important exception to the absence of data on learning-related activity over time. This is the information NAEP collects on the transcripts of high school seniors and its questionnaires on courses that students have taken and schools provide. These do show prior instruction before current exams. The trends in course taking have been of great public interest and it is reasonable to relate them to student achievement.

NAEP reports should present information on the patterns and trends of non-cognitive variables known from other sound research to have a relationship to academic achievement. These presentations should be straightforward and impartial, and care must be taken to avoid stating conclusions as to cause and effect relationships. Further analysis of any relationships should be left to researchers.

Chapter Six

Using NAEP in Educational Research

As a cross-sectional survey without longitudinal data, the National Assessment of Educational Progress is able to document school conditions and practices. It can report on achievement results. But it cannot properly be used to establish direct cause-and-effect relationships. Still, over the past three 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 doing its main task of monitoring achievement well, NAEP has provided sound data for researchers to use. NAEP results have been critical in identifying hypotheses for other research to pursue. Its large data sets 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 the National Assessment can be of value to educational research and that the interest of researchers in the assessment should be encouraged. The NCES program of secondary analysis grants for researchers to use NAEP data should continue. Education researchers should be involved, under the auspices of NCES and its contractors, in developing NAEP background questionnaires and other data collection efforts to carry out the provisions of this framework.

This chapter considers the limitations and strengths of NAEP for educational research and discusses research that has made use of NAEP data. The chapter draws on papers by David Grissmer, senior research scientist at RAND, who has used NAEP extensively in analyzing educational factors and trends.

NAEP's Limitations and Strengths for Research

The primary purpose of NAEP is to ***accurately and fairly monitor achievement*** over time ***and accurately and fairly compare achievement across states and important sub-groups of students.***

Beyond providing such data, any research based on NAEP, particularly into the causes of academic achievement, is severely limited by its design.

As a representative sample survey, in which no individual student takes more than a small part of the full exam, NAEP has shortcomings in most of the elements commonly used to evaluate academic achievement (Podgursky, 2002):

- It provides no prior data on student achievement, and cannot be changed to gather longitudinal data.
- It can only collect contemporaneous information on school practices and resources, and has no way of ascertaining how students were taught or what school experiences they may have had in previous years.
- There is considerable measurement error in survey responses obtained from teachers and schools because they may well give the expected "right" answers rather than report accurately what they do.
- The current classroom practices that teachers report may be a response to student achievement levels, not the cause of such achievement, and it is difficult to disentangle causation.
- It is difficult for NAEP to get good information on socioeconomic status or family background factors, but these are powerfully correlated with academic achievement, and must be controlled for in any analysis of school effects.

On the other hand, NAEP does have unique strengths and comparative advantages (Grissmer, 2003), and thus has the potential to address some important research and public policy questions with its cognitive data and background information:

- NAEP is the only data set on student achievement that has collected data from nationally representative samples of students continuously from 1969–70 to the present.
- It is the only data set that has collected academic achievement data simultaneously, repeatedly, and consistently from three separate age groups.
- It is the only data set that collects from statistically reliable samples at the state level, and within states for different types

of communities (central city, suburban and rural) and for racial/ethnic groups within most states.

- NAEP has far larger sample sizes than any other nationally representative survey of student achievement, such as the National Education Longitudinal Study (NELS) and the Early Childhood Longitudinal Study (ECLS). These surveys are only approximately 10 to 20 percent as large as NAEP in any single application, and 1 to 5 percent as large as NAEP in any repeated data collection.
- NAEP is the only survey that tests a wide range of academic subjects.
- NAEP achievement measures at 4th and 8th grade fill an important void in measuring the well-being of children during this developmental period.
- NAEP generally incorporates a higher quality and unique design of test instruments, administrative procedures, and scoring methodology, compared to other data sets.

Previous Use of NAEP in Research

As a result of its strengths, NAEP has been used in important educational research by authors such as David Grissmer, Alan Krueger, David Armor, and Christopher Jencks. These studies point to an important comparative advantage of NAEP, namely, that it is the only representative sample data in existence on student achievement in the United States from 1969 to 2002. Thus, research into important historical questions about the effects of changing families, communities, and schools on achievement almost require NAEP data. Without NAEP, it is unlikely that the significant narrowing of the black-white score gap would be known and its possible causes the subject of research.

