DOCUMENT RESUME

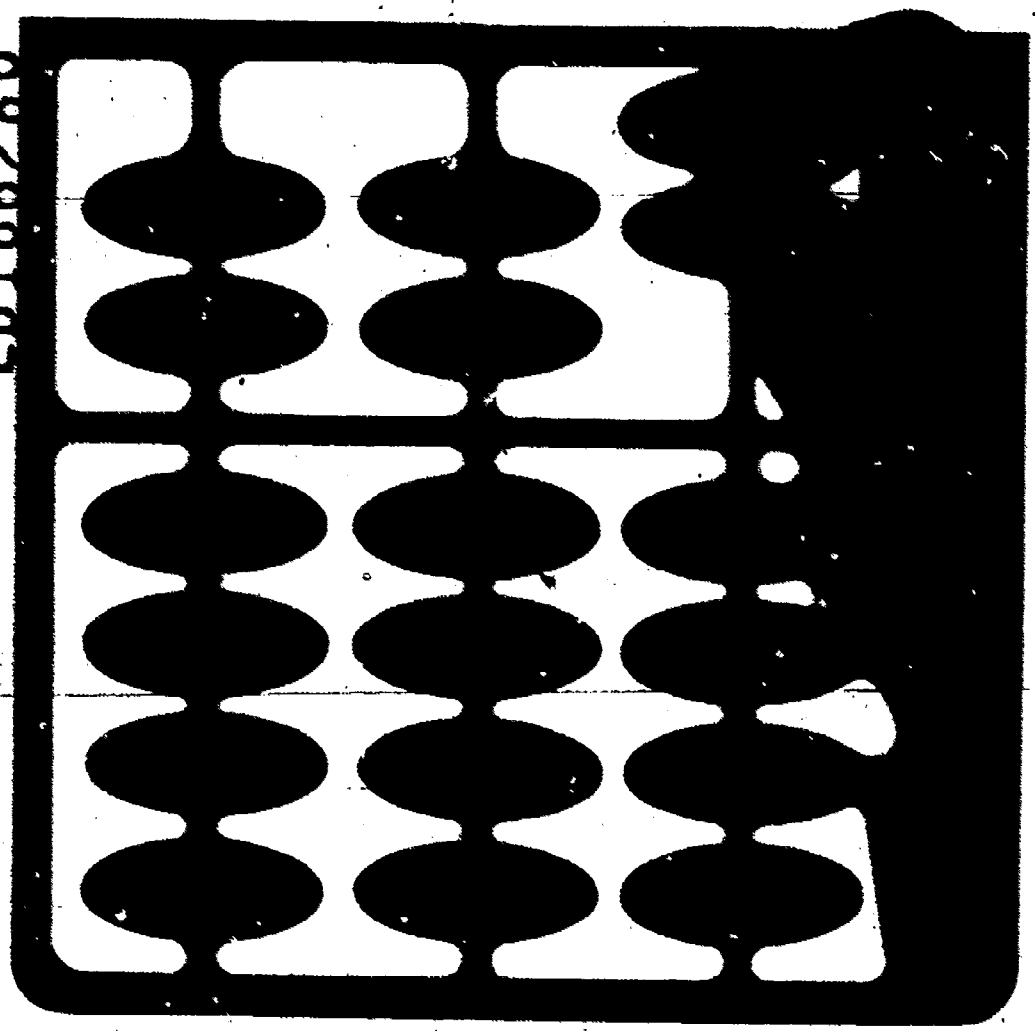
ED 186 280 SE 030 759

AUTHOR Ward, Barbara
TITLE Procedural Handbook: 1977-78 Mathematics Assessment.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY National Center for Education Statistics (DHEW), Washington, D.C.: National Inst. of Education (DHEW), Washington, D.C.
REPORT NO NAEP-09-MA-40
PUB DATE Apr 80
CONTRACT OEC-0-74-0506
GRANT NIE-G-80-0003
NOTE 74p.
AVAILABLE FROM National Assessment of Educational Progress, 1860 Lincoln St., Suite 700, Denver, CO 80295 ($3.70); and Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402 (no price quoted).

