National Assessment Governing Board

Assessment Development Committee

March 2-3, 2017

AGENDA

Thursday, March 2		
1:00 – 1:30 pm	**Closed Session** Welcome and Introductions *Cary Sneider, ADC Vice Chair* Review of NAEP Long Term Trend Items • Discussion of Item Comments • Q & A *ADC Members* *Ina Mullis, Professor of Educational Research, Measurement, and Evaluation, Boston College*	Attachment A
1:30 – 4:00 pm	Review of NAEP Cognitive Items in Reading, Mathematics, and Science	Secure material provided under separate cover

Friday, March 3		
10:00 – 11:00 am	Joint Session with the Committee on Standards, Design and Methodology on Dynamic Frameworks *Andrew Ho, COSDAM Chair* *Cary Sneider, ADC Vice Chair*	Attachment B
11:00 – 11:10 am	Break	
11:10 am –12:00 pm	ADC's Role in Implementing the Strategic Vision *Cary Sneider, ADC Vice Chair*	See material sent under separate cover
Information Item	Item Review Schedule	Attachment C

Discussion: NAEP Long-Term Trend Assessments in Reading and Mathematics

Overview

As stated in the NAEP statute (P.L. 107-279), the Commissioner for Education Statistics shall "continue to conduct the trend assessment of academic achievement at ages 9, 13, and 17 for the purpose of maintaining data on long-term trends in reading and mathematics."

The Governing Board has previously reviewed policy issues for the NAEP Long-Term Trend (LTT) assessments, and the Board's Strategic Vision, adopted November 2016, calls the Board to further exploration and discussion on NAEP LTT in order to:

> "Research policy and technical implications related to the future of NAEP Long-Term Trend in reading and mathematics."

Continuing from the November 2016 Committee discussion of the history, design, and content of the LTT assessments, the purpose of this closed ADC discussion is to further discuss the content of the assessment. Under separate cover, the Committee received secure access to the reading and math test items used in the 2012 LTT assessment – see the executive summary from the 2012 Long-Term Trend report here. This item review provides the ADC with a closer look at the content of the assessment.

LTT content expert Ina Mullis will join the ADC at this session for Q&A as the Committee considers:

- How does LTT content compare to the content of main NAEP assessments in reading and mathematics?
- Does this content belong in the Nation's Report Card?

The content of the LTT assessments is an important consideration in the upcoming discussions on how to implement the Board's Strategic Vision for LTT. These discussions will include the Board's planned LTT symposium in March 2017. In upcoming March and May 2017 Board deliberations on the future of the Long Term Trend assessments, the ADC will provide guidance to the Board in identifying and grappling with the content issues.

Reference materials for this session include:

- Long Term Trend: History and Next Steps
- Comparison Chart of Long-Term Trend and Main NAEP
- Long-Term Trend Content Objectives: Reading and Mathematics

Long Term Trend: History and Next Steps

History of LTT and Main NAEP Assessments

NAEP includes two national assessment programs—Long-Term Trend (LTT) NAEP and Main NAEP. While both assessments enable NAEP to measure student progress over time, there are similarities and differences between the two assessments. Both assessments measure reading and mathematics. The NAEP LLT assessment measures national educational performance in the United States at ages 9, 13 and 17. In contrast, the Main NAEP assessments focus on populations of students defined by grade, rather than age, and go beyond the national level to provide results at the state and district level. LTT trend lines date back to the early 1970s and Main NAEP trend lines start in the early 1990s. The content differs as well—for example, LTT math measures more "traditional" mathematics than the current Main NAEP math content.

The Main NAEP assessments in reading and mathematics are administered every two years, as required by law. The administration of NAEP LTT assessments in reading and mathematics at ages 9, 13, and 17 is also required by law, but the periodicity is not specified. The NAEP LTT assessments had been administered approximately every four years over the past two decades (and more frequently prior to that), but were last administered in 2012. The Governing Board postponed the NAEP LTT planned administration for 2016 to 2020, and then to 2024 due to budgetary constraints. Some stakeholders have expressed concern with the gap of 12 years between assessment administrations, which represents a cohort's entire length of schooling. Other stakeholders argue that the NAEP LTT is not very useful now that Main NAEP provides trend information back to the early 1990s, and that it should be eliminated altogether.

Next Steps

In 2012, the Future of NAEP panel recommended exploring ways of consolidating or combining Long-Term Trend and Main NAEP data collections. This is a complex challenge due to the many differences in content, sampling, and administration of the assessments. To explore the feasibility of combining the data collection efforts, and to debate the relative merits of NAEP LTT, the Governing Board is organizing a symposium on the future of NAEP Long-Term Trend. The symposium will take place on the morning of March 2, 2017, immediately preceding this Committee Session.

In advance of the symposium, Edward Haertel of Stanford University prepared a white paper of approximately 30 pages on the history of NAEP Long-Term Trend and a consideration of current issues. Four other symposium participants have prepared a shorter responses (8-10 pages) on their perspective on the future of NAEP LTT. The papers will be disseminated in advance of the symposium and will serve as the basis for discussion during the March 2nd event. The papers, speaker bios, and event details are available here.

Symposium participants will also discuss their perspectives and solicit external input at a planned session during the annual American Educational Research Association (AERA) conference in April 2017.

During the May 2017 quarterly meeting, the Governing Board will discuss key takeaways and potential next steps regarding the future of the NAEP Long-Term Trend assessments.

What Are the Differences Between Long-Term Trend NAEP and Main NAEP?

Although long-term trend and main NAEP both assess mathematics and reading, there are several differences, particularly in the content assessed, how often the assessment is administered, and how the results are reported. These and other differences mean that results from long-term trend and main NAEP cannot be compared directly.

	Long-Term Trend Assessment	Main NAEP Assessment
Origin	Reading series began in 1971. Mathematics series began in 1973.	Reading series began in 1992. Mathematics series began in 1990.
Frequency	Since 2004, long-term trend NAEP has measured student performance in mathematics and reading every four years. Last reported for 2008, it will be reported next for 2012.	Main NAEP assessments measure student performance in mathematics and reading every two years.
Content Assessed	Long-term trend NAEP has remained relatively unchanged since 1990. In the 1970s and '80s, the assessments changed to reflect changes in curriculum in the nation's schools. Continuity of assessment content was sufficient not to require a break in trends. Mathematics focuses on numbers and numeration, variables and relationships, shape and size and position, measurement, and probability and statistics. Basic skills and recall of definitions are assessed. Reading features short narrative, expository, or document passages, and focuses on locating specific information, making inferences, and identifying the main idea of a passage. On average, passages are shorter in long-term trend reading than in main NAEP reading.	Main NAEP assessments change about every decade to reflect changes in curriculum in the nation's schools; new frameworks reflect these changes. Continuity of assessment content was sufficient not to require a break in trends, except in grade 12 mathematics in 2005. Mathematics focuses on numbers, measurement, geometry, probability and statistics, and algebra. In addition to basic skills and recall of definitions, students are assessed on problem solving and reasoning in all topic areas. Reading features fiction, literary nonfiction, poetry, exposition, document, and procedural texts or pairs of texts, and focuses on identifying explicitly stated information, making complex inferences about themes, and comparing multiple texts on a variety of dimensions.
Question formats	Students respond to questions in multiple-choice format; there are also a few short answer questions (scored on a two-point scale). In reading, there are also a few questions requiring an extended answer (usually scored on a five-point scale).	Students respond to questions of several possible types: multiple choice, short answer, and extended answer. Constructed-response questions may be scored as correct or incorrect, or they may be scored on a multi-level scale that awards partial credit.

	Long-Term Trend Assessment	Main NAEP Assessment
Students Sampled	Students are selected by age (9, 13, and 17) to represent the nation and to provide results for student groups such as Black, Hispanic, White, and sometimes others, by gender, family income, school location, and school type (public or private). Students with disabilities (SD) and English language learner (ELL) students are included using the same participation guidelines and with the same accommodations (as needed) in main NAEP. Since 2004, accommodations have been provided to enable participation of more SD and ELL students.	Students are selected by grade (4, 8, and 12). Students represent the nation and provide results for student groups such as Black, Hispanic, White, and sometimes others, by gender, family income, and school location and school type. In some assessments, samples are chosen to report on states or selected large urban districts and as a result, more students must participate. The inclusion and accommodation treatment is the same for main and for long-term trend assessments.
Administration	Long-term trend is assessed every four years, throughout the school year: in October through December for 13-year-olds, January through March for 9-year-olds, and March through May for 17-year-olds. See the schedule for all assessments (long-term trend as well as main NAEP). Test booklets contain three 15-minute blocks of questions, plus one section of student questions concerning academic experiences and demographics. There are no ancillary materials, such as calculators or manipulatives, provided.	Main NAEP mathematics and reading are assessed every two years (the odd-numbered years) at grades 4, 8, and 12. The administration takes place from late January through early March. Test booklets contain two 25-minute blocks, plus student questions concerning academic experiences and demographics. There may be ancillary materials provided with the test booklets.

	Long-Term Trend Assessment	Main NAEP Assessment
Results Reported	National-level performance and how it has changed since the 1970s is reported using scores on a 0-500 scale. Long-term trend also reports descriptive performance levels (150, 200, 250, 300, and 350) that have the same meaning across the three age levels. There are no achievement levels to correspond with those used in main NAEP. There are student questionnaires, but no teacher or school questionnaires.	Main NAEP has been reported since the 1990s for the nation and participating states and other jurisdictions, and since 2002 for selected urban districts. Performance and how it has changed over the past several years is reported using scale scores and achievement levels. Scores are reported using either a 0-300 or 0-500 scale, depending on the subject. The achievement levels reported are *Basic*, *Proficient*, and *Advanced.* Student results are reported in the context of the questionnaires given to the students' teachers and principals.

Source: https://nces.ed.gov/nationsreportcard/about/ltt_main_diff.aspx

Reading Objectives
1983-84 Assessment

No. 15-RL-10

by the
National Assessment of Educational Progress
CN 6710
Princeton, NJ 08541-6710

3

The National Assessment of Educational Progress is funded by the National Institute of Education under a grant to Educational Testing Service (ETS). It is the policy of ETS to take affirmative action to prevent discrimination in its policies, programs, and employment practices.

Library of Congress Catalog Card Number 72-169008

Although objectives booklets produced by the National Assessment of Educational Progress between 1969 and 1972 have their own individual catalog card numbers, the number above is a *series number* assigned to all National Assessment objectives booklets published since then.

ISBN 0-88685-022-3

National Assessment of Educational Progress, 1984

The National Assessment of Educational Progress is an education project mandated by Congress to collect and report data, over time, on the performance of young Americans in various learning areas. National Assessment makes available information on assessment procedures and materials to state and local education agencies and others.

The work upon which this publication is based was performed pursuant to Grant Nos. NIE-G-80-0003 (with Educational Commission of the States) and NIE-G-83-0011 (with Educational Testing Service) of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

Contents

Introduction

The reading objectives presented in this booklet are the most recent in a series that has included one previous set of combined reading and literature objectives (1979-80), two sets of reading objectives (1970 and 1974), and two sets of literature objectives (1970 and 1975).

With each successive set of objectives, the National Assessment of Educational Progress (NAEP) has tried to reflect advances in educational theory and practice. (See page 9 for a description of the process used to determine NAEP objectives.) The combination of reading and literature in the 1979-80 objectives marked a major shift in orientation as well as a recognition that the two areas involve many of the same goals. The present set of objectives carries forward this integration of objectives. In particular, separate objectives that dealt with the reader's comprehension (primarily of expository passages) and the reader's response (primarily to literary passages) have been reorganized. The objectives now reflect the current view that both the processes of comprehension and the extension of that comprehension through interpretation and analysis have a place in the reading of passages of all kinds. Objectives related to skills that support comprehension have also been reorganized in the present booklet; that is, those objectives are now incorporated as a part of the process of managing the reading experience. Included among the skills reorganized in this way are many previously grouped with study skills and with skills relating to awareness of text conventions and self-awareness.

The objectives are not defined in terms of age appropriateness. It is assumed that each objective and subobjective represents a continuum of difficulty. As students gain knowledge and experience, the complexity of the materials they read and of the tasks they are expected to perform increases. In addition, it is assumed that no fixed hierarchical relationship exists between objectives or between subobjectives.

Finally, the 1983-84 objectives were conceived as educational objectives that reflect the interactions of reader, text, and process rather than definitions of discrete units that can be directly translated into observable behaviors.

Objective I

Comprehends What Is Read

The first objective, *Comprehends What Is Read*, is central since every other objective is an outgrowth of that one.

Three factors apply to every reading situation: the type of material being read, the reader's purpose, and the background knowledge that the reader brings to the reading experience. Comprehension is an interactive process by which the reader constructs meaning both from the passage, which has a whole range of characteristics, and from the various kinds of background knowledge brought to the reading experience. Readers also bring their own purposes to the reading experience. These purposes guide them in setting expectations and deriving meaning consistent with their own goals. Thus, in discussing reading achievement, it is not enough to look at questions or tasks related to a particular passage. It is also necessary to ascertain the particular purposes for which the passage is to be read and to account for the kinds of knowledge that readers may already have that will help them more fully understand what they are reading. If concepts in the passage are new, they may need to be elaborated before readers will understand and remember them. If the concepts are familiar, readers may find it relatively easy to understand the passage—that is, to apply the concepts to new or more complex situations.

A. *Comprehends Various Types of Written Materials*

In their personal as well as their school lives, students encounter a wide variety of written materials; each of these poses its own problems of comprehension and interpretation. Making sense of the perhaps cryptic notes on a shopping list is different from understanding a complex essay or interpreting a literary work. Reading a science textbook differs from reading an historical essay. Letters, reports, inventories, and a wide range of record-keeping systems are integral to many businesses in today's "information society." To learn to manage problems of comprehension and interpretation, students need to read, discuss, and write about these different types of materials.

B. *Comprehends Materials Read for a Particular Purpose*

Reading purpose should determine the way something is read. The kind of attention required for skimming through a mail-order catalog to pick up relatively isolated bits of information differs from the kind of attention required for following detailed instructions line by line to assemble a new bicycle. These kinds of reading, in turn, differ markedly from the careful reading of integrated concepts that is required for preparing to write a research report. Similarly, the level and kind of attention needed for reading a play purely for enjoyment is quite different from that required for reading to prepare for directing or staging a play. Experience in reading for a variety of purposes can help the student develop varied strategies.

Objective II

Extends Comprehension

Whenever people read, to some degree they analyze, interpret, and evaluate the material they are reading. Objective II, however, has to do with deliberate, conscious kinds of analysis, interpretation, and evaluation of the sort, say, that a student undertakes when participating in a class discussion or that the reader is involved with when developing a viewpoint for a talk or a paper.