Similarly, NAEP data have been used to help analyze the effects of differences in resources, systemic reform initiatives, differential opportunity for learning, and other educational policies on state-level academic achievement. Such research has concluded that the rates of improvement in achievement varied markedly across states in the 1990s, and that changing resources or demographics cannot account for the gains in the states with most rapid improvement. This

research points to another strong comparative advantage of NAEP. State NAEP is the only survey that includes representative samples of students in many different states, and thus plays a central role in monitoring and explaining the differences in academic achievement and achievement trends across the states. NAEP can identify where positive trends are occurring so researchers can puzzle out causation.

A review of research studies using NAEP (Grissmer, 2003) suggests that only a small proportion of the non-cognitive items collected by the assessment have been utilized in productive research. Also, such research has often supplemented NAEP with data from other sources, such as the U.S. Census and the Common Core of Data (CCD) and Schools and Staffing Survey (SASS), both conducted by the National Center for Education Statistics. However, the National Assessment played such a crucial role in these studies that they could not have been conducted without NAEP data, including some of its non-cognitive variables, principally those on socio-economic status, family structure, and school resources.

On the other hand, NAEP data have also been misused for simplistic and weak research. Many background items on school practice and student behavior have been used in a simplistic way to imply a direct, causal relationship to achievement while ignoring the complex mix of other, more fundamental factors that may well have a stronger impact. NAEP has encouraged such associations by presenting one-way tabulations in its reports, such as average scale score by hours of television watched, type of reading instruction, or books read per week, and these have been disseminated widely to support particular beliefs or public policy positions. Simple, single-variable linkages can often be misleading because of the strong correlations between many background variables, particularly with socio-economic status, prior academic achievement, or family background. They should only be included in NAEP reports when there is strong justification based on previous research.

Also, most of the hundreds of background questions in NAEP have never been used for either public reporting or research. Many come from the early 1980s, and would be difficult to justify in a sound research design today.

Secondary Analysis Grants and District Samples

For many years NCES has been making awards to education researchers for secondary analyses of NAEP data. These have explored a range of topics, often in combination with other data sets. Many of the studies have focused on state-to-state differences in student achievement and the impact of state-level policies, relying on NAEP academic data, a few background questions for SES controls, and much additional information from other sources. The program has been valuable as a means of encouraging the use of NAEP for research, and, in a few cases, notably the Grissmer studies, has had considerable impact. As in any grant program, all findings are the responsibility of the individual researchers, not of the agency making the grant.

The program should continue, and now that NCES has become part of the Institute for Education Sciences, the leadership of the new agency should ensure that the analysis grants are aligned with the research priorities of the Institute.

In addition, data from the school district NAEP samples in the Trial Urban District Assessment, started in 2002, will provide important new opportunities for research. NAEP results for school districts can readily be combined with Census data, which include pertinent information on family background and socio-economic status. The school district samples can also be tied to important education policy variables, such as per-pupil spending, for which information is available at this level but not for schools.

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

Chapter Seven

Review and Improvement of Non-Cognitive Questions

This chapter discusses several mechanisms for the review and improvement of NAEP's non-cognitive questions and for implementation of the NAEP Background Information Framework.

Independent Validity Studies

Since the early 1990s NAEP has had the benefit of independent outside advice on topics of urgency or interest. These studies have been very helpful to the Governing Board and NCES as they made decisions about the future of the NAEP program. For example, several years ago research was conducted on the possibility of combining the national and state samples in NAEP to achieve greater efficiency and cost-savings. Starting in 2003 NAEP has moved in that direction. The decisions surrounding such change, however, can only be as good as the research that informs them. The work of the current NAEP Validity Panel, in conjunction with the current NAEP operations contractors, contributed significantly to making the change possible.