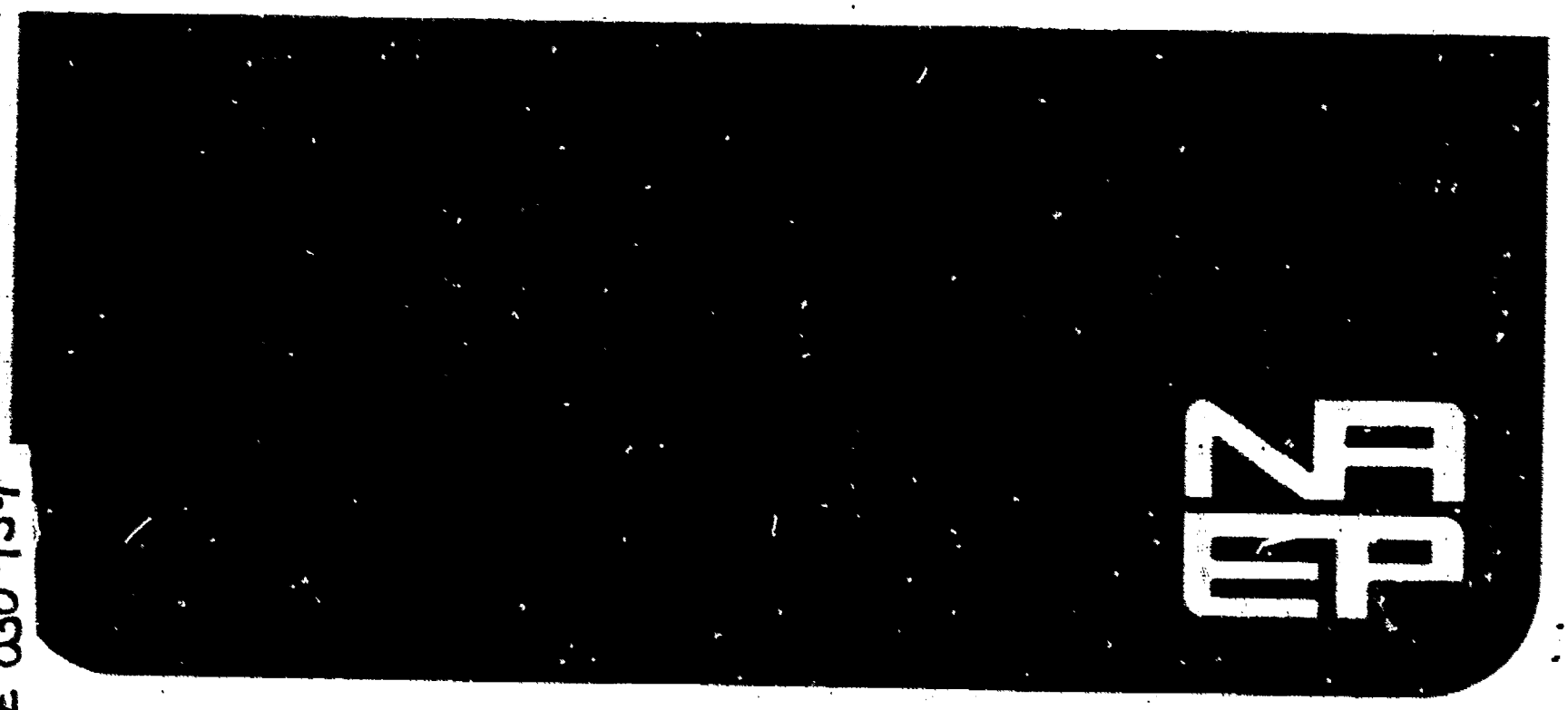
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; Data Analysis; *Data Collection; *Educational Assessment; Elementary School Mathematics; Elementary Secondary Education; Item Sampling; *Mathematics Education; *National Competency Tests; National Surveys; Sampling; Secondary School Mathematics; Statistical Analysis; *Test Construction; Testing Programs

ABSTRACT

The National Assessment of Educational Progress (NAEP) has completed two assessments of mathematics, the first conducted in 1972-73 and the second during 1977-78. Each assessment surveyed the mathematics achievement of American 9-, 13-, and 17-year-olds, using a deeply stratified, multi-stage probability sample design. This report documents procedures used in the 1977-78 mathematics assessment and also describes changes in procedures between the assessments. Described are specific procedures to develop objectives and exercises, draw the assessment sample, prepare materials for the assessment, administer and score the items, and analyze results. Availability of materials providing information about the results of the assessment is summarized in the final chapter. (Author/MK)