There are several major avenues that readers use in expanding their comprehension. They can examine their personal experience to increase their understanding of particular ideas, characters, or situations. They can use their awareness of the emotional impact of a passage as a source of information about its purpose and quality. They can make a general comparison of what they are reading with other materials they have read or they can examine particular ideas in light of specific information from other sources. They can examine the structure and conventions of a passage. They can judge the validity of the ideas and information presented. Such activities are not necessarily separate from one another; some or all may take place as readers extend their comprehension of any particular passage.

A. Analyzes What Has Been Read

When they analyze what they have read, readers may clarify their initial interpretations by employing increasingly explicit ways of communicating their views to others. Analysis can take many different forms. It may involve tracking the logic of an argument, identifying the emotional appeals underlying a political statement, explaining the motivations of a character in a story, or tracing the causes of a sequence of historical events. Such activities can lead to the discovery of inconsistencies in an initial interpretation (and hence to a reinterpretation of the passage) or they can lead to the discovery of additional evidence for explaining or defending an initial point of view.

B. Interprets What Has Been Read

Fluent readers use a variety of skills to deepen their understanding of what they have read. These include relating the concepts to their own experiences, to other works they have read, and to their own initial reactions to a passage. After putting a passage aside, readers may reflect on their own experiences with similar problems or events and may, in the process, form opinions concerning the validity or worth of what has been written. They may also compare what they are reading with something they have read before. Sometimes this means relating two books by the same author. Sometimes it means exploring other sources of information on the same topic. Sometimes it means relating a work to other works dealing with the same historical, cultural, or ideological theme. Such explorations are important steps in extending comprehension of any set of new ideas or experiences.

Reading involves both intellectual understanding and personal response. Many works are intended to entertain, persuade, or illustrate through emotional appeals. Therefore, another goal of reading instruction is to help students become aware of their emotional reactions in interpreting what they read. By articulating their personal reactions through discussion or writing, students can become more involved with characters, events, and ideas. They can also better understand the subtle ways in which writers influence their audiences. One way is to present a serious message within the context of a humorous piece. Another is to use an emotional appeal to promote a cause that cannot stand rationally on its own merits.

C. Evaluates What Has Been Read

One part of a reader's reaction to any passage is a judgment or evaluation of its usefulness or quality. At the simplest level, such a judgment controls the initial selection of reading material as well as the decision about continuing once the reading is under way. At a more formal level, readers judge the success of a work against either their specific purposes for reading or more general criteria of successful writing.

In most situations, evaluation is intertwined with a reader's comprehension of a passage and continues throughout interpretation and analysis. Defending or explaining an evaluation helps the reader articulate the criteria upon which an evaluation is based and relate characteristics of the work to those criteria.

Instruction in reading and literature should not lead students to a single scale of values by which to judge what they read. Rather, it should lead students to develop their own values and apply them appropriately to a variety of reading experiences.

Objective III

Manages the Reading Experience

Good readers develop a variety of strategies to help them comprehend what they read. Applied throughout the reading experience, these strategies vary according to the characteristics of particular passages, the reader's knowledge and experience with similar materials, and the reader's purpose for reading.

A. Uses the Structure and Organization of the Text

Comprehension of a passage is based on information drawn from many different elements at many different levels. Traditionally, teachers have tended to view these elements hierarchically, beginning with words, then moving to relationships among words and sentences, and then to devices that give structure to the passage as a whole. Actually, these elements cannot stand alone. They are all interrelated; and they also are related to the reader's previous experience. Indeed, in reading an entire passage or a complete work, good readers are aware of and sensitive to relationships and structures that govern larger units of a text. For example, sensitive readers develop an awareness of an evolving plot and of the relationships among the characters. In general, a good reader is guided by a sense of the structure of the particular genre (story, newspaper article, letter, research report) as well as by a growing understanding of the author's purpose and direction.

In longer works, paragraphs, clauses, and sentences are typically linked together to express relationships among the ideas or events that are being presented. Sometimes the relationships are stated, as in the following sentence: "The table wobbled because one leg was shorter than the other three." At other times, the relationship is simply implied: "Sarah hit Jim. Jim went home crying." Good readers look for these relationships to help them understand the passage they are reading.

Word meanings are, of course, dependent on context. The word *fly* has one meaning in the context of getting from New York to Chicago and quite another in the context of a baseball game. Vocabulary skills involve both the understanding of various dictionary meanings and the ability to choose from among those meanings according to the context in which the word is used.

B. Uses Readers' Aids

Many books provide a variety of aids that can simplify their use. These include typography (e.g., boldface, italics), layout (e.g., headings, subheadings), illustration (e.g., charts, graphs, photographs), and various kinds of listings and guides (e.g., table of contents, index, footnotes, bibliography, glossary). Although an experienced reader may automatically make use of such aids, a novice may need to have them pointed out and explained.

C. Shows Flexibility in Approach to Reading

Different purposes for reading require different approaches. For example, a reader may study a textbook carefully to remember details, read a mystery story quickly to get the gist of the plot, skim a newspaper article for an overall impression, or scan an encyclopedia entry to locate specific information. Notetaking, outlining, summarizing, or other study techniques can increase understanding and retention of what has been read. Good readers choose from among a variety of approaches, depending on their specific purpose in reading.

D. Selects Reading Materials Appropriate to the Purpose

From the vast array of reading materials available, readers must learn to select those appropriate for their purposes. Sometimes their selections are guided by the suggestions of parents, teachers, or friends. At other times, readers have to turn to the reference materials available in their school and community libraries. Some reference tools, such as dictionaries or encyclopedias, provide the reader with all the information that is needed. Others, such as bibliographies, card catalogs, indexes, and abstracts, may point them toward the required sources. In any case, readers must learn how to find the relevant materials and how to evaluate the usefulness of particular information.

Objective IV

Values Reading

Students should acquire a growing appreciation of the ways reading can affect their lives. At one level of appreciation, readers are marginally aware that reading can be pleasurable or informative. They choose reading over other activities only when the other activities are limited or unrewarding.

At another level of appreciation, readers actively seek opportunities to read or write. In their spare time at home or at school, they are often deep in a book they have chosen. They buy books or borrow them from the library and discuss what they read with friends and family. Some may even volunteer to tutor other students in reading.

A. Values Reading as a Source of Enjoyment

If students enjoy reading, they are likely to continue to read after their formal schooling is over. Thus, students should be encouraged to read for pleasure and to enjoy a wide variety of literary and expository materials.

B. Values Reading to Expand Understanding and Fulfill Personal Goals

Reading can enrich people's understanding of themselves and the world. Ideas or situations encountered in reading can help readers understand themselves, the people they meet, and the situations in which they find themselves. Some reading may be directly psychological, inspirational, or philosophical. Some may allow the reader to appreciate historical, contemporary, or fictional personalities. In some cases, reading can help develop a personal sense of justice and an understanding of the ranges of choice open to every individual.

C. Values Reading as a Means of Acquiring Knowledge and Learning New Skills

Reading serves a variety of utilitarian functions. People must read to choose groceries at the store, select a movie from the entertainment section of the paper, or complete income tax forms. They also must read to plan vacation trips, keep up with the daily news, and learn new skills.

The current popularity of "how to" books dramatizes the importance of written materials for acquiring knowledge and solving problems. Throughout the school years, textbooks provide students with information about new topics and once formal schooling is completed, reading continues to be a primary source of new information.

D. Values the Cultural Role of Written Language

Students should learn to appreciate the critical role written materials play in society. Words can profoundly affect individuals; and individuals, independently and collectively, change societies. As students mature, they gain an increasing sense of the importance of the interaction between written materials and society and of the importance of protecting and sustaining this interaction.

The Development Process

The reading objectives in this booklet were developed in preparation for the fourth national assessment of reading. Mail reviews and conferences organized by NAEP staff were conducted during the period between November 1982 and December 1983 to obtain information about the current thinking on reading from a variety of constituencies. Subject-matter specialists, teachers, school administrators, researchers, parents, and members of the lay public were asked to react to previous objectives and to comment on a draft of the new objectives. Participants in the objectives development process were:

Arthur Applebee	National Council of Teachers of English, Urbana, IL
Fernie Baca	University of Colorado, Denver, CO
Richard Beach	University of Minnesota, Minneapolis, MN
Barbara Bianchi	Paideia School, Atlanta, GA
Robin Butterfield	Northwest Regional Educational Laboratory, Portland, OR
Robert Calfee	Stanford University, Stanford, CA
Jeanne Chall	Harvard Graduate School of Education, Cambridge, MA
Carita Chapman	Swift Elementary School, Chicago, IL
Ruth Coleman	North Side High School, Mothers Alumni Club, Fort Wayne, IN
Larry Coon	Hamburger University (McDonald's), Oakbrook, IL
Bernice Cullinan	New York University, New York, NY

Mary E. Curtis	Harvard Graduate School of Education, Cambridge, MA
Jacqueline Danzberger	Youthwork Inc., Washington, DC
Philip DiStefano	University of Colorado, Boulder, CO
Priscilla Drum	University of California at Santa Barbara, Santa Barbara, CA
William Eller	State University of New York at Buffalo, Amherst, NY
Claryce Evans	Boston Public Schools, Boston, MA
Marjorie Farmer	School District of Philadelphia, Philadelphia, PA
Roger Farr	University of Indiana, Bloomington, IN
Edmund Farrell	University of Texas, Austin, TX
Edward Fry	Rutgers University, New Brunswick, NJ
Carol Gibson	National Urban League, New York, NY
Kenneth Goodman	University of Arizona, Tucson, AZ
Donald Graves	University of New Hampshire, Durham, NH
Doris Hankins	Germantown High School, Germantown, TN
Jerome Harste	University of Indiana, Bloomington, IN
David Hayes	University of Georgia, Athens, GA
Paul Heffernan	Star Market, Newtonville, MA
Harold Herber	Syracuse University, Syracuse, NY
Shu-in Huang	Personnel Department, City of Thornton, Thornton, CO
Judith Langer	University of California, Berkeley, CA
Diane Lapp	Boston University, Boston, MA
Charles Moody	University of Michigan, National Alliance of Black School Educators, Ann Arbor, MI
Edwin Newman	NBC News, New York, NY
Anthony Petrosky	University of Pittsburgh, Pittsburgh, PA
Beverly Roller	Jefferson County Public Schools, Lakewood, CO
Glenn E. Rotz	Highland Elementary School, Clarkson, WA
Sarah Saint-Onge	Godine Publishing Co., Boston, MA
Adan C. Salgado	Johnston High School, Austin, TX
S. Jay Samuels	University of Minnesota, Minneapolis, MN
Robert Schreiner	University of Minnesota, Minneapolis, MN
John Stewig	University of Wisconsin, Milwaukee, WI
Robert Tierney	University of Illinois, Champaign, IL
Jaap Tuinman	Simon Fraser University, Burnaby, B.C., Canada
Richard Vacca	Kent State University, Kent, OH
Sheila Valencia	University of Colorado, Boulder, CO
Thomas Vallejos	University of Colorado, Boulder, CO
Richard Venezky	University of Delaware, Newark, DE

Maria Watkins	University of Pennsylvania, Graduate School of Education, Philadelphia, PA
Kathy Yen	San Francisco Public Schools, San Francisco, CA
Seymour Yesner	Brookline High School, Brookline, MA

References

Bowes, J. E., Ed., & Norris, E. L., Ed. *Literature Objectives*. Ann Arbor, Michigan: National Assessment of Educational Progress, 1970.

Bowes, J. E., Ed., & Norris, E. L., Ed. *Reading Objectives*. Ann Arbor, Michigan: National Assessment of Educational Progress, 1970.

Literature Objectives; Second Assessment. Denver, Colorado: Education Commission of the States, National Assessment of Educational Progress, 1975.

Reading and Literature Objectives, 1979-80 Assessment. Denver, Colorado: Education Commission of the States, National Assessment of Educational Progress, 1980.

Reading Objectives; Second Assessment. Denver, Colorado: Education Commission of the States, National Assessment of Educational Progress, 1974.

MATHEMATICS OBJECTIVES

1981-82 Assessment

No. 13-MA-10

by the
National Assessment of Educational Progress

Education Commission of the States
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295

No. 13-MA-10
ISBN 0-89398-147-8

2

TABLE OF CONTENTS

4

LIST OF FIGURES

CHAPTER 1

THE FIRST ASSESSMENT OF MATHEMATICS (1972-73)

The National Assessment of Educational Progress conducted the first assessment of mathematics during the 1972-73 school year. The objectives for that assessment were developed by two educational testing contractors: Educational Testing Service and Psychological Corporation. Each contractor independently developed a set of objectives, relying on its staff, mathematicians and mathematics educators. The final sets of objectives were reviewed by panels of interested lay citizens to decide which set to use in the assessment. Panel members were evenly divided in their preference for the two sets of objectives. In the absence of a strong preference, the Psychological Corporation was asked to continue the development of objectives, and in 1968, completed its revision of the objectives. The revision, together with objectives selected from the Educational Testing Service's version, was compiled into a final statement of objectives for the first assessment of mathematics. A booklet containing the statement was published in 1970.

When the objectives for mathematics were first formulated, they were compared with other statements of objectives that had appeared in mathematics education literature during the preceding 25 years. The objectives for the first assessment were consistent with objectives appearing in the literature. This outcome was both desired and expected since one criterion for the National Assessment objectives was that they be central to prevailing teaching efforts.

A three-dimensional classification scheme was used to categorize the mathematics objectives for the first assessment. One dimension of the scheme was "Uses of Mathematics," which was divided into three major categories:

1. Social mathematics (the mathematics needed for personal living and effective citizenship in our society).
2. Technical mathematics (the mathematics necessary for various skilled jobs and professions).
3. Academic mathematics (the formally structured mathematics that provide the basis for an understanding of various mathematical processes).

Another dimension of the matrix was "Content." The content areas were:

1. Numbers and numeration concepts.
2. Properties of numbers and operations.
3. Arithmetic computations.
4. Sets.
5. Estimation and measurement.
6. Exponents and logarithms.
7. Algebraic expressions.
8. Equations and inequalities.
9. Functions.
10. Probability and statistics.
11. Geometry.
12. Trigonometry.
13. Mathematical proof.
14. Logic.
15. Miscellaneous topics.
16. Business and consumer mathematics.
17. Attitude and interest.

The third dimension of the classification scheme consisted of six cognitive "Objectives or Abilities":

1. To recall and/or recognize definitions, facts and symbols.
2. To perform mathematical manipulations.
3. To understand mathematical concepts.
4. To solve mathematical problems — social, technical and academic.
5. To use mathematics and mathematical reasoning to analyze problem situations, define problems, formulate hypotheses, make decisions and verify results.
6. To appreciate and use mathematics.

During the development and review of the exercises, the content and ability dimensions of the classification scheme were the most useful. The exercise developers tended not to use the first dimension, uses of mathematics, when classifying exercises. This first dimension tended to pose too many restrictions on exercise development to make its use worthwhile.