The value of this kind of applied research cannot be overestimated. Neither can the value of the independent nature of such work. The NAEP program is very large and complex and demands a commitment of many resources from the NAEP contractors. NAEP contractors should not be burdened with conducting simultaneous research studies while carrying out the requirements of the operations contracts. There is a precedent for this approach in the current separation of responsibilities for operations and research in separate NAEP contracts. There are two reasons why independent validity studies on topics associated with the non-cognitive framework are recommended. First, there are some non-cognitive variables that will need validation, particularly if those variables are new or are new composite indices of existing variables. Second, following the approach already

established for the cognitive components of NAEP, recommendations from research studies should be truly independent and free from any conflict of interest.

Review of Background Information Framework

This background information framework should be reviewed on a periodic basis. The NAEP cognitive frameworks are reviewed every ten years. This policy was adopted at the time of the NAEP redesign in 1996. Reviewing a NAEP framework can result in major revision, minor revision, or even no revision and re-adoption. Since the background framework is a new undertaking, a required review after five years is appropriate with additional reviews every ten years thereafter.

References

Allen, N.L., Carlson, J.E., and Zelenak, C.A. (1999). *The NAEP 1996 technical report.* Washington, DC: U.S. Government Printing Office.

Armor, D.J. (December 18, 2002). *Comments on NAEP non-cognitive questions.* Available from the National Assessment Governing Board, Washington, DC.

Barton, P.E. (2002, September). *Perspectives on non-cognitive questions in the National Assessment of Educational Progress.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Braswell, J.S., Lutkus, A.D., Grigg, W.S., Santapau, S.L., Tay-Lim, B., and Johnson, M. (2001). *The Nation's report card: Mathematics 2000.* Washington, DC: National Center for Education Statistics.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., and York, R.L. (1966). *Equality of Educational Opportunity.* Report of the Equality of Educational Opportunity Commission (EEOC). Washington, DC: National Center for Education Statistics.

Donahue, P.L., Finnegan, R.J., Lutkus, A.D., and Allen, N.L. (2001). *The Nation's report card: Reading 2000.* Washington, DC: National Center for Education Statistics.

Donahue, P.L., Voelkl, K.E., Campbell, J.R., and Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states.* Washington, DC: National Center for Education Statistics.

Forsyth, R., Hambleton, R.K., Linn, R., Mislevy, R., and Yen, W. (1996). *Design/feasibility team report to National Assessment Governing Board.* Washington, DC: National Assessment Governing Board. Available at http://nagb.org.

Greenberg, E., Stancavage, F., Farr, B., and Bohrnstedt, G. (January 2001 draft). *Making comparisons: Using background information from the Nation's report card to understand student achievement.* Washington, DC: American Institutes for Research.

Grissmer, D. (2002, September). *Improving NAEP for research and policymaking.* Paper presented at workshop on NAEP background

questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Grissmer, D. (2003, January). *Utilizing NAEP for educational research.* Paper prepared for National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Mayer, D.P., Mullens, J.E., and Moore, M.T. (2001). *Monitoring school quality: An indicators report.* Washington, DC: National Center for Education Statistics.

Mullis, I.V.S. (2002, September). *Non-cognitive questions in TIMSS and PIRLS: An overview.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

National Assessment Governing Board. (1996, August). *Policy statement on redesigning the National Assessment of Educational Progress.* Washington, DC: Author.

National Assessment Governing Board. (2002, May). *Policy statement on collection and reporting of non-cognitive data by the National Assessment of Educational Progress.* Washington, DC: Author.

Pellegrino, J.W., Jones, L.R., and Mitchell, K.J. (eds.) (1999). *Grading the Nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* Washington, DC: National Academy Press.

Podgursky, M. (2002, September). *NAEP non-cognitive questions: What can we learn from NAEP about the effect of schools and teachers on student achievement?* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Public Law 107-110. *No Child Left Behind Act of 2001.* Section 602. Available at http://nagb.org.

Reckase, M.D. (2002, September). *Contributions of non-cognitive questions to improving the precision of NAEP results.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Reese, C.M., Miller, K.E., Mazzeo, J., and Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states.* Washington, DC: National Center for Education Statistics.

Rust, K. (September 26, 2002). *Technical memo: Sample design for 2003 NAEP.* Available from NAGB Board Meeting briefing book, November 14–16, 2002. National Assessment Governing Board, Washington, DC.