PROCEDURAL HANDBOOK:
1977-78 MATHEMATICS ASSESSMENT



PROCEDURAL HANDBOOK:
1977-78 MATHEMATICS ASSESSMENT

Report No. 09-MA-40

by the
National Assessment of Educational Progress
Education Commission of the States
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295

April 1980

The National Assessment of Educational Progress is funded by the National Institute of Education. It is under contract with the Education Commission of the States. It is the policy of the Education Commission of the States to take affirmative action to prevent discrimination in its policies, programs and employment practices.

Library of Congress
Catalog Card Number: 72-183665

Although a few early National Assessment reports have individual catalog card numbers, all recent reports have been assigned this series number.

ISBN 0-89398-143-5

The work upon which this publication is based was performed pursuant to Contract No. OEC-0-74-0506 of the National Center for Education Statistics and Grant No. NIE-G-80-0003 of the National Institute of Education. It does not, however, necessarily reflect the views of those agencies.

TABLE OF CONTENTS

LIST OF EXHIBITS, TABLES AND FORMS

FOREWORD

When the U.S. Office of Education was chartered in 1867, one charge to its commissioners was to determine the nation's progress in education. The National Assessment of Educational Progress (NAEP) was initiated a century later to address, in a systematic way, that charge.

Each year since 1969, National Assessment has gathered information about levels of education achievement across the country and reported its findings to the nation. NAEP surveys the education attainments of 9-year-olds, 13-year-olds and 17-year-olds in 10 learning areas: art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies and writing. Different learning areas are assessed every year, and all areas are periodically reassessed in order to measure possible changes in education achievement. National Assessment has interviewed and tested more than 810,000 young Americans since 1969.

Learning-area assessments evolve from a consensus process. Each assessment is the product of several years of work by a great many educators, scholars and lay persons from all over the nation. Initially, these people design objectives for each subject area, proposing general goals they feel Americans should be achieving in the course of their education. After careful reviews, these objectives are given to exercise (item) writers, whose task it is to create measurement instruments appropriate to the objectives.

When the exercises have passed extensive reviews by subject-matter specialists, measurement experts and lay persons, they are administered to probability samples. The people who compose these samples are chosen in such a way that the results of their assessment can be generalized to an entire national population. That is, on the basis of the performance of about 2,500 9-year-olds on a given exercise, we can make generalizations about the probable performance of all 9-year-olds in the nation.

After assessment data have been collected, scored and analyzed, National Assessment publishes reports to disseminate the results as widely as possible. Not all exercises are released for publication. Because NAEP will readminister some of the same exercises in the future to determine whether the performance level of Americans has increased, remained stable or decreased, it is essential that they not be released in order to preserve the integrity of the study.

ACKNOWLEDGMENTS

Many organizations and individuals have made substantial contributions to the mathematics assessments. Not the least of those to be gratefully acknowledged are the administrators, teachers and students who co-operated so generously during the collection of the data.

Special acknowledgment must go to Jane Armstrong of the National Assessment staff for coordination of development efforts and to Stuart Kahl and George Moreno for their assistance. Also to be acknowledged are the substantial contributions in guiding item development and report planning made by the members of the Mathematics Advisory Board: Dr. Sarah Herriot, Dr. Prince Jackson, Dr. Jeremy Kilpatrick, Dr. Donald Kreider, Dr. John LeBlanc and Dr. James Wilson.

Administration of the mathematics assessment was conducted by the Research Triangle Institute, Raleigh, North Carolina. Scoring and processing were carried out by the Measurement Research Center, Iowa City, Iowa, and by the National Assessment staff.

The actual preparation of this report was a collaborative effort of the National Assessment staff. Special thanks must go to the following people: Donald T. Searls and Eugene Johnson for information on sampling and data analysis; Marci Reser and Carmen Nietes for production. Scoring and statistical analyses for the mathematics assessment were supervised by Donald Phillips. This report was written by Barbara Ward.