Although the exercises were classified by content and ability, not all content areas or abilities were assessed equally. Certain content topics were purposely measured in more detail than others. Furthermore, even though the objectives were intended to include all the mathematics taught in the nation's schools, it was impossible to measure every objective in depth. Little emphasis, for example, was placed on the topics of trigonometry and logic. The content area of "attitude and interest" and the related ability of "appreciation and use of mathematics" were not measured because the exercises developed to assess these were considered inadequate.

Approximately half of the exercises used in the mathematics assessment were released to the public. These exercises were included in various mathematics reports and made available to individuals, groups and states for their own uses. The unreleased exercises were used again in the second assessment of mathematics to measure changes in educational attainments. A general survey of the results of the first mathematics assessment is provided in *The First National Assessment of Mathematics: An Overview* (1975).[1] The text of each released exercise and accompanying documentation including results can be found in the *Mathematics Technical Report: Exercise Volume* (1977). Data are provided for all of the mathematics exercises, but the exact text and scoring guides are provided for the released exercises only.

Results concerning computational abilities of young Americans are presented and discussed in a special report, *Math Fundamentals: Selected Results From the First National Assessment of Mathematics* (1975). The results concerning consumer mathematics are presented and discussed in *Consumer Math: Selected Results From the First National Assessment of Mathematics* (1975). In addition, National Assessment has produced computer data tapes containing respondent-level data for the exercises that were reassessed in 1977-78. These data tapes are available to any researcher who wishes to conduct further analysis of the data.[2]

National Assessment has worked closely with a panel of mathematics educators from the National Council of Teachers of Mathematics (NCTM), who studied the data in order to draw implications from the results of the first mathematics assessment. The NCTM panel presented summaries of its findings in the October 1975 issues of *The Arithmetic Teacher* and *The Mathematics Teacher*. Additional brief articles on specific content topics were presented in the October 1975 through May 1976 issues of *The Arithmetic Teacher*. These articles suggest some of the ways mathematics teachers might use information from the first assessment to improve teaching. References for these articles are found in the section at the back of this booklet entitled "List of References."

---

[1]National Assessment reports can be ordered through the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 or the National Assessment of Educational Progress, Suite 700, 1860 Lincoln Street, Denver, Colorado 80295.

[2]Data tapes are available, at a charge, through the Department of User Services, National Assessment of Educational Progress, Suite 700, 1860 Lincoln Street, Denver, Colorado 80295.

CHAPTER 2

THE SECOND ASSESSMENT OF MATHEMATICS (1977-78)

Unlike the first assessment, which made use of outside contractors, the objectives and exercises for the second assessment were developed through conferences organized by the National Assessment of Educational Progress. The conference procedure was intended to give the assessment greater flexibility, involve more professionals in mathematics and education and be more efficient in cost and time.

Several types of consultants participated in the developmental conferences: college or university mathematics educators, mathematicians, classroom teachers and interested lay citizens. The objectives had to be acceptable to these groups. The mathematics experts generally had to agree that the objectives were worthwhile and important to assess. The classroom teachers had to consider the objectives to be desirable teaching goals in most schools. Finally, the objectives had to be considered desirable by the lay citizens. The lay group, including parents and others with an interest in education, had to agree that an objective be important for America's youth to achieve and that it be of value in today's society.

The objectives were organized into a content-by-process matrix (see Figure 1). This matrix resembles the classification scheme developed for the first assessment and was used extensively in the developmental process. Fewer, but more inclusive, content and process headings in the matrix for the second assessment resulted in fewer cells. The complicated task of exercise development was subdivided into units corresponding to the cells of the matrix. The cells were weighted in proportion to their importance. The number of exercises to be assessed in each cell was determined by the relative weights, and review and selection were done by cells.

After the initial objectives matrix was developed, an advisory board was formed to give direction and advice to the National Assessment staff for further refinement of the objectives and the development of the assessment. The six-member advisory board included three university mathematics educators, two mathematicians and a mathematics teacher. This board was instrumental in organizing the final set of objectives, planning the development of exercises, selecting the final exercises and planning the subsequent reports.

FIGURE 1. Objectives Framework for the Second Assessment

PROCESS	CONTENT: A. Numbers and Numeration	B. Variables and Relationships	C. Shape, Size and Position	D. Measurement	E. Other Topics
I. Mathematical knowledge					
II. Mathematical skill					
III. Mathematical understanding					
IV. Mathematical application					

The first task of the advisory board was to review the new objectives matrix and put it into final form. The board and NAEP staff thought the objectives should be used as a plan or framework for exercise development and for reporting. They thought a reporting scheme should exist prior to exercise development to help organize and improve the comprehensiveness of the second assessment. The advisory board devised a set of questions that related to each of the four cognitive processes (knowledge, skill, understanding and application) described in Chapter 3. For example, under mathematical skills, one of the questions was "How well can students perform computation?" Under mathematical applications, one of the questions was "How well can students solve typical textbook problems?" Each series of questions was intended to be "answered" by the results from the assessment of a set of exercises. This planning helped insure that the questions could be adequately covered by the assessment. These questions and more detailed information on the objectives appear in *Mathematics Objectives, Second Assessment* (1978).

While the objectives were being formulated, conferences were held to discuss special topics reflecting current trends in mathematics education. A special topic that received considerable attention was the measurement of attitudes toward mathematics. Attitudes were not measured during the first mathematics assessment because of the difficulty of developing adequate exercises. However, consultants for the second assessment encouraged the development of attitudinal exercises, and an effort was made to develop such measures.

Assessment Results and Reports

Approximately one-third of the exercises from the second assessment were released, many of them appearing as examples in four reports on the results of the second assessment. The reports, reflecting the objectives matrix, are *Mathematical Knowledge and Skills* (1979), *Mathematical Understanding* (1979) and *Mathematical Applications* (1979). A fourth report, *Changes in Mathematical Achievement, 1973-78* (1979), discusses the changes in mathematical achievement during the five years between the first and second assessments.

The text of each released exercise and accompanying documentation including selected results can be found in *The Second Assessment of Mathematics, 1977-78, Released Exercise Set* (1979). Summary data augmenting the four selected reports mentioned above appear in *Mathematics Technical Report: Summary Volume* (1980). This report includes information on mean performance levels on various sets of items for the nation and various population subgroups. A detailed description of the developmental process, sampling, data collection, scoring and data analysis can be found in *Procedural Handbook: 1977-78 Mathematics Assessment* (1980). For researchers who are interested in doing further analyses of the assessment data, data tapes containing respondent-level data for all exercises in the second assessment are available.[1]

In addition to the reports produced by National Assessment, interpretive articles aimed primarily at teachers were written by a panel of mathematics educators from the National Council of Teachers of Mathematics (NCTM). The NCTM panel presented summaries of its findings in the April 1980 issue of *The Arithmetic Teacher* and the May 1980 issue of *The Mathematics Teacher*. Succeeding issues contain brief articles on specific topics, suggesting ways mathematics teachers might use the implications of the results in their teaching. Titles of these articles appear in the references listed at the back of this booklet.

[1]Data tapes are available, at a charge, through the Department of User Services, National Assessment of Educational Progress, Suite 700, 1860 Lincoln Street, Denver, Colorado 80295.

CHAPTER 3

THE THIRD ASSESSMENT OF MATHEMATICS (1981-82)

The development of the objectives and assessment exercises proceeded along lines described in Chapter 2. College or university mathematics educators, mathematicians, classroom teachers and interested lay citizens contributed their views to the development of objectives and assessment exercises. A nine-member advisory committee was instrumental in reviewing the objectives and giving guidance throughout the developmental process. A list of the advisory board members and consultants who aided the developmental process appears in Appendix A.

The objectives for the third mathematics assessment are based on the framework used for the second assessment with some revisions that reflect current content and trends in school mathematics. To update the objectives, numerous mathematicians and mathematics educators reviewed the second assessment objectives and made suggestions for revisions and new content.

The objectives framework uses a two-dimensional content-by-process matrix for organizing the objectives (see Figure 2). One dimension includes five process levels and the other dimension includes six content areas. Attitudes toward mathematics, ability to use the calculator and computer literacy (which were considered "special topics" in the last assessment) have been incorporated into the matrix for the third assessment.

Content

The content domain for the third assessment of mathematics draws primarily from the current curriculum of elementary and secondary schools, although some projection of future mathematics emphases is acknowledged (for example, assessment of problem solving strategies, use of calculators and computers). Mathematics up to, but not including, calculus is included in the assessment exercises, which are classified according to the six content categories shown in Figure 2.

These content categories help to organize the domain, but are not intended to be represented equally in the assessment. Each is discussed below.

FIGURE 2. Objectives Framework for the Third Assessment

PROCESS	CONTENT A. Numbers and Numeration	B. Variables and Relationships	C. Shape, Size and Position	D. Measurement	E. Statistics/ Probability	F. Technology
I. Mathematical knowledge						
II. Mathematical skill						
III. Mathematical understanding						
IV. Mathematical application and problem solving						
V. Attitudes toward mathematics						

Numbers and Numeration

This category contains the largest number of exercises because of its importance in th curriculum. Exercises deal with the ways numbers are used, processed or written. Knowl edge and understanding of numeration and number concepts are assessed for whole num bers, common fractions, decimal fractions, integers and percents. Considerable emphasis i placed on operations. Number properties and order relations are also included. Most of th exercises included here are to be done with paper and pencil; however, in some instances students are asked to use other computational methods such as mental computation o estimation. Exercises include typical one- and two-step application problems, nonroutin problems and consumer problems. Nonroutine problems are exercises not normally taugh or encountered in the curriculum, but understandable to the age group. Consumer prob lems deal primarily with the uses of mathematics in commercial situations (for exampl the mathematics needed for buying and selling, including loans, percent, discount, financ charge and reading advertisements) and are emphasized more at the 17-year-old level tha at the two younger age levels.

An important new assessment category within numbers and numeration is estimatior There is an increasing need for students to be able to make good estimates. With the wide spread use of the calculator, students need to be able to check the reasonableness of the: answer by estimating. As consumers, they face daily decisions requiring estimation skill Two types of estimation skills are assessed: computational estimation (which involve working with numerical data alone) and application estimation (which requires workin with numerical data embedded in a real-world context).

To encourage students to estimate rather than use paper and pencil, only a few seconds ar allowed to complete each exercise. Additional information on the assessment of estimatio is included in Appendix B.

Variables and Relationships

The use of variables and relationships corresponds to an important part of the schoo mathematics curriculum. The exercises assessing skills in this area deal with the recogni tion of facts, definitions and symbols of algebra; the solution of equations and inequalitie: the use of variables to represent problem situations and elements of a number system; th evaluation and interpretation of functions and formulas; the graphing of points and lines i a coordinate system; the use of exponential and trigonometric functions; and logic. Ther are very few exercises appropriate for 9-year-olds in this category, and only a few topics ar appropriate for 13-year-olds. However, most exercises are appropriate at the 17-year-ol level, where students have had the opportunity to study algebra.

Shape, Size and Position

The exercises in this content category measure objectives related to school geometry. Th emphasis in the assessment is not on geometry as a formal, deductive system. Rather, th exercises concern plane and solid shapes, congruence, similarity, properties of triangle properties of quadrilaterals, constructions, sections of solids, basic theorems and relatior ships, and rotations and symmetry.

Measurement

The measurement exercises cover appropriate units; equivalence relations; instrumer reading; length, weight, capacity, time and temperature, perimeter, area and volume; nor standard units; and precision and interpolation. A substantial number of the measurement exercises require the use and understanding of metric units.

Probability and Statistics

This content area reflects a greater emphasis on statistics and probability in the school mathematics curriculum. The exercises assess collecting data; organizing data with tables, charts, graphs; interpreting and analyzing data; drawing inferences; making generalizations; using basic statistics; predicting outcomes and determining combinations.

Technology

The impact of new technology on school mathematics is measured in this content area by assessing the use of the calculator and computer literacy.

Calculator

The increasing availability and popularity of calculators has made it important for National Assessment to gather information on their use by students. This was begun in the 1977-78 mathematics assessment.

Five categories of exercises are identified for assessment. They are: (1) routine computation, (2) more difficult computations, (3) understanding concepts, (4) exploration and (5) application or problem solving. Some calculator activities such as understanding and exploration are more appropriate for instructional use in the classroom and are not emphasized in the assessment. Thus, of the five categories of exercises, computation, nonroutine computation and application are measured in the greatest depth. Additional information on these categories can be found in Appendix B.

Computer Literacy

An increasing number of schools have computers or computer terminals available for students' use. This led National Assessment to begin collecting computer literacy data in the 1977-78 mathematics assessment.

Thirteen- and 17-year-olds are asked to provide background information on their experiences with computers: whether they have access to one in their school, what programming language and computer topics they have studied and what specific activities and problems they have solved using a computer. The assessment of computer literacy includes measures of students' attitudes toward the uses, effect and role of computers and their knowledge of specific terms, flow charts or BASIC programs to determine the output.

A summary outline of these content categories is provided in Appendix C. The desired approximate percentages of exercises by content category and age group are shown in Figure 3.

FIGURE 3. Percentages of Exercises by Age and Content*

		Age 9	Age 13	Age 17
A.	Numbers and numeration	40%	40%	35%
B.	Variables and relationships	10	10	20
C.	Shape, size and position	10	10	10
D.	Measurement	15	15	10
E.	Probability and statistics	5	5	10
F.	Technology	10	10	10

**These percentages do not add to 100% because the attitudinal exercises are not included.*

Process

The process domain for the third assessment has five categories, as shown in Figure 4.

Like the content domain, the process domain can be used to classify either objectives of mathematics instruction or exercises to assess the learning of mathematics. Although each category suggests a type of mental process, neither objectives nor exercises falls neatly into a single process category — if only because the process has to be inferred; and different students may use different processes or different combinations of processes. Arbitrary decisions must be made in using any system of process categories. Such a system is helpful, however, in ensuring consideration of the diversity possible within a given content category.

Mathematical Knowledge

Mathematical knowledge refers to the recall and recognition of mathematical ideas expressed in words, symbols or figures. Mathematical knowledge relies, for the most part, on memory processes. It does not ordinarily require any other more complex mental processes.

FIGURE 4. Percentages of Exercises by Age and Process Level

		Age 9	Age 13	Age 17
I.	Mathematical knowledge	20%	15%	15%
II.	Mathematical skill	25	25	25
III.	Mathematical understanding	20	25	25
IV.	Mathematical application	25	30	30
V.	Attitudes toward mathematics	10	5	5

Exercises that assess mathematical knowledge require that a student recall or recognize one or more items of information. An example of an exercise involving recall would be one that asks for a multiplication fact such as the product of five and two. Another example would be an exercise asking for the statement of a mathematical relationship such as the law of cosines. An example of an exercise involving recognition would be one that presents several symbols and asks which symbol means "parallel."