Walberg, H.J. (2002, September). *The National Assessment's most useful non-cognitive items.* Paper presented at workshop on NAEP background questions. National Assessment Governing Board, Washington, DC. Available at http://nagb.org.

Appendix

Adopted May 18, 2002

National Assessment Governing Board

Policy Statement on Collection and Reporting of Background Data by the National Assessment of Educational Progress

Introduction

As the Nation's Report Card, the National Assessment of Educational Progress (NAEP) is an on-going, Congressionally-authorized program to collect data through surveys on the academic knowledge and skills of American students. Its primary goal is to report fair and accurate information on student achievement in reading, mathematics, and other subjects taught in elementary and secondary schools. This information is to be made available in a clear and timely manner to members of the public, policymakers, and educators throughout the country.

Since it began in 1969–70, NAEP has administered, in addition to cognitive questions, background questionnaires that provide information for reporting categories and collect non-cognitive data on students, their family background, teachers, and schools. These have

enriched reporting of the National Assessment and increased the precision of NAEP results. The background data have also been used in secondary analyses. However, because NAEP tests a cross-section of students at a particular time with no follow-up of the students tested, the assessment can only show correlations or associations rather than causal relationships between background factors and achievement.

By statute (P.L. 107-110), the National Assessment shall include, "whenever feasible, information collected, cross-tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency." The statute provides that NAEP may "not evaluate or assess personal or family beliefs and attitudes" and may "only collect information that is directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information." These provisions are intended to prevent intrusive, inappropriate, or unnecessary questions being asked about students and their families.

The law requires that the Governing Board take steps to ensure that all NAEP questions are "free from racial, cultural, gender, or regional bias, and are secular, neutral, and non-ideological." However, a House-Senate Conference report, accompanying the legislation, says the law does not preclude the use of "non-intrusive, non-cognitive questions," with a direct relationship to academic achievement.

The National Assessment is conducted by the Commissioner of Education Statistics under the policy guidance of the National Assessment Governing Board. The Board's specific areas of responsibility include: (1) assessment objectives and test specifications; (2) the methodology of the assessment; (3) guidelines for reporting and disseminating results; and (4) "appropriate actions needed to improve the form, content, use, and reporting" of the National Assessment. Under the statute, the Board has "final authority" on the appropriateness of all NAEP items—both cognitive and non-cognitive.

To carry out these responsibilities, the National Assessment Governing Board hereby adopts guiding principles, policies, and procedures for the collection and reporting of background data by the National Assessment of Educational Progress.

Guiding Principles

1. Background data on students, teachers, and schools is needed to fulfill the statutory requirement that NAEP include information, whenever feasible, disaggregated by race or ethnicity, socioeconomic status, gender, disability, and limited English proficiency. In addition, background data is collected to enrich the reporting of NAEP results by examining factors related to academic achievement. However, the collection of such data should be limited, and the burden on respondents kept to a minimum. It must always be considered in light of NAEP's primary purpose: providing sound, timely information on the academic achievement of American students.
2. All background questions must be directly related to academic achievement or to the fair and accurate presentation of achievement results.
3. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining the background questions to be asked and the samples to which they shall be administered.
4. In accordance with law, questions shall be non-intrusive and free from bias, and must be secular, neutral, and non-ideological.
5. No personally identifiable information shall be included in NAEP reports or data releases.
6. Decisions on the retention or addition of background items shall draw on technical studies on the reliability and validity of current and proposed questions and on the contribution such items make to the precision of NAEP results.
7. Consideration should be given to obtaining background information from non-NAEP sources and to avoiding duplication with other federal surveys.
8. Questionnaires should be revised to keep background questions timely and related to academic achievement. Those questions showing little change over time and/or a stable relationship to achievement should be deleted or asked less frequently and to limited samples, unless required to ensure the precision of NAEP results.

9. Questions should not address personal feelings and attitudes.
10. Since security considerations do not apply, background questionnaires shall be readily available to the public.
11. Interpretation of results shall be limited in official NAEP reports and must be strongly supported by NAEP data. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships.
12. Background questions for NAEP assessments shall be prepared in accordance with frameworks and specifications adopted by the Governing Board.
13. The Governing Board shall review and approve all background items before they are administered in NAEP surveys or pilot and field tests.