Roy H. Forbes
Director

INTRODUCTION

The National Assessment of Educational Progress (NAEP) has completed two assessments of mathematics, the first conducted in 1972—73 and the second during 1977—78. Each assessment surveyed the mathematics achievement of American 9-, 13- and 17-year-olds, using a deeply stratified, multistage probability sample design. This report documents procedures used in the 1977—78 mathematics assessment and also describes changes in procedures between the assessments.

To measure changes in performance between 1972—73 and 1977—78, approximately half of the exercises assessed in the first assessment were reassessed in the second under almost identical administrative conditions. To measure the status of mathematics achievement in 1977—78, National Assessment consultants revised the objectives used in the first assessment and developed additional exercises to provide complete coverage of the revised objectives.

Approximately 17,000 9-year-olds, 27,000 13-year-olds and 27,000 17-year-olds participated in the 1977—78 mathematics assessment. National Assessment reports results for groups of students, not individuals. Therefore, it is not necessary for each student to respond to every item (exercise).[1] Each respondent completed only one item booklet of about 45 minutes in length. Between 2,100 and 2,500 students responded to each booklet. In 1977—78, there were 7 exercise booklets for 9-year-olds, 11 booklets for 13-year-olds and 12 booklets for 17-year-olds.

In each assessment, 13-year-olds were assessed in October through December; 9-year-olds in January and February; and 17-year-olds in March and April. Thus, the amount of school experience for each age group was approximately the same in each assessment.

The exercises for each assessment were administered by a professional data collection staff to minimize the burden on participating schools and to maximize uniformity of assessment conditions. Instructions and items were recorded on a paced audio tape and played back to students to reduce the potential effect of reading difficulties and to insure that all students moved through the packages at the same speed.

Each item included one or more item parts; these item parts were analyzed separately. Over 900 item parts were included in the 1977—78 mathematics assessment. About half the items were multiple-choice and about half were open-ended, or free response, items. Changes in achievement between 1972—73 and 1977—78 were based upon 55 item parts for 9-year-olds, 77 item parts for 13-year-olds and 102 item parts for 17-year-olds. Approximately 20% of the items used to measure change were multiple-choice; the remaining 80% were open-ended.

Multiple-choice items were scored by an optical scanning machine; open-ended items were hand-scored by trained scorers using scoring guides, which defined categories of acceptable and unacceptable responses. These scoring guides were developed following field testing of the items and then revised and

[1] National Assessment uses the term "exercise" to mean an assessment item. The terms "exercise" and "item" are used interchangeably in this report.

refined during receipt of initial assessment data.

National Assessment reports estimated percentages of correct responses for single items. When a report indicates that "85% of the 17-year-olds gave a correct response," it means that an estimated 85% of the 17-year-olds would have given a correct response if all the 17-year-olds in schools across the country had been assessed. In addition to reporting national results, National Assessment provides data on the performance of various population subgroups within the national population, defined by sex, race, region of the country, size and type of community lived in, and level of parental education. For the 1977–78 mathematics assessment, some data are available on the performance of 17-year-old students with varying levels of mathematics course-taking experience.[2] National Assessment aggregates percentages of success on various sets of items to provide data on changes in performance between assessments and on the differential performance of population subgroups.

The following chapters describe specific procedures used to develop objectives and exercises, draw the assessment sample, prepare materials for the assessment, administer and score the items and analyze results. This information should prove useful to those wishing to replicate National Assessment procedures and to researchers and data users with an interest in how the data were obtained. Materials providing information about the results of the assessment are described in Chapter 8.

[2] See Appendix B for definitions of the subgroups reported upon.

CHAPTER 1

OBJECTIVES REDEVELOPMENT

The primary goal of the National Assessment of Educational Progress (NAEP) is to report on the current education status of young Americans and to monitor any changes in achievement over time. For each learning area to be assessed, NAEP asks consultants to develop objectives that define the subject area; specify knowledge, skills and attitudes to be assessed; and serve as guides for exercise developers.

Education in America is a collaborative enterprise involving a great many people with widely differing philosophies. Providing information about education nationwide would be considerably easier if there were consensus about the means and ends of American education, but the fact is that Americans have conflicting and sometimes contradictory values regarding the goals of education and the means for achieving them. To develop an assessment that is truly national in scope and takes into account the diversity of curricula, values and goals across the country, National Assessment employs a consensus process for developing objectives, involving representation of many different groups of people.