Mathematical Skill

Mathematical skill refers to the routine manipulation of mathematical ideas and relies on algorithmic processes that are standard procedures leading to answers. Exercises assessing mathematical skill assume that the required algorithm has been learned and practiced. They do not require that the student decide which algorithm to use or that he or she apply the algorithm to a new situation. Such exercises aim at measuring proficiency in carrying out the algorithm rather than understanding how or why it works. Mathematical skill is assessed by exercises that require the performance of specified tasks, such as making measurements, multiplying two fractions, performing mental computations, graphing a linear equation or reading a table.

Mathematical Understanding

Mathematical understanding refers to the explanation and interpretation of mathematical knowledge and relies primarily on translation processes. The mathematical knowledge can be expressed in words, symbols or figures; and the translation may be within or between any of these modes of expression. Mathematical understanding involves memory processes of associating one item of knowledge with another.

Mathematical understanding may also require judgment in selecting the appropriate uses of different tools or processes. For example, students should understand appropriate times to use a calculator, computer, estimation or paper-and-pencil computation.

Exercises that assess mathematical understanding require that a student provide an explanation, an illustration for one or more items of knowledge or the transformation of knowledge. They do not require the application of that knowledge to the solution of a problem. An example of an exercise involving explanation is one that asks why a certain graph is not the graph of the function. Exercises involving transformation might ask for a drawing of an array to represent six times seven or ask for an equation to represent the information in a word problem.

Mathematical Application and Problem Solving

Mathematical application and problem solving refer to the use of mathematical knowledge, skill and understanding in solving both routine and nonroutine problems. Mathematical application and problem solving rely on memory and algorithmic, translative and judgmental processes. The student is not told how to solve the problem; reasoning and decision-making processes must be used.

Exercises that assess mathematical application and problem solving require a sequence of processes that relate to the formulation, solution and interpretation of problems. The processes may include recalling and recording knowledge, selecting and carrying out algorithms, making and testing conjectures, and evaluating arguments and results.

Exercises assessing mathematical application may vary from routine textbook problems to exercises dealing with mathematical arguments. An exercise might require the solution of a standard problem or proportion, the demonstration that two geometric figures are congruent, an estimate of the amount of carpet needed for a room, or the formulation of a problem, given a graph of statistical data.

Exercises assessing problem solving require the use of strategies in solving nonroutine problems. Students may need to be able to use such strategies as drawing diagrams, trial and error, modeling, simplification and estimation in order to correctly solve the problems.

Attitudes Toward Mathematics

National Assessment assessed attitudes toward mathematics for the first time during the second assessment in 1977-78, and these exercises are reassessed in the third mathematics assessment. Five categories of attitudinal measures were developed: (1) mathematics in school, (2) mathematics and oneself, (3) mathematics and society, (4) mathematics as a discipline and (5) attitudes toward computers. These categories were not developed as scales but rather as sets of exercises with related content. Additional information on the assessment of attitudes appears in Appendix B.

The desired approximate percentages of exercises by age group and process category are shown in Figure 4.

Questions To Be Answered by the Assessment

In the development and selection of exercises for the assessment, care was taken to assure an appropriate balance of emphasis on both the content and process dimensions. Achievement of this balance was facilitated by a set of questions organized according to the categories of the process dimension. The questions were based on the combined priorities of the interested public, mathematicians, mathematics educators (including teachers) and educational administrators.

I. Mathematical knowledge

A. How well can students recall and recognize facts, definitions and symbols?

II. Mathematical skill

A. How well can students perform paper-and-pencil computations, including computations with whole numbers, integers, fractions, decimals, percents, and ratios and proportions?

B. How well can students perform algebraic manipulations?

C. How well can students perform geometric manipulations like constructions and spatial visualizations?

D. How well can students make measurements?

E. How well can students read graphs and tables?

F. How well can students compute statistics, probabilities or combinations?

G. How well can students perform mental computations, including computation with whole numbers, fractions, decimals and percents?

H. How well can students estimate the answers to computations and measurements?

I. How well can students perform computations involving whole numbers, decimals, fractions and percents using calculators?

J. How well can students read flow charts or basic computer programs?

III. Mathematical understanding

A. How well can students translate a verbal statement into symbols or a figure, and vice versa?

B. How well do students understand mathematical concepts and principles?

C. How well can students select the appropriate uses of computers?

D. How well can students select an appropriate computational method such as paper and pencil, mental, estimation or calculator?

IV. Mathematical application

A. How well can students solve routine textbook problems?

B. How well can students solve nonroutine problems?

C. How well can students apply problem-solving strategies?

D. How well can students estimate the answers to application problems?

E. How well can students interpret data and draw conclusions?

F. How well can students use mathematics, including logic, in reasoning and making judgments?

G. How well can students use a calculator to solve application problems?

V. Attitudes

A. How do students feel about the mathematics they encounter in school?

B. How do students feel about the various activities in mathematics classes?

C. How do students feel about their personal experience with mathematics?

D. What are students' beliefs about the nature of mathematics as a discipline?

E. What are students' beliefs about the value of mathematics to society?

F. What are students' beliefs about computers?

The development and selection of exercises is primarily organized around these questions, which express the main objectives of the third mathematics assessment and will serve to organize the reports of assessment results.

To answer these questions as comprehensively as possible, sets of related or "nested" exercises appear in the assessment. For example, the same numbers may be used in a computational exercise and in an application exercise, or identical data may be provided in several different formats, or an intermediate step in a multistep problem may be assessed separately in another exercise. Nested exercises are an attempt to identify the mathematical processes that cause students difficulty.

APPENDIX A

CONSULTANTS FOR DEVELOPMENT

Advisory Committee Members

Carl, Iris M., Houston Independent School District, Houston, TX
Edwards, Edgar L. Jr., State Department of Education, Richmond, VA
Flores, Emelina G., San Jose Unified School District, San Jose, CA
Hill, Shirley A., University of Missouri, Kansas City, MO
Honicky, Donald U., AT&T, New York, NY
Keyser, Tamara, Grapevine Elementary School, Vista, CA
Koch, Patricia A., Ford Junior High School, Brook Park, OH
Kreider, Donald L., Dartmouth College, Hanover, NH
Wilson, James W., University of Georgia, Athens, GA

Consultants

Alley, Janet C., National Congress of Parents and Teachers, Chester, VA
Anderson, Ronald E., University of Minnesota, Minneapolis, MN
Anthony, Marguerite, College Hills Elementary School, College Station, TX
Armstead, James, Francis C. Hammond Jr. High School, Alexandria City School Board, Alexandria, VA
Bestgen, Barbara J., Parkway School District, St. Louis, MO
Blume, Glendon W., University of Iowa, Iowa City, IA
Boucree, Olympia E., New Orleans Public Schools, New Orleans, LA
Busic, Edward L., Richmond, VA
Carpenter, Thomas P., University of Wisconsin, Madison, WI
Carry, L. Ray, University of Texas, Austin, TX
Carter, Heather, University of Texas, Austin, TX
Chambers, Donald L., Wisconsin Department of Public Instruction, Madison, WI
Choate, Stuart A., Midland Public Schools, Midland, MI
Clark, Robey J., Portland Public Schools Indian Education Project, Portland, OR
Clemens, J. Philip, Cherry Creek High School, Englewood, CO
Coburn, Terrence, Oakland Schools, Pontiac, MI

Corbitt, Mary Kay, University of Kansas, Lawrence, KS
Cowick, Ronald M., Houston, TX
Crosswhite, F. Joe, Ohio State University, Columbus, OH
Crowley, Jill F., First National Bank of Boston, Boston, MA
Davis, Edward J., University of Georgia, Athens, GA
Davis, Robert B., University of Illinois, Urbana, IL
Dean, Jerilyn, Jefferson County Schools, Lakewood, CO
Dilworth, Robert P., California Institute of Technology, Pasadena, CA
Downs, Floyd, Hillsdale High School, Hillsdale, CA
Dunnagan, Catherine B., Control Data Corporation, Minneapolis, MN
Edwards, Steven Craig, Flint Public Schools, Flint, MI
Egbers, Elden B., Mathematics/Metrics, Olympia, WA
Eisenberg, Theodore, Northern Michigan University, Marquette, MI
Flexer, Abraham S., Boulder, CO
Flexer, Roberta, University of Colorado, Boulder, CO
Frissell, Lee, Consumers Education & Protective Association, Philadelphia, PA
Furtado, Thomas, Pratt and Whitney Aircraft, East Hartford, CT
Ganley, Marilyn G., East Hartford Public Schools, East Hartford, CT
Graeber, Anna O., Research for Better Schools, Philadelphia, PA
Greenes, Carole E., Boston University, Boston, MA
Gross, Marguerite S., San Jose Unified District Advisory Committee, San Jose, CA
Hale, Jerry Jean, Southmoor Elementary School, Denver, CO
Harner, Carol F.H., University of Denver, Denver, CO
Hernandez, Norma G., University of Texas at El Paso, El Paso, TX
Herriot, Sarah T., Henry M. Gunn Sr. High School, Palo Alto, CA
Hewitt, Susan M., Marquette Public Schools, Marquette, MI
Holloway, Linda J., Dublin High School, Dublin, CA
Johnson, Alzena B., Clarke County Public Schools, Athens, GA
Johnson, Larry Robert, Hilsman Middle School, Athens, GA
Johnson, Martin L., University of Maryland, College Park, MD
Kantowski, Mary Grace, University of Florida, Gainesville, FL
Kilpatrick, Jeremy, University of Georgia, Athens, GA
Klitz, Ralph H. Jr., Cincinnati Country Day School, Cincinnati, OH
Lappan, Glenda, Michigan State University, East Lansing, Mi
LeBlanc, John F., Indiana University, Bloomington, IN
Lebo, Carol D., Cranbrook School, Bloomfield Hills, MI
Lew, Linda Fay, Richmond Community High School, Richmond, VA
Lindquist, Mary M., National College of Education, Evanston, IL
Mack, Rita Brooks, IBM, Manassas, VA
McClintock, Edwin, Florida International University, Miami, FL
McCluskey, Murton L., Indian Education, Great Falls, MT
McKillip, William D., University of Georgia, Athens, GA

McNair, Rose Marie, Santa Cruz, CA
Morgan, William H., St. Andrews Presbyterian College, Laurinburg, NC
Moser, James M., University of Wisconsin, Madison, WI
Musser, Gary L., Oregon State University, Corvallis, OR
Myers, Warren S., H.S. District No. 214, Arlington Heights, IL
Ockenga, Earl, University of Northern Iowa, Cedar Falls, IA
Osborne, Alan, Ohio State University, Columbus, OH
Perazzo, Pauline E., California Demonstration Program in Mathematics, Herbert Hoover Jr. High, San Jose, CA
Pouncy, Earnest, Flint Community Schools, Flint, MI
Powell, Leonard, Martin Marietta Aerospace, Denver, CO
Rapley, Frank E., Jefferson County Schools, Louisville, KY
Reys, Robert E., University of Missouri, Columbia, MO
Robertson, Jack M., Washington State University, Pullman, WA
Rogers, Margaret, Yakima School District No. 7, Yakima, WA
Romberg, Thomas, University of Wisconsin, Madison, WI
Rosenberger, Naomi, University of Colorado, Colorado Springs, CO
Rowan, Thomas E., Montgomery County Public Schools, Rockville, MD
Sachar, Jane, Rand Corporation, Washington, DC
Salinas, Frank A., F.M. Black Middle School, Houston, TX
Schoen, Harold L., University of Iowa, Iowa City, IA
Scott, Wayne R., Michigan Department of Education, Lansing, MI
Sharpe, Glyn H., Jefferson County Schools, Lakewood, CO
Shaw, Betty VandenBosch, Flint Community Schools, Flint, MI
Shulte, Albert P., Oakland Schools, Pontiac, MI
Sokol, Louis F., U.S. Metric Association, Boulder, CO
Spencer, Janet Young Man, Education Professionals for Indian Children, Tahlequah, OK
Strausz, Richard P., Farmington Public Schools, Farmington, MI
Sullivan, Frances E., Jackson State University, Jackson, MS
Sutton, Anthony L., Bell & Howell Education Group, Inc., Evanston, IL
Swadener, Marc, University of Colorado, Boulder, CO
Swafford, Jane O., Northern Michigan University, Marquette, MI
Swift, Jim, Nanaimo Senior Secondary School, Nanaimo, B.C.
Tarr, John, University of Northern Iowa, Cedar Falls, IA
Taylor, B. Ross, Minneapolis Public Schools, Minneapolis, MN
Thomas, James R., Ishpeming Public Schools, Ishpeming, MI
Thompson, Alba G., San Diego State University, San Diego, CA
Tittle, Carol Kehr, University of North Carolina at Greensboro, Greensboro, NC
Tobin, Alexander, Mathematics Education, School District of Philadelphia, Philadelphia, PA
Trafton, Paul R., National College of Education, Evanston, IL
Travers, Kenneth J., University of Illinois, Urbana, IL
Usiskin, Zalman, University of Chicago, Chicago, IL

Van Beynen, John G., Northern Michigan University, Marquette, MI
Weaver, J.F., University of Wisconsin-Madison, Madison, WI
Wheatley, Grayson H., Purdue University, West Lafayette, IN
Whitaker, Karen, Evanston, IL
Zimpfer, Richard P., Bayside Planning & Engineering, Houston, TX

APPENDIX B

DETAILS OF OBJECTIVES FOR THE THIRD MATHEMATICS ASSESSMENT

For the interested reader, more information is provided here on certain assessment topics in the third assessment. Topics include assessment of attitudes, the calculator and estimation skills.

Attitudes

The attitudinal measures included in the third assessment were originally developed for the second assessment and will be reassessed to provide information on changes in students' attitudes. Five categories of attitudinal measures were developed.

The first category, mathematics in school, consists of exercises assessing attitudes toward the mathematics courses students have encountered in school. Subcategories of exercises include a school subject comparison, questions about the frequency of various classroom activities and students' attitudes toward those activities, and a breakdown of mathematics content activities.

The subject-comparison exercises list the most commonly encountered school subjects (e.g., science, mathematics, social studies) and ask respondents to rate each subject on the basis of whether they like or dislike it, whether they find it easy or hard and whether or not they view it as important.

Students are presented with a list of activities that might occur in a mathematics classroom and are asked to indicate how often (often, sometimes, never) they have participated in such activities. They are also asked to state whether they like or dislike each activity and whether they find the activity useful in helping them learn mathematics. Typical exercises are the following:

Last year in your mathematics course how often did you

A. watch the teacher work mathematics problems on the board?

Often ○ Sometimes ○ Never ○

B. get individual help from the teacher on your mathematics?

Often ○ Sometimes ○ Never ○

C. help a classmate do mathematics?

Often ○ Sometimes ○ Never ○

How do you feel about each of these activities in learning mathematics? First, how much do you like or dislike them? Second, how useful are they in learning mathematics? Indicate your feeling by filling in one oval on each line.