Policies and Procedures

1. Framework and Specifications

The Governing Board will adopt a general framework for background questionnaires and specifications for the questions on selected topics and in specific subject areas.

Since this is a new area of responsibility for the Board, the process of developing a framework for background questions and specifications will begin with commissioned papers on relevant issues, such as the reliability and validity of current background questions, their contribution to improving the precision of NAEP results, their value and limitations for educational research, and changes that may be needed in response to the No Child Left Behind legislation. Following consideration of these issues, the Board will define the scope of background questionnaires and adopt a process for preparing a framework and specifications. This work will include the active participation of teachers, education researchers, state and local school administrators, assessment specialists, parents of children in elementary and secondary schools, and interested members of the public.

2. Background Question Development

In preparing background questions, the National Center for Education Statistics shall follow adopted frameworks and specifications, and consider the review criteria adopted by the Governing Board. NCES may use cognitive laboratories of students, teachers, and school officials to help determine the clarity and burden of proposed questions. Ad hoc advisory committees may also be established, comprised of teachers, parents, technical experts, and others interested in NAEP. Steps shall be taken to determine the reliability of questions used.

3. Governing Board Review and Approval of Background Questions

Background questions for all NAEP pilot tests, field tests, and operational use shall be reviewed and approved by the Governing Board. The category of respondents, e.g. students, schools, and grade level, shall clearly be designated, as will the NAEP samples, e.g. national, state, or district, in which the questions will be asked.

For each questionnaire there shall be an explanation of its intended use in NAEP reporting and analysis and of the hypothesized relationships between the background items and student achievement that demonstrates the need to know such information. Technical data shall be presented on the reliability and validity of questions and, if applicable, on their contribution to improving the precision of NAEP results. The Board will use the explanations and data presented along with the review criteria in this policy statement in determining the appropriateness of background questions.

The Reporting and Dissemination Committee will have primary responsibility for the review and approval of background questions. The Assessment Development Committee will participate in the approval of questions relating to specific subject-matter assessments. Ad hoc committees of Board members may be established by the Board Chairman for background question review. Questions may also be reviewed by external advisors, including teachers, parents, and technical experts. Recommendations on background questionnaires shall be subject to final approval by the full Governing Board.

4. Criteria for Governing Board Review

The following criteria for review and approval of background questions are based on the most recent revision of the authorizing statute of the National Assessment of Educational Progress (P.L. 107-110) and the Guiding Principles of this policy statement:

A. Background information is needed to fulfill the statutory requirement that NAEP report and analyze achievement data, whenever feasible, disaggregated by race or ethnicity, gender, socio-economic status, disability, and limited English proficiency. Non-cognitive data may enrich the reporting and analysis of academic results, but the collection of such data should be limited and the burden on respondents kept to a minimum.

B. All background questions must be related to the primary purpose of NAEP: the fair and accurate presentation of academic achievement results.

C. Any questions on conditions beyond the school must be non-intrusive and focused on academic achievement and related factors.

D. Questions shall be free from racial, cultural, gender, or regional bias.

E. All questions must be secular, neutral, and non-ideological. Definitions of these terms, accompanied by clarifying examples, are presented in Attachment A, as adopted in the Governing Board Policy on NAEP Item Development and Review.

F. NAEP must not evaluate or assess personal feelings or family beliefs and attitudes unless such questions are non-intrusive and have a demonstrated relationship to academic achievement.

G. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining which questions to include in background questionnaires. These factors must also be considered in determining the frequency with which various questions shall be administered and whether they shall be included in both national and state samples.

H. Background questions that do not differentiate between students or have shown little change over time should be deleted or asked less frequently and to limited samples.

5. Public Access to Background Questions

Since security considerations do not apply, all background questionnaires shall be readily available to parents, teachers, state and local officials, and interested members of the public. Such questionnaires shall be available before field tests and operational assessments or at any other time members of the public wish to obtain them. Background questions in operational use shall be posted on the Internet prior to each assessment, accompanied by explanations and rationales.

6. Reporting of Background Information

The presentation of background data in official NAEP reports shall be straightforward and impartial. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships. Any composite indices including demographic and socioeconomic factors shall be presented to the Board for approval before use as reporting categories in NAEP data releases and reports.