A. Taking mathematics tests

Like A Lot ○ Like ○ Undecided ○ Dislike ○ Dislike A Lot ○

Very Useful ○ Useful ○ Undecided ○ Not Very Useful ○ Useless ○

B. Doing mathematics homework

Like A Lot ○ Like ○ Undecided ○ Dislike ○ Dislike A Lot ○

Very Useful ○ Useful ○ Undecided ○ Not Very Useful ○ Useless ○

Respondents are also given a list of content activities and asked to respond on an important/not-important, easy/hard, like/dislike basis. A sample exercise is:

> How do you feel about each of these mathematics activities? First, how important are they? Second, how easy are they? Third, how much do you like them? Indicate your feeling by filling in one oval on each line.

A. Solving word problems

Very Important	Important	Undecided	Not Very Important	Not Important At All
Very Easy	Easy	Undecided	Hard	Very Hard
Like It A Lot	Like It	Undecided	Dislike It	Dislike It A Lot

B. Working with fractions

Very Important	Important	Undecided	Not Very Important	Not Important At All
Very Easy	Easy	Undecided	Hard	Very Hard
Like It A Lot	Like It	Undecided	Dislike It	Dislike It A Lot

C. Estimating answers to problems

Very Important	Important	Undecided	Not Very Important	Not Important At All
Very Easy	Easy	Undecided	Hard	Very Hard
Like It A Lot	Like It	Undecided	Dislike It	Dislike It A Lot

Exercises within the "mathematics and oneself" category assess a respondent's perception of himself or herself in relation to mathematics. Different attitudinal components such as anxiety, motivation, self-concept and enjoyment of mathematics are reflected in these Likert-type exercises. A sample exercise follows:

> This exercise asks how you feel about mathematics or mathematics activities. There are no correct answers. The answer choices are "True about me," "Sometimes true about me" and "Not true about me." For each part, choose the one response that best describes how you feel about the statement. Be sure to fill in one oval in each box.

A. Mathematics is boring for me.

True About Me	Sometimes True About Me	Not True About Me
O	O	O

B. I usually understand what we are talking about in mathematics.

True About Me	Sometimes True About Me	Not True About Me
O	O	O

C. Doing mathematics makes me nervous.

True About Me	Sometimes True About Me	Not True About Me
O	O	O

The four examples above are for 9-year-olds. For 13- and 17-year-olds, the exercises have five response options, from "Strongly Disagree" to "Strongly Agree."

The "mathematics and society" category reflects two major concerns: the value of mathematics to the individual as a member of society and the value of mathematics to society in general. Likert-type exercises are designed to assess attitudes toward both the usefulness and importance of mathematics to society. A typical exercise follows:

> This exercise asks how you feel about mathematics or mathematics activities. There are no correct answers. The answer choices are "Strongly Disagree," "Disagree," "Undecided," "Agree" or "Strongly Agree." For each part, choose the one response that best describes how you feel about the statement. Be sure to fill in one oval in each box.

A. Most people do not use mathematics in their jobs.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
⬭	⬭	⬭	⬭	⬭

B. I would like to work at a job that lets me use mathematics.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
⬭	⬭	⬭	⬭	⬭

Attitudes toward mathematics as a field of study are the focus of the exercises in the "mathematics as a discipline" category. Students' views toward mathematics as a cumulative or compartmentalized subject, the status of mathematics as a fixed or changing subject and mathematics as a process are among the different aspects of attitude assessed in this category. Items such as the following are included:

This exercise asks how <u>you feel</u> about mathematics or mathematics activities. There are no correct answers. The answer choices are "Strongly Disagree," "Disagree," "Undecided," "Agree" or "Strongly Agree." For each part, choose the one response that <u>best</u> describes how <u>you</u> feel about the statement. Be sure to fill in one oval in each box.

A. Mathematics is made up of unrelated topics.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
⬭	⬭	⬭	⬭	⬭

B. Mathematics helps one to think logically.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
⬭	⬭	⬭	⬭	⬭

C. Doing mathematics requires lots of practice in following rules.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
⬭	⬭	⬭	⬭	⬭

Calculator

Like the assessment of attitudes, assessment of the use of the calculator was first done during the second mathematics assessment. At that time the increasing availability and popularity of calculators made it important for National Assessment to collect information on their use by students. The small, inexpensive calculator has the potential to dramatically influence the mathematics curriculum in the United States.

Several conferences were organized by National Assessment to discuss the place of the calculator in the second mathematics assessment. The participants at these conferences identified five categories of exercises for which calculators might be used. These categories are:

1. *Routine Computation* — This category includes typical computations with whole numbers, decimals, fractions and integers that are routinely taught at a particular age.

2. *More Difficult Computations* — Students might be asked to perform difficult computations or computations for which algorithms have not been formally taught. For example, 9-year-olds might be asked to do computations with decimals or difficult division problems. Thirteen-year-olds might be given chaining operations or conversions between fractions and decimals. Students at all ages might be required to work with very large numbers or complicated decimals that would make computation without a calculator tedious.

3. *Understanding Concepts* — Nine-year-olds might use the calculator to learn more about place value; and 13-year-olds might learn estimation for order of magnitude. Seventeen-year-olds might use the calculator to facilitate learning order of operations, approximating square roots, graphing functions and understanding properties of functions.

4. *Exploration* — Exercises in this category might deal with topics in number theory including series, summations, patterns or divisibility problems.

5. *Applications and Problem Solving* — This category includes routine and more difficult word problems and also multistep problems. For 9-year-olds, some problems might use larger numbers than they are accustomed to working with. Thirteen-year-olds might work problems involving percent, unit pricing and other more difficult word problems. Seventeen-year-olds might be given a variety of realistic consumer problems as well as exercises dealing with mathematical formulas.

In an effort to collect data on the use of the calculator, as wide a variety of topics as possible was assessed. Although it was agreed that the five categories should be assessed at each

Sample exercises for computation estimation are:

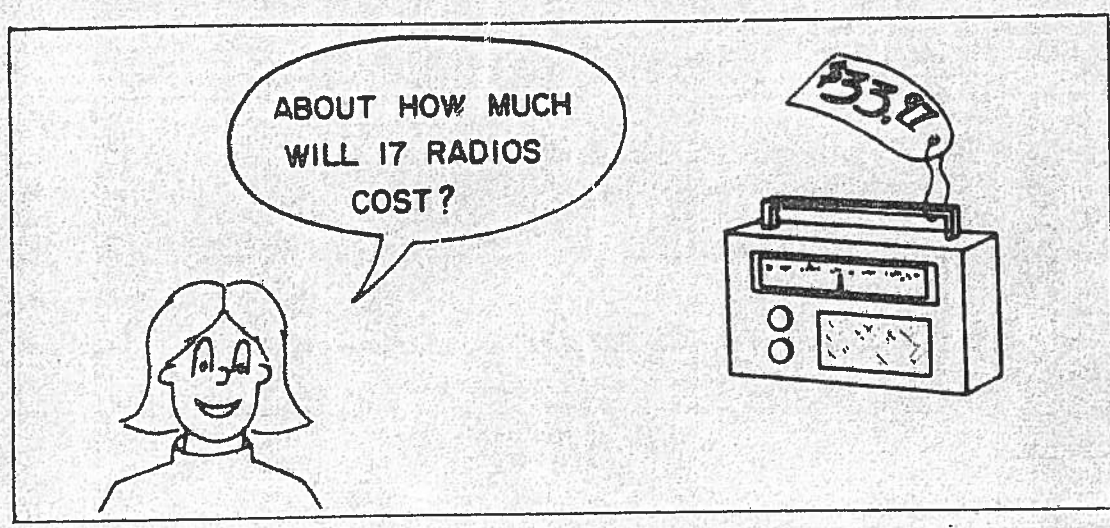
A. 48 X 25 X 12

Estimate ______________

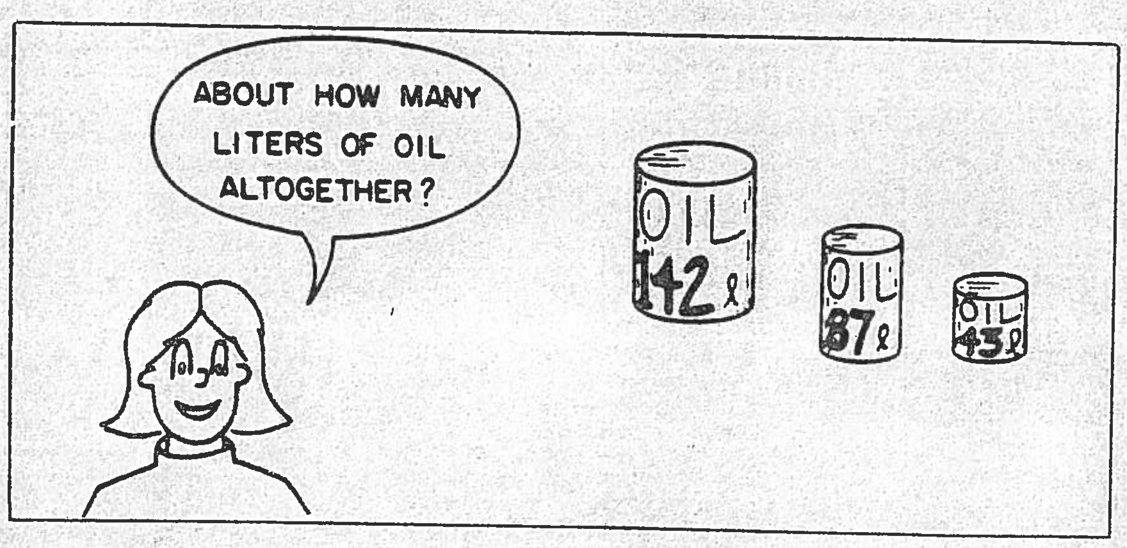
B. 613 - 490

- O 100
- O 200
- O 300
- O 400
- O 1,200
- O I don't know.

Sample exercises for application estimation are:



Estimate ______________



- 200
- 280
- 350
- 1,400
- I don't know.

An effort will be made to relate the arithmetic required in the mental computation exercises with the arithmetic needed in the computation and application estimation exercises. For example, a mental computation exercise might be 20 x 40. The related computation estimation exercise might be 19 x 42, which the student would probably solve by rounding each number to the nearest 10. A related application estimation exercise might ask students to estimate the area of a rectangle with a width of 19 and a length of 42. By embedding related or identical numbers in different types of problems, more information can be obtained on students' estimation skills.

APPENDIX C

CONTENT OUTLINE

A. Numbers and Numeration

1. Numeration (whole numbers, fractions, decimals, percent, integers, scientific notation)

2. Number concepts (whole numbers, fractions, decimals, percent, integers)

3. Operations (whole numbers, fractions, decimals, percents, integers)

4. Mental computation

5. Estimation

6. Properties

7. Relations

B. Variables and Relationships

1. Facts, definitions and symbols

2. Use of variables in equations and inequalities (solutions, equivalences and translations)

3. Operations with variables

4. Use of variables to represent elements of a number system

5. Functions and formulas

6. Coordinate systems

7. Exponential and trigonometric functions

8. Logic

C. Shape, Size and Position

1. Recognition of figures

2. Constructions and drawings

3. Visualization (static and dynamic)

4. Recognition of relationships (congruence, similarity and symmetry)

5. Identification of properties from given visual information within, between or among figures

6. Relationships involving classes of figures

7. Definitions, postulates and theorems (recall, inference and application)

D. Measurement

1. Unit (appropriate size and type of unit, unit equivalents, conversions within a system)

2. Instrument reading (English and metric rulers, scales, thermometers, clocks, etc.)

3. Linear measure (including nonstandard units)

4. Area, perimeter and volume

5. Precision

6. Estimation of measurements

E. Probability and Statistics

1. Organizing, displaying and interpreting information (tallies, graphs, charts and tables)

2. Measures of central tendency (mean, median, mode)

3. Measures of spread and position (range, percentile, standard deviation)

4. Sampling and polling

5. Probability (simple, compound and independent events; odds)

6. Combinations and permutations

F. Technology

1. Hand calculator

2. Computer literacy

Reporting Relevant Progress: Exploring Dynamic Frameworks for NAEP

Overview

According to the NAEP statute (P.L. 107-279), the Governing Board is responsible for developing assessment objectives and test specifications for each NAEP subject area. Since 1989 the Governing Board has developed assessment frameworks and specifications in more than 10 subjects through comprehensive, inclusive, and deliberative framework projects. The Board's Framework Development Policy is included in this attachment.

Three models have been used to account for the need to update framework content over time:

1. New Framework/Start New Trend

In some cases, the Board has determined through research, outreach, content and policy input, and other means that a new framework is warranted in a subject area. In these subject area assessments, the new assessment framework defines a new construct, includes different content and skills, adds new item types, changes the assessment delivery mode (i.e., digital-based assessment (DBA)), and other modifications. Examples of this model include 2011 NAEP Writing, where the new construct was writing on a computer and using word processing tools. This was judged to represent a different construct from writing in the previous framework's paper and pencil assessment. The new construct definition motivated a break in trend reporting from the old assessment's results. A similar break in trend occurred for the 2009 NAEP Science Framework, which reflected several enhancements from advancements in science and science curricular standards, such as crosscutting content and deeper integration of science practices.

2. New Framework/Maintain Trend

In this model, the new framework is designed to be different in many ways from the previous framework; however, empirical investigation reveals that the construct does not differ substantially. The interest in maintaining trend prompts linking studies and other research to try to ensure trend lines can be maintained. Board adoption of the 2009 NAEP Reading Framework was under similar circumstances as the new frameworks for NAEP Writing and NAEP Science, because the old NAEP Reading Framework had several sub-elements that were no longer relevant to the field's conceptualization of reading comprehension. This framework update occurred during the No Child Left Behind Act (NCLB) era. Given the NCLB statute's requirement to use NAEP as a monitoring tool for states, there was substantial interest in establishing a bridge to maintain the reading trend despite changes to the construct being measured on NAEP. Empirical investigation revealed that trend reporting could be maintained, and so the NAEP Reading trend remained intact from its beginning in 1992.

3. Updated Framework/Maintain Trend

This model is defined by gradual changes to a framework over time so that trend is maintained. For mathematics, the framework has been "tweaked" over time to more clearly define the objectives, shift content emphases, and refine the process dimension while not redefining the construct. NAEP has been able to maintain the mathematics trend line for grades 4 and 8 since 1990. The framework "tweaks" have occurred sporadically rather than on an ongoing basis, often prompted by less dramatic but important curricular and assessment advances for a subject area. A more ongoing and systematic model for these updates could be included in the concept of dynamic frameworks.

Dynamic Framework Model

The Board's Strategic Vision, adopted at the November 2016 quarterly meeting, includes a goal to:

> Develop new approaches to update NAEP subject area frameworks to support the Board's responsibility to measure evolving expectations for students, while maintaining rigorous methods that support reporting student achievement trends.

This description in the Strategic Vision suggests a fourth model for making continuous, gradual changes to NAEP frameworks using empirical evidence to avoid compromising the ability to maintain trend. This more systematic and ongoing approach to updating assessment content is novel and has been referred to as *dynamic frameworks*. First described in The Future of NAEP (attached), a dynamic framework incorporates continuous, incremental changes to content rather than periodic abrupt shifts in content.

According to The Future of NAEP:

> "Dynamic frameworks would balance dual priorities of trend integrity and trend relevance. As an analogy, the Consumer Price Index (CPI) tracks inflation by deliberately conflating two concepts: change in the cost of a fixed basket of goods and change in the composition of the basket itself. As time passes, an increase in the cost of a product that is no longer relevant should contribute less to estimated inflation. By adopting dynamic frameworks, NAEP would similarly conflate increases in student proficiency with a change in the definition of proficiency itself. Although this conflation may seem undesirable, it may be the best way to balance desires for both an interpretable trend and a relevant trend" (p. 17).

There are several issues and questions that need to be resolved before the Governing Board can make a determination about the feasibility of a dynamic framework model. Issues related to the reasons for updating frameworks and what content to add or delete is in the domain of the Assessment Development Committee (ADC). Issues related to the speed of change and methods for maintaining trend with continuous, incremental changes to content would be in the domain of the Committee on Standards, Design and Methodology (COSDAM) and the National Center for Education Statistics (NCES).

During the November 2016 quarterly Board meeting, the ADC and COSDAM met to begin discussing how to approach the idea of dynamic frameworks. An excerpt of the minutes from that joint committee meeting discussion is included in this attachment. The committees agreed that additional time was needed to discuss how to approach a dynamic framework model.

The following suggested discussion questions are provided to support the March 2017 joint session of the committees and address some of the issues involved in pursuing this approach.

Discussion questions

1. What are the conceptual differences between the dynamic framework and the other 3 models (New Framework/New Trend; New Framework/Maintain Trend; Updated Framework/Maintain Trend) that the Board and NCES have employed in the past?

   Some possible factors to consider:
   - The extent to which proposed content changes are judged to represent a new construct
   - An a priori decision about the importance of maintaining trend, versus deciding whether trend can be maintained post hoc
   - The scope of proposed content changes
   - The speed of proposed content changes
   - The extent to which the approach distinguishes between adding versus dropping objectives
   - The extent to which the operationalization of the frameworks (i.e., item specifications and item pools) changes over time

2. What does an assessment development schedule currently look like?
   - The Board, through ADC, currently decides on the scope of proposed content changes through contractors that convene educators, parents, and the general public, for active and broad participation. The current framework development policy states that frameworks and test specifications shall remain stable for at least 10 years.
     - Can we make the scope of changes and related outreach more continuous, smooth, and systematic?
   - Under current operational procedures, it takes approximately 4.5 years between Board adoption of a framework (or changes to a framework) and NCES development and administration of new items under that framework.
     - Can we make this transition more continuous, smooth, and systematic?
   - With each operational administration of an assessment, several items are released and replacement items are developed by NCES. The released and replaced items may vary somewhat in terms of the objectives covered.
     - Are there implications for this process under a dynamic framework model?

3. What are the “must haves” for dynamic frameworks?
   - For example, should we posit that we must document and communicate any changes to the public and various stakeholders clearly and by a certain time?
   - For example, should we commit to upholding the current framework development policy to include the active participation of educators, parents, and members of the general public?

4. Would the possible updates being considered for the NAEP Mathematics Framework be a good time to try out a dynamic framework model?

   Possible next steps related to assessment content:

   - Determine whether there is a compelling rationale to pursue content updates
   - Determine which objectives (if any) should be dropped (procurement underway)

- o Determine which objectives (if any) should be added (procurement underway)
- o Identify how to ensure content updates are determined through an inclusive and deliberative process with active participation from states, educators, and parents
- o Determine how quickly a revised framework can be developed and adopted
- o Gather information about the number of items related to dropped objectives that have appeared on recent NAEP assessments
- o Determine how quickly an assessment can be administered based on a revised framework that drops some objectives
- o Determine how quickly an assessment can be administered based on a revised framework that adds some new objectives
- o Determine whether each increment of change is meaningful and defensible from a content perspective

Possible next steps related to methodology:

- o Reassess Board commitment to maintaining trends in NAEP Mathematics in 2021 and beyond
- o Determine how (if at all) recent NAEP results would have been different if the assessment had not included items associated with objectives that will be dropped
- o Determine what factors affect the speed with which a framework can be revised while still maintaining trends:
    - i. Number or proportion of objectives to be deleted
    - ii. Number or proportion of objectives to be added
    - iii. Number or proportion of items associated with deleted/added objectives
    - iv. Difficulty of items associated with deleted/added objectives
    - v. Potential changes to other aspects of the framework (e.g., cognitive processes)
    - vi. Other
- o Consider implications of content changes for achievement levels

Except of minutes from November 2016 joint session on Dynamic Frameworks

Mary Crovo provided an overview with historical context about ways in which the Board has changed frameworks while maintaining or breaking trend lines. In these instances, NAEP has either continued to report trends on new assessment results connecting with previous results or started a new reporting trend relative to previous assessment results. She noted that NAEP's practice has been to reflect broad-based input from many stakeholders. Ms. Crovo summarized there are three different ways that NAEP has dealt with framework changes: starting a new framework and breaking the trend line for the assessment results; starting a new framework and maintaining the trend line connecting to the previous framework; and implementing smaller framework updates while maintaining the trend line.

Ms. Crovo also reviewed the current timeline for development of an assessment, from framework development to reporting of results. Joe Willhoft made a note of the long lead time of nearly 4.5 years between a framework's completion and the final operational assessment being administered, but Ms. Crovo noted that smaller or more incremental framework changes could shorten this timeline with fewer items to develop.

As part of this session, the Committees also heard a presentation from Dan McGrath of NCES to summarize how NCES has considered the concept of dynamic frameworks for NAEP as part of the NCES Future of NAEP initiative, and how international assessments have approached this concept of updating frameworks.

Cary Sneider noted that the Board could foreseeably identify rationales for shifting the percentages of content or having content that repeats in multiple grades. For example, such changes could address cases where there are NAEP alignment issues resulting primarily from different sequencing of content across grades, and these changes provide helpful information on how learning progresses on the same content, from grade 4 to 8.

Lucille Davy noted that the grade 4 NAEP Mathematics Assessment has some content most students are not learning by the 4th grade, as indicated by several states' adoption of Common Core State Standards. She acknowledged the need to study how much change is too much and to study the ideal rate of change over time, in order to optimize both measurement of student performance and relevance to education policy.

Dale Nowlin commented that even when we do not change the measure, i.e., the assessment, what is being measured is changing. The NAEP Writing Assessment shows this clearly—the current NAEP Writing Framework reflects a construct focused on writing in a digital environment with common word processing tools, but if NAEP continued to assess students in the traditional paper-pencil format today, the assessment would not collect the same information compared to the student performance data gathered from the last paper-pencil assessment because this is increasingly not the way students write.

In addition to the rate of implemented changes, the Committees noted several issues that need to be carefully considered and balanced. Mitchell Chester suggested reviewing how shifting the context of items can represent desired changes, without changing the construct. Ms. Garrison noted that time limitations for assessment administrations are an important factor, as well as assuring that current NAEP items remain relevant to students in future administrations. Joe Willhoft suggested we examine how new changes may interact with general content drift over time or the accumulation of year-to-year trend inferences over time. Finally, Linda Rosen and Mr. Willhoft noted that different stakeholders may react to changes differently.

Mr. Willhoft also noted that the Board should carefully consider how communications with educators are framed so that messages do not create a sense that students are chasing a moving target, with an assessment that is constantly changing. Jim Popham encouraged the Board to promote educational progress in how the concept of dynamic NAEP assessment frameworks is defined and pursued.

Several Committee members agreed on the importance of clarifying and articulating the problem that the Board is hoping to address with a dynamic assessment framework model. Mr. Chester asked the Board to consider changes in the field that NAEP is not detecting in the current more static framework model, and whether these changes are important for NAEP to capture. Generally, the Committees agreed about the need to study how much change is too much, i.e., what level of change would potentially compromise NAEP's ability to report trends over time. Another important issue is how to implement proposed changes.

The framework updates that the Board will eventually consider for the NAEP Mathematics Assessment will be a first case where the concept of dynamic frameworks can be applied. Ms. Crovo noted that the Board is commissioning research to comprehensively survey state mathematics standards, including the 15 percent of additional state-level standards. This research will inform decisions on whether and how to change the current NAEP Mathematics Framework.

Ms. Davy also reminded the Committee that several of these issues are time sensitive to best support states, and so Board discussion should be deliberate and also reflect this urgency. Chasidy White agreed that states need guidance on these issues. The Committees requested continued joint Committee discussion to grapple with these issues and open questions, with a next meeting that focuses more on understanding current processes and considering how they could be changed.

Adopted: May 18, 2002

National Assessment Governing Board

Framework Development

Policy Statement

It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to determine the content and format of all subject area assessments under the National Assessment of Educational Progress (NAEP). Objectives developed and adopted by the Governing Board as a result of this process shall be used to produce NAEP assessments that are valid and reliable, and that are based on widely accepted professional standards. The process shall include the active participation of educators, parents, and members of the general public. The primary result of this process shall be an assessment framework to guide NAEP development at grades 4, 8, and 12.

The Governing Board, through its Assessment Development Committee, shall carefully monitor the framework development process to ensure that all Governing Board policies are followed; that the process is comprehensive, inclusive, and deliberative; and that the final Governing Board-adopted framework, specifications, and background variables documents are congruent with the Guiding Principles, Policies, and Procedures that follow.

Introduction

Since its creation by Congress in 1988, the Governing Board has been responsible for determining the content and format of all NAEP subject area assessments. The Governing Board has carried out this important statutory responsibility by engaging a broad spectrum of educators, policymakers, business representatives, and members of the general public in developing recommendations for the knowledge and skills NAEP should assess in various grades and subject areas. From this comprehensive process, the Governing Board develops an assessment framework to outline the content and format for each NAEP subject area assessment.

Under provisions of the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279), Congress has authorized the Governing Board to continue its mandate for determining the content and format of NAEP assessments by requiring that:

- "the purpose [of NAEP] is to provide…a fair and accurate measurement of student academic achievement;"
- "[NAEP shall]…use widely accepted professional testing standards, objectively measure academic achievement, knowledge, and skills, and ensure that any academic assessment authorized….be tests that do not evaluate or assess personal or family beliefs and attitudes or publicly disclose personally identifiable information;"
- "[NAEP shall]…only collect information that is directly related to the appraisal of academic achievement, and to the fair and accurate presentation of such information;"
- "the Governing Board shall develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards;"
- "the Governing Board shall have final authority on the appropriateness of all assessment items;"
- "the Governing Board shall take steps to ensure that all items selected for use in the NAEP are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological;" and
- "the Governing Board shall develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public."

Given the importance of these mandates it is incumbent upon the Governing Board, in the design, conduct, and final action on the assessment framework, to ensure that the highest standards of test development are employed. The validity of educational inferences made using NAEP data could be seriously impaired without high standards and rigorous procedures for framework development.

Historically, the task of developing the framework for a NAEP assessment has been conducted by the Governing Board through competitive procurements. It is imperative that contractors be fully informed of the Governing Board's policy regarding framework development, so that all deliverables under the contract meet statutory requirements and are acceptable to the Governing Board. The purpose of the Policy on Framework Development, therefore, is to articulate the Guiding Principles, Policies, and Procedures that will direct the framework development process.

Each of the following Guiding Principles is accompanied by Policies and Procedures. Full implementation of this framework development policy will require the appropriate framework contractor(s), to provide assurances to the Governing Board, through the Governing Board staff, that all aspects of the Policies and Procedures for which they are responsible have been successfully completed. These assurances will be in writing, and may require supporting information prepared by the contractor and/or the Governing Board staff.

This policy complies with the documents listed below which express widely accepted technical and professional standards for test development. These standards reflect the agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.

Standards for Educational and Psychological Testing. (1999). Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.

Code of Fair Testing Practices in Education. (2004). Washington, DC: Joint Committee on Testing Practices.

National Center for Education Statistics (NCES) Statistical Standards, September 2002.

Guiding Principles – Framework Development

Principle 1

The Governing Board is responsible for developing an assessment framework for each NAEP subject area. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and preliminary achievement level descriptions.

Principle 2

The Governing Board shall develop an assessment framework through a comprehensive, inclusive, and deliberative process that involves the active participation of teachers, curriculum specialists, local school administrators, parents, and members of the public.

Principle 3

The framework development process shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and assessments, and other pertinent factors and information.

Principle 4

The Governing Board, through its Assessment Development Committee, shall closely monitor all steps in the framework development process. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the assessment framework; assessment and item specifications; and background variables that relate to the subject being assessed.

Principle 5

Through the framework development process, preliminary achievement level descriptions shall be created for each grade being tested. These preliminary descriptions shall be an important consideration in the item development process and will be used to begin the achievement level setting process.

Principle 6

The specifications document shall be developed during the framework process for use by NCES and the test development contractor as the blueprint for constructing the NAEP assessment and items in a given subject area.

Principle 7

NAEP assessment frameworks and test specifications generally shall remain stable for at least 10 years.

Policies and Procedures for Guiding Principles

Principle 1

The Governing Board is responsible for developing an assessment framework for each NAEP subject area. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and preliminary achievement level descriptions.

Policies and Procedures

1. The assessment framework shall determine the extent of the domain and the scope of the construct to be measured for each grade level in a NAEP assessment. The framework shall cover grades 4, 8, and 12, where applicable, in a given subject area. The framework shall provide information to the public and test developers on three key aspects of the assessment: a) <u>what</u> should be measured; b) <u>how</u> that domain of content is most appropriately measured in a large-scale assessment; and c) <u>how much</u> of the content domain, in terms of knowledge and skills, should students know and be able to do at the basic, proficient, and advanced levels.

2. More specifically, the framework shall: a) articulate the purpose and scope of the assessment; b) define the content and skills to be tested at each grade; c) define the weighting of the item pool in terms of the content and process dimensions; d) describe the format requirements of the items and the assessment; e) include preliminary achievement level descriptions for each grade at the basic, proficient, and advanced levels; and f) contain sample items for each grade to be tested.