Background data should be available for extensive secondary analyses by scholars and researchers, who are responsible for conclusions reached. Responses to background questions shall be presented and tabulated on the Internet, although, if necessary, posting may be delayed for a brief period after release of the principal NAEP results.

Attachment A

Definitions of Secular, Neutral, and Non-Ideological Item Review Criteria

From Governing Board Policy on NAEP Item Development and Review—5/18/02

Items shall be secular, neutral, and non-ideological. Neither NAEP nor its questions shall advocate a particular religious belief or political stance. Where appropriate, NAEP questions may deal with

religious and political issues in a fair and objective way. The following definitions shall apply to the review of all NAEP test questions, reading passages, and supplementary materials used in the assessment:

Secular—NAEP questions will not contain language that advocates or opposes any particular religious views or beliefs, nor will items compare one religion unfavorably to another. However, items may contain references to religions, religious symbolism, or members of religious groups where appropriate.

Examples: The following phrases would be acceptable: "shaped like a Christmas tree," "religious tolerance is one of the key aspects of a free society," "Dr. Martin Luther King, Jr., was a Baptist minister," or "Hinduism is the predominant religion in India."

Neutral and Non-Ideological—Items will not advocate for a particular political party or partisan issue, for any specific legislative or electoral result, or for a single perspective on a controversial issue. An item may ask students to explain both sides of a debate, or it may ask them to analyze an issue, or to explain the arguments of proponents or opponents, without requiring students to endorse personally the position they are describing. Item writers should have the flexibility to develop questions that measure important knowledge and skills without requiring both pro and con responses to every item.

Examples: Students may be asked to compare and contrast positions on states rights, based on excerpts from speeches by X and Y; to analyze the themes of Franklin D. Roosevelt's first and second inaugural addresses; to identify the purpose of the Monroe Doctrine; or to select a position on the issue of suburban growth and cite evidence to support this position. Or, students may be asked to provide arguments either for or against Woodrow Wilson's decision to enter World War I. A NAEP question could ask students to summarize the dissenting opinion in a landmark Supreme Court case.

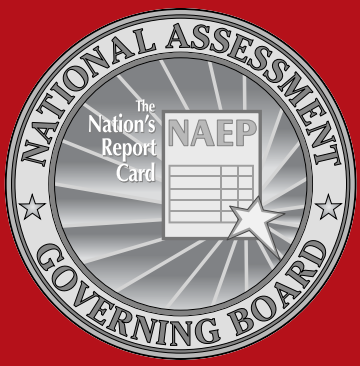
The criteria of neutral and non-ideological also pertain to decisions about the pool of test questions in a subject area, taken as a whole. The Board shall review the entire item pool for a subject area to ensure that it is balanced in terms of the perspectives and issues presented.

Acknowledgments

The Ad Hoc Committee on NAEP Background Questions of the National Assessment Governing Board was chaired by Board member John H. Stevens. He also serves as chairman of the Board's standing Committee on Reporting and Dissemination, which has responsibility for reviewing all core NAEP background questionnaires and making recommendations on them to the full Board.

The Ad Hoc Committee also included members of the Board's two other standing committees—the Assessment Development Committee and the Committee on Standards, Design, and Methodology—with a wide range of backgrounds and perspectives: Amanda Avallone, Dwight Evans, Thomas Fisher, Sheila Ford, Jo Ann Pottorff, and Sister Lourdes Sheehan. The Board chairman, Darvin Winick, participated in many of the lively discussions that marked the Committee's deliberations.

Among the many discussants and presenters at the workshop and public forum, we wish to recognize the care and thoughtfulness of Robert Mislevy, of the University of Maryland, and Harold Wenglinsky, of Baruch College of the City University of New York. The comments submitted by Sandra Feldman, president of the American Federation of Teachers, were particularly perceptive and helpful. This project also benefited greatly from the continuing advice and insight of Paul Barton and David Grissmer, both of whom have used NAEP data for many years to understand and explain American education to its public.


NATIONAL ASSESSMENT
GOVERNING BOARD
The
Nation's
Report
Card
NAEP