3. The primary audience for the assessment framework shall be the general public. Technical and subject-specific terminology should be used only when necessary, and shall be defined in the body of the framework or in a glossary. Where appropriate, the framework should use tables, charts, and graphics to clearly and concisely communicate necessary information pertaining to the various assessment elements. The framework shall contain sufficient information to inform policymakers, educators, and others about the nature and scope of the assessment in a given subject area.

4. NAEP frameworks shall continue to be developed with the active participation of states. Content coverage in each subject and grade shall be broad, inclusive of content valued by states as important to measure, and reflect high aspirations for student achievement.

5. The framework shall not endorse or advocate a particular pedagogical approach to the subject area being assessed, but shall focus on important, measurable indicators of student achievement to inform the nation about what students know and are able to do. While the framework shall not endorse pedagogy, it may facilitate reporting on various types of skills essential to achievement in the grade and subject area.

6. Where appropriate, the framework shall describe additional requirements of the assessment and administrative conditions which may be unique to a given subject area. For example, this may include a brief discussion of ancillary materials, use of technology, and other conditions.

7. Special studies, if any, to be conducted as part of the assessment in a given subject area shall be described in the framework. This description shall provide an overview of the purpose and rationale for the study, the nature of the student sample(s), and a discussion of the instrument and administration procedures.

8. Following Governing Board adoption, the framework shall be widely disseminated in print and electronic versions.

Principle 2

The Governing Board shall develop an assessment framework through a comprehensive, inclusive, and deliberative process that involves the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public.

Policies and Procedures

1. The guiding statute calls for the "active participation" of various NAEP audiences in the framework development process. Because this is a public endeavor it is important that all major constituents are represented in a fair and open process. The Governing Board's framework development process shall be comprehensive in its scope and outreach; inclusive in its involvement of broad-based panel members and reviewers; and deliberative in considering all viewpoints and debating all pertinent issues in formulating the content and design of a NAEP assessment.

2. The framework development committees shall be constituted in such a way as to be representative in terms of gender, race/ethnicity, region of the country, and viewpoints regarding the content of the assessment under development. In addition, many different views shall be sought from various segments of the population in the review of materials and in soliciting public input and feedback. The level of "active participation" shall be documented in a report of the framework development process.

3. The framework development environment shall be open, balanced, and even-handed. To the greatest extent possible, the project deliberations will be protected from inappropriate influences of various interest groups. All issues and agendas shall be considered in a careful, objective, and respectful manner by all project committees and the Governing Board.

4. Prior to implementation of the framework development process, the contractor shall identify procedures that will be used to clarify positions and views, roles and responsibilities of all project staff and committees, as well as how the process will work toward reaching an understanding of the scope, content, and design of the framework.

5. While the NAEP statute no longer requires a “national consensus process,” the Governing Board will develop frameworks through involvement of broadly representative groups and individuals with diverse viewpoints, open discussion and deliberation of issues, and careful consideration, and revision when necessary, of framework recommendations prior to final Governing Board action. The Governing Board shall make the final decision on a framework and shall not delegate decisions on the content and format of NAEP assessments.

6. It is a requirement throughout the framework development process to obtain reviews of draft materials and general public input from a wide audience of stakeholders, including content experts (outside of the framework committees), curriculum and assessment staff of state and local education agencies, users of assessment data, those who are employed in the specific content area under consideration, policymakers, parents, and the general public. The constituency of “users and consumers” mentioned above may include scientists, mathematicians, journalists, civic leaders, authors, and others.

7. Written summaries of all hearings, forums, surveys, and committee meetings shall be made available to the framework committees in a timely manner, so that such information can best inform the decisionmaking process. The Assessment Development Committee and the Governing Board shall receive written documentation and regular briefings on all project activities at their quarterly meetings.

8. Framework development panels shall consist of a policy oversight or steering committee comprised of representatives from key policy groups, business and industry, content experts, educators at the state and district level, users and consumers, parents, and the general public. At least 30 percent of this committee shall be composed of users and consumers in the subject area under consideration. Both public and private schools shall be represented on this committee.

9. The steering committee will receive the project charge directly from the Governing Board, and shall formulate guidelines for the conduct of the framework development process, consistent with statutory requirements and Governing Board policy. This oversight committee shall monitor the progress of the development work via meetings, teleconferences, and electronic communication. The final recommended documents from the project shall be reviewed by the oversight panel for recommendation to the Governing Board at the completion of the deliberative process.

10. Development of the project documents shall be the responsibility of a project planning committee composed of content experts, educators at the state and district level, curriculum specialists, university professors, policymakers, users and consumers, business representatives, and members of the public. Classroom teachers shall be well represented on this committee at all grade levels designated for the assessment under development. Teachers, administrators, and curriculum specialists shall be drawn from schools across the nation, including individuals who work with students from high-

poverty and low-performing schools. Both public and private schools shall be represented on this committee.

11. The planning committee shall carefully consider the charge from the Governing Board and guidelines set forth by the project oversight committee in developing the assessment framework. The committee shall carry out its work through meetings, conference calls, and electronic communication. It shall be responsible for developing the major deliverables of the project: the framework, specifications, and background variables documents, under the direction of project staff.

12. Where appropriate, a third committee of technical experts shall be involved in the framework development process. This committee shall consist of psychometricians, state testing experts, and individuals involved in developing assessments in the content area under consideration. It shall be this panel's responsibility to uphold the highest technical standards for development of the NAEP framework and specifications. The committee shall respond to technical issues raised during the process and provide guidance to project staff and the project committees on technical aspects of the assessment specifications. As with the steering and planning committees, the technical panel will meet in-person, via teleconference, and through electronic communication.

13. The preceding Policies and Procedures for conducting the framework development process constitute one model of committee structure. A prospective contractor may propose an alternative plan; however, the committees must be broad-based and representative of the type of groups and individuals identified above.

Principle 3

The framework development process shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and assessments, and other pertinent factors and information.

Policies and Procedures

1. The NAEP framework development process shall be informed by a broad, balanced, and inclusive set of factors. The framework shall maintain a balance between curriculum reform in a field, exemplary research regarding cognitive development and instruction, and the nation's future needs and desirable levels of achievement. This delicate balance between "what is" and "what should be" is the essence of the NAEP framework development process.

2. The framework development process shall begin by thoroughly identifying major policy and assessment, issues in the content area, to be summarized in an issues paper. The primary audiences for the issues paper are the Governing Board and the project committees. Designed to serve as a springboard for committee deliberations and framework development, this paper shall elaborate on major issues providing both pros and cons, summarize the research, and cite trends in state standards and assessments.

3. The framework panels shall consider a wide variety of resources as the deliberations proceed, including but not limited to curriculum guides and assessments developed by states and local districts, widely accepted professional standards, scientific research, other types of exemplary research studies in the literature, key reports having significant national and international interest, international standards and assessments, other assessment instruments in the content area, and prior NAEP frameworks, if available.

4. In considering the relative importance of these sources of information in developing the framework, the project committees shall consider the charge as delivered by the Governing Board, the role and purpose of NAEP in informing the public about student achievement, constraints of a large-scale assessment, technical assessment standards, issues of burden and cost-effectiveness in designing the assessment, and other factors unique to the content area.

Principle 4

The Governing Board, through its Assessment Development Committee, shall closely monitor all steps in the framework development process. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the assessment framework; assessment and item specifications; and background variables that relate to the subject being assessed.

Policies and Procedures

1. When the framework development process is conducted for the Governing Board by an outside contractor, the process shall be managed in an efficient, cost-effective manner, shall be completed in a timely fashion, and shall adhere to sound measurement practice.

2. The Governing Board's Assessment Development Committee (ADC) shall be responsible for monitoring the framework development process that results in recommendations to the Governing Board on the content and format of each NAEP assessment. Direction will be provided to the framework development contractor by the ADC and the Governing Board, via Governing Board staff, to assure compliance with the NAEP law, Governing Board policies, Department of Education and government-wide regulations, and requirements of the framework contract.

3. The performance of work for the framework development process shall be subject to the technical direction of a Governing Board staff member, designated as the Contracting Officer's Representative. This individual shall work under the guidance of the ADC and the Governing Board during all phases of the framework process.

4. During the framework process, the Governing Board shall review work-in-progress and make modifications as necessary. The Governing Board shall receive regular updates on the framework development process at its quarterly meetings. Updates

shall be provided to the ADC as necessary during the framework development process via in-person meetings, teleconferences, printed material, and electronic communication.

5. At the conclusion of the framework development process, the Governing Board will take final action on the recommended framework, specifications, and background variables documents. This action may result in modifications to one or more of the documents, which will be incorporated prior to dissemination.

6. The framework process shall also result in recommendations to the Governing Board on background variables to be collected from students, teachers, and schools related to a particular subject area. Such variables shall be related to academic achievement and to the fair and accurate presentation of achievement information. Background variables shall meet criteria for being secular, neutral, and non-ideological, as stated in the Governing Board's Policy on NAEP Item Development and Review, and will not assess personal or family beliefs and attitudes, or publicly disclose personally identifiable information. In recommending background variables, the Governing Board's Policy on Collecting and Reporting Background Data shall also be followed. Recommendations on background variables shall take into account burden, cost, quality of the data to be obtained, and other factors.

7. Following adoption by the Governing Board, the final framework, specifications, and background variables documents shall be provided to NCES at least 12 months prior to pilot or field testing, except in the case of unforeseen circumstances related to congressional action, budget limitations, or other extraordinary events.

Principle 5

Through the framework development process, preliminary achievement level descriptions shall be created for each grade being tested. These preliminary descriptions shall be an important consideration in the item development process and will be used to begin the achievement level setting process.

Policies and Procedures

1. The framework panels shall draft preliminary descriptions for basic, proficient, and advanced performance for all applicable grades in the content area under development. The panels shall use the Governing Board's policy definitions for basic, proficient, and advanced achievement in developing the preliminary descriptions. The descriptions shall provide statements of what students should know and be able to do, as derived from the content and process dimensions of the assessment at each grade.

2. The preliminary descriptions shall be included in the framework draft that is widely circulated for public review and comment, to obtain broad input on the draft descriptions prior to Governing Board action on the framework.

3. Once the Governing Board has approved the framework document, NCES shall be provided with the preliminary achievement levels descriptions so that these definitions can guide development of NAEP test questions.

4. The preliminary descriptions approved by the Governing Board shall also be provided to the achievement levels contractor to begin the level-setting process.

Principle 6

The specifications document shall be developed during the framework process for use by NCES and the test development contractor as the blueprint for constructing the NAEP assessment and items in a given subject area.

Policies and Procedures

1. The assessment and item specifications shall produce an assessment that is valid and reliable, and based on relevant widely accepted professional standards. The specifications shall also be consistent with Governing Board policies regarding NAEP design such as booklet and block (item sets within a booklet) structure, test administration conditions, and accommodations for special needs students.

2. The primary audience for the specifications, or assessment blueprint, shall be the contractor(s) responsible for developing the assessment and test questions. The specifications shall be written in sufficient detail so that item writers can develop high-quality questions based on the framework objectives for grades 4, 8, and 12, where applicable, in a given subject area.

3. The specifications shall include, but not be limited to: a) detailed descriptions of the content and process dimensions, including the weighting of those dimensions in the pool of questions at each grade; b) types of items; c) guidelines for stimulus material; d) types of response formats; e) scoring procedures; f) preliminary achievement level descriptions; g) administration conditions; h) description of ancillary or additional materials, if any; i) considerations for special populations; j) detailed information on special studies, if any; k) a substantial number and range of sample items with scoring guidelines for each grade level; and l) any unique requirements for the given subject area.

4. The specifications shall evolve from the framework document, and be carefully reviewed by technical experts involved in the process, prior to submission to the Governing Board.

Principle 7

NAEP assessment frameworks and test specifications generally shall remain stable for at least 10 years.

Policies and Procedures

1. Development of a new subject area framework shall be guided by the schedule of NAEP assessments adopted by the Governing Board.

2. In deciding when to conduct a new framework development process for an existing NAEP assessment, the Board shall consider factors such as exemplary research, curriculum and assessment reform, widely accepted professional standards, implications for existing trendlines, cost and technical issues, and other factors.

3. In rare circumstances, such as where significant changes in curricula have occurred, the Governing Board may make changes to assessment frameworks and specifications before 10 years have elapsed.

4. In those subjects and grades for which NAEP would provide confirmatory evidence about progress in achievement on state tests, the Governing Board shall revise frameworks only when the rationale for doing so is compelling.

MAY 2012

NAEP:
LOOKING AHEAD

LEADING ASSESSMENT INTO THE FUTURE

Recommendations to the Commissioner
National Center for Education Statistics

NAEP: Looking Ahead
Leading Assessment into the Future

NCES Initiative on the Future of NAEP

The National Assessment of Educational Progress (NAEP) has undergone a series of notable changes in the past decade. The NAEP program has expanded to meet new demands. All 50 states, the District of Columbia, the Department of Defense schools, and (on a trial basis) 21 urban districts are now participating in the mathematics and reading assessments at grades 4 and 8. In addition, thirteen states are participating in trial state 12^{th}-grade assessments in reading and mathematics. NAEP is also reporting in record time to ensure that the findings are highly relevant upon release. Technology has taken on a bigger role in the development and administration of NAEP, including computer-based tasks in the science and writing assessments. These are just a few of the major developments; the program has grown and matured in almost all respects.

There is also growing interest in linking NAEP to international assessments so that NAEP scores can also show how our nation’s students measure up to their peers globally. Additionally, there is increasing interest in broadening assessments in the subject areas to incorporate college and career readiness, as well as what are often called “21^{st}-century skills” (communication, collaboration, and problem-solving).

The National Center for Education Statistics (NCES), which administers NAEP, is dedicated to moving the program forward with its upcoming procurement cycle which will take the program to 2017. Under the leadership of NCES Commissioner Jack Buckley, NCES convened a diverse group of experts in assessment, measurement, and technology for a summit in August 2011. These experts discussed and debated ideas for the future of NAEP. NCES convened a second summit of state and local stakeholders in January 2012. Participants at both gatherings were encouraged to “think big” about the role that NAEP should play in the decades ahead.

NCES assembled a panel of experts from the first summit, chaired by Edward Haertel, an expert in educational assessment, to consider and further develop the ideas from the two discussions and make recommendations on the role of NAEP in the future—10 years ahead and beyond. Based on summit deliberations and their own extensive expertise, the panel developed a high-level vision for the future of the NAEP program, as well as a plan for moving toward that vision.

This paper contains the panel’s recommendations to the NCES Commissioner. NCES will consider these recommendations in their mid- and long-range planning for the program.

Panel Members

Edward Haertel (chair)
Jacks Family Professor of Education
School of Education
Stanford University

Russell Beauregard
Research Scientist & Director of Design
Education Market Platforms Group
Intel Corporation

Jere Confrey
Joseph D. Moore Distinguished University Professor
College of Education
North Carolina State University

Louis Gomez
MacArthur Chair in Digital Media and Learning
Graduate School of Education and Information Sciences
University of California, Los Angeles

Brian Gong
Executive Director
National Center for the Improvement of Educational Assessment

Andrew Ho
Assistant Professor
Graduate School of Education
Harvard University

Paul Horwitz
Senior Scientist and Director
Concord Consortium Modeling Center
Concord Consortium

Brian Junker
Professor
Department of Statistics
Carnegie Mellon University

Roy Pea
David Jacks Professor of Education
School of Education
Stanford University

Robert Rothman
Senior Fellow
Alliance for Excellent Education

Lorrie Shepard
Dean and Distinguished Professor
School of Education
University of Colorado, Boulder

3. NAEP's Assessment Frameworks and Learning Outcomes

3.1 Background and History

Assessment frameworks are conceptual, overview documents that lay out the basic structure and content of a domain of knowledge and thereby serve as a blueprint for assessment development. Typically, assessment frameworks, for NAEP and for other large-scale assessments, are constructed as two-dimensional matrices of content strands and cognitive processes. For example, the current NAEP mathematics framework includes five content areas: number properties and operations; measurement; geometry; algebra; and data analysis, statistics and probability. These are assessed at different levels of cognitive complexity, which include mathematical abilities such as conceptual understanding, procedural knowledge, and problem-solving. In geography, the content areas include: space and Earth places; environment and society; and spatial dynamics and connections. The levels of the cognitive dimension consist of knowing, understanding, and applying.

NAEP Assessment Frameworks are developed under the auspices of the Governing Board through an extensive process involving subject matter experts, who consider how research in the discipline and curricular reforms may have shifted the conceptualization of proficiency in a given knowledge domain. The development process also requires multiple rounds of reviews by educators, policy leaders, members of the public, and scholars. It is expected that assessment frameworks will need to be changed over time. However, the decision to develop new frameworks is approached with great caution because measuring change requires holding the instrument constant. Introducing new frameworks—while providing a more valid basis for the assessment—could threaten one core purpose of NAEP, which is to monitor "progress." In the past, when relatively minor changes have been made in assessment frameworks, as judged by content experts, trend comparisons over time have been continued and bridge validity studies have been conducted to verify that conclusions about gains have not been conflated with changes in the measuring instrument or redefinition of the construct being assessed.

When more profound changes occur in the conceptualization of an achievement domain, then a new framework is essential, and correspondingly the beginning of a new trend line. The adoption by nearly all states of the CCSS in English language arts and literacy and mathematics and the new Science Education Framework developed by the National Research Council (NRC) could be the occasion for a substantial enough change in conceptualization of these domains that new NAEP frameworks and new trend comparisons are warranted. Still, the future of NAEP—as a statistical indicator and as an exemplar of leading-edge assessment technology—requires great care and attention to the implications of new trend comparisons rather than merely acceding to the hoopla surrounding the new standards.

In the history of NAEP, few changes have been made in the assessment frameworks for reading and for mathematics. The old frameworks in these two core subjects, begun in 1971 and 1973 respectively, were replaced in the early 1990s, and then again in 2009 for reading. The old assessments have been continued on a less frequent cycle and are referred to as long-term trend NAEP. The 1990's mathematics framework and 2009 reading framework guide the present-day assessments, referred to as main NAEP. While NCES has been careful to insist that the old and new frameworks measure different things and therefore cannot be compared, the existence of the two trends provides a critically important example to illustrate how changing the measure can change interpretations about educational progress (e.g., see Beaton & Chromy, 2010). The earlier assessments focused much more on basic skills. Reading passages were generally shorter compared to today's NAEP and did not require students to demonstrate so wide a range of reading skills or answer extended-response questions. In mathematics, long-term trend NAEP had a greater proportion of computational questions and items asking for recall of definitions, and no problems where students had to show or explain their work. In a 2003 study, researcher Tom Loveless complained that the new NAEP mathematics assessment exaggerated progress in mathematics during the 1990s because gains on the basic skills test over the same period were much smaller (when compared in standard deviation units of the respective tests). Because the two assessments are administered entirely separately, Loveless then had to rely on comparisons based on the less than satisfactory item-percent-correct metric to try to track progress in subdomains of the test. A more recent study using more sophisticated methods has largely confirmed his general conclusions, but that same study has highlighted the technical challenges of comparing trends for two assessments administered under such different conditions (Beaton & Chromy, 2010).

3.2 New Approaches for Assessment Frameworks

3.2.1 Designing frameworks and assessments to evaluate directly the effects of changing domain definitions

NAEP cannot be a research program and in particular cannot be structured to investigate the effectiveness of various instructional interventions. However, it can and should be attentive to the ways that shifting definitions of subject matter competence can affect claims about progress or lack of progress (cf. Section 3.2.3). In the CCSS context, it will be especially important to pay attention directly to potential differences between consortium-based conclusions and NAEP trends. Taking this on as a role for NAEP continues its important function as a kind of monitoring instrument. For example, when some state assessment results have shown remarkable achievement gains and closing of achievement gaps, achievement trends for the same states on NAEP have helped to identify inflated claims. These disparities might exist because of teaching-the-test practices on state tests (Klein, Hamilton, McCaffrey, & Stecher, 2000; Koretz & Barron, 1998), state content or achievement standards that do not rise to NAEP levels (Bandeira de Mello, Blankenship, & McLaughlin, 2009), exclusion of low-performing students on NAEP, or lower motivation on NAEP. More direct linking by carefully accounting for the consortium frameworks within new NAEP frameworks, would allow NAEP to act somewhat like an external monitor for CCSS assessment results. While the current NAEP frameworks do cover many of the same skills as the CCSS, they can be enhanced with some shifts in content.

"21st-century skills" aren't actually new in this century, but it is a relatively new idea (beginning in the 1990s) that these reasoning skills should be more broadly attained and expected of all students. More importantly, it is indeed new that policy leaders would move toward a view of learning that calls for reasoning and explaining one's thinking from the earliest grades, in contrast to outmoded theories of learning predominant in the 20th century that postponed thinking until after the "basics" had been mastered by rote. In addition, the CCSS firmly ground reasoning, problem-solving, and modeling in relation to specific content, not as nebulous generalized abilities. While there is widespread enthusiasm for designing new assessments that capture these more rigorous learning goals, we should note that promises like this have been made before. In the case of the current NAEP mathematics assessment, item developers acknowledge that the proportion of high complexity items actually surviving to the operational assessment is much smaller than is called for in the NAEP Mathematics Framework, and a validity study at both grades 4 and 8 found that the representation of high-complexity problems was seriously inadequate at grade 8, especially in the Algebra and Measurement strands (Daro, Stancavage, Ortega, DeStefano, & Linn, 2007).

Good intentions to measure "higher order thinking skills" are often undermined for three interrelated reasons. First, test questions at higher levels of cognitive complexity are inherently more difficult to develop. Because the dimensions of the task are intended to be ill-specified, such problems are often perceived to be ambiguous. But as soon as the item developer provides clarifying parameters, the challenge of the problem is diminished. Second, because "21st-century skills" involve applying one's knowledge in real world contexts, prior experience with particular contexts (or lack thereof) can create very large differences in performance simply because students unfamiliar with the context are unable to demonstrate the intended content and reasoning skills. In fact, application or generalization can only be defined in relation to what is known to have been taught. This is the curriculum problem that haunts large-scale assessments like NAEP that seek to be curriculum independent. Finally, well designed items can fail on statistical criteria if too few students can do them.

These are all cautionary tales. They do not imply that NAEP should be less ambitious in developing new assessment frameworks that reach as far as possible in representing these higher levels of subject matter proficiency. But they do suggest a hedging-one's-bets approach that does not discard old frameworks wholesale in favor of the new. Rather, as mentioned previously, some conscious combination of old and new would create an assessment better equipped to track progress over time. Later we discuss Innovations Laboratory studies like those NAEP has used historically to explore the feasibility of new assessment strategies. However, we should emphasize that studies of innovative assessment strategies that tap complex skills should not merely be new assessment formats administered to random samples of students. Rather, in recognition of the fact that opportunities to learn particular content and skills may affect whether an assessment looks psychometrically sound, studies should be undertaken with carefully selected populations where relevant opportunities to learn can be established. This will help determine whether more advanced performance can be accurately documented to exist within the parameters of the new standards.

3.2.2 Standing subject-matter panels

To aid in this process, provide substantive oversight, and ensure meaningful interpretation of trends, we elaborate a recommendation for the future of NAEP previously made by a National Academy of Education Panel, which called for standing subject-matter committees. We recommend an expanded role whereby standing committees of subject matter specialists would review field test data, for example, and call attention to instances when after-the-fact distortions of the intended domain occur because more ambitious item types fail to meet statistical criteria. These committees would also have a role in ongoing incremental updates to content frameworks. They might include at least one member with psychometric expertise to aid in formulating technical specifications. The role of these committees is further described in Section 6.1.3.

3.2.3 Dynamic assessment frameworks and reporting scales

As just explained in Section 3.1, NAEP assessment frameworks have historically been held fixed for a period of years and then changed. It might be added that historically, NAEP item pools have been constructed according to test specifications derived from assessment frameworks. NAEP reporting scales, in turn, have reflected the resulting mix of NAEP items. Periodic small revisions to assessment frameworks have been made while maintaining trend lines; major breaks requiring new trend lines have occurred only rarely. With standing subject-matter panels, assessment frameworks for each subject-grade combination might be adjusted more frequently, defining a gradually changing mix of knowledge and skills, analogous to the Consumer Price Index (cf. Section 5.3). At the same time, item pools might be expanded somewhat, including everything in the assessment framework but also covering some additional material. Assessment frameworks would still define the intended construct underlying NAEP reporting scales, but not all items in the NAEP exercise pool would be included in the NAEP reporting scales. For example, content required to maintain long-term trend NAEP, to assure sufficient representation of the CCSS, or to improve the linkage to some other assessment could be introduced into the pool without affecting NAEP reporting scales. With somewhat broader exercise pools, alternative construct definitions could be investigated in special studies. The panel assumes that broader exercise pools, supporting modestly different construct definitions, will increase the value of NAEP by highlighting distinctions among achievement patterns under different construct definitions. Of course, there would still be one main NAEP reporting scale for each subject/grade combination. Clarity in communicating NAEP findings would remain a priority.

Different assessment frameworks may imply different definitions of the same broad subject area achievement construct (e.g., "reading" or "mathematics"), and achievement trends may differ depending on the construct definition chosen. Incremental changes in assessment frameworks and the corresponding set of items on which NAEP reporting scales were based would afford local (i.e., near-term) continuity in the meaning of those scales, but over a period of decades, constructs might change substantially. This was seen by the panel as a potential strength, but also a potential risk. Policymakers and the public should be aware of how and when the construct NAEP defines as "reading," for example, is changed. Not every small, incremental change would need to be announced, but it would be important to establish and to enforce clear policies concerning the reporting of significant changes in assessment frameworks, so as to alert stakeholders when constructs change and to reinforce the crucially important message that not all tests with the same broad content label are measuring the same thing. As small content framework adjustments accumulate over time, standing committees, using empirical studies, would need to determine when the constructs measured have changed enough to require establishing new trend lines.

Dynamic frameworks would balance dual priorities of trend integrity and trend relevance. As an analogy, the Consumer Price Index (CPI) tracks inflation by deliberately conflating two concepts: change in the cost of a fixed basket of goods and change in the composition of the basket itself. As time passes, an increase in the cost of a product that is no longer relevant should contribute less to estimated inflation. By adopting dynamic frameworks, NAEP would similarly conflate increases in student proficiency with a change in the definition of proficiency itself. Although this conflation may seem undesirable, it may be the best way to balance desires for both an interpretable trend and a relevant trend.

3.2.4 Learning progressions as possible guides to assessment frameworks

Learning progressions or trajectories represent descriptions of how students' knowledge, skills, and beliefs about the domain evolve from naïve conceptions through gradual transformations to reach proficiency with target ideas at high levels of expertise over a period of years (Heritage, 2008). They entail the articulation of intermediate proficiency levels that students are likely to pass through, obstacles and misconceptions, and landmarks, of predictable importance as students' knowledge evolves over time. Empirical study of learning progressions highlights the key roles of instruction, use of tools, and peer interactions in supporting learning. Because the process of evolving understanding can take multiple years, learning progressions bridge formative and summative assessment.

A learning progression can provide much more information than a typical assessment framework. A learning progression ideally specifies both what is to be learned as well as how that learning can take place developmentally over time. It often integrates content and cognition. It includes not only the learning targets but also common less-than-ideal states that many students pass through. It is ordered developmentally. It provides a domain-based interpretation of development or growth that is useful to educators. The 2009 NAEP Science Framework already contains a section on learning progressions; however, learning progressions may offer guidance for the development of future NAEP assessment frameworks, especially in mathematics.

Learning progressions are closely entwined with instructional decisions regarding the sequencing of key concepts and skills. In the Netherlands, for example, the related constructions are referred to as "learning-teaching trajectories." However, few empirically supported "learning progressions" as yet exist, and developing more has proven challenging. In addition, because of NAEP's role as a curriculum-independent monitor, it may be more difficult to develop assessment frameworks that are entirely built as a collection of learning progressions. More likely some particular sequences, if proven to be valid across curricula, could be embedded within more general assessment frameworks.

IES NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Assessment Development Committee
Item Review Schedule
February 2017 - June 2017
February 2, 2017

Review Package to Board	Board Comments to NCES	Survey/ Cognitive	Review Task	Approx. Number Items	Status
2/2/17	3/3/17	Cognitive	2021 Reading (4, 8) Pilot (SBT) Concept Sketches	4-8	
2/24/17	3/13/17	Cognitive	2019 Science (12) Pilot (ICT Wave 2) Draft Builds	1 task	
2/24/17	3/13/17	Cognitive	2021 Math (4, 8) Pilot (SBT) Concept Sketches	6-8	
4/3/17	4/26/17	Cognitive	2019 Reading (12) Pilot (DI)	30-35	
4/20/17	5/02/17	Survey	2019 Science (4, 8, 12) Pilot	80-100	
5/3/17	5/26/17	Cognitive	2019 Science (4, 8, 12) Pilot (DI)	157	
5/3/17	5/26/17	Cognitive	2019 Math (12) Pilot (DI)	180	
5/3/17	5/26/17	Survey	2018 Civics, US History, and Geography (8) Operational	130-140	
5/8/17	5/26/17	Cognitive	2019 Reading (12) Pilot (SBT) Final Builds	2 tasks	
5/8/17	5/26/17	Cognitive	2019 Math (12) Pilot (SBT) Final Builds	2 tasks	
6/28/17	7/19/17	Cognitive	2019 Science (4, 8) Pilot (ICT) Final Builds	4 tasks	

NOTE: "SBT" indicates Scenario-Based Task
"DI" indicates Discrete Item