Several types of consultants helped develop the 1977—78 mathematics objectives. College and university specialists in mathematics insured that the objectives included important concepts that the schools should be teaching. Educators, including classroom teachers, curriculum supervisors and persons involved in teacher education, made sure that the objectives included concepts, skills and attitudes that the schools should be teaching and those that they presently are teaching. Concerned citizens, parents and other interested lay persons had to agree that the objectives were important for young people to achieve, were free of education jargon and were not biased or offensive to any groups. Consultants were representative of the different regions of the country, minority groups, both sexes, and various types of communities, age levels and education philosophies.

The objectives for the 1972—73 mathematics assessment were developed by an outside contractor, utilizing the types of consultants described above. Objectives for the second assessment were developed through conferences organized and conducted by the National Assessment staff. This procedure increased interaction between assessment staff and consultants, involved more professionals in the areas of mathematics and education, allowed greater flexibility and was more efficient in terms of cost and time.[1]

Redevelopment of the mathematics objectives for the second assessment was built upon the framework established by the objectives for the first assessment. The 1972—73 mathematics objectives were originally organized into a three-dimensional scheme: (1) uses of mathematics, (2) mathematical content and (3) cognitive ability levels. The first dimension — defined as social, technical or academic uses — proved difficult to apply in guiding exercise development and so was abandoned as a functioning dimension of the objectives. The remaining two dimensions included 17 content categories and 6 cog-

[1] For a more complete description of the procedures used in redeveloping the objectives, see *Mathematics Objectives, Second Assessment* (1978).

nitive ability categories.

The 1977–78 mathematics objectives were also organized in a content-by-process (or cognitive ability) matrix. This matrix resembled the classification scheme developed for the first assessment, but fewer and more inclusive content and process categories were included in the matrix for the second assessment. The second assessment objectives included five content categories and four cognitive-process levels[2] (Exhibit 1).

In developing the mathematics assessment, the consultants felt that several topics not included in the matrix were still highly important to assess. Some of these topics looked toward future emphases in mathematics education: use of the hand-held calculator, computer literacy and computer-related mathematics. Exercises addressing these topics will provide baseline data against which to measure the future impact of these topics in the schools. The measurement of attitudes toward mathematics also received considerable attention. Attitudes were not measured during the first assessment because of the difficulty of developing adequate measures. However, consultants for the second assessment strongly encouraged the development of attitudinal exercises, so a concentrated effort was made to include such measures.

After consultants developed the initial objectives, National Assessment formed an advisory board to give direction and advice to NAEP staff for further refinement of the objectives and development of the assessment. The six-member advisory board included three mathematics educators, two mathematicians and a mathematics teacher (see Appendix A for members). This board was instrumental in organizing the final set of objectives, planning the development of exer-

[2] For a description of the content and process levels, see *Mathematics Objectives, Second Assessment* (1978).

EXHIBIT 1. Content by Process Matrix for the 1977–78 Assessment of Mathematics

PROCESS	CONTENT: Numbers and Numeration	Variables and Relationships	Geometry (Shape, Size and Position)	Measurement	Other Topics
Mathematical knowledge					
Mathematical skill					
Mathematical understanding					
Mathematical application					

cises, selecting the final exercises and planning the subsequent report topics.

The objectives were also reviewed by a panel of lay persons to insure that they covered topics generally considered to be important and that they were not biased. In addition, they were looked upon as guides to structure reporting. National Assessment staff and the advisory board felt that reporting of results would be better organized and more comprehensive if a reporting scheme existed prior to exercise development. Accordingly, the advisory board recommended structuring the reporting by cognitive-process levels, and they devised a set of questions that relate to each of the four cognitive processes (Exhibit 2).

EXHIBIT 2. Questions Used To Structure Reporting of the 1977–78 Mathematics Assessment

I. Mathematical knowledge

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

II. Mathematical skill

- A. How well can students perform computations, including computations with whole numbers, integers, fractions, decimals, percents, ratios and proportions?
- B. How well can students make measurements?
- C. How well can students read graphs and tables?
- D. How well can students perform geometric manipulations like constructions and spatial visualizations?
- E. How well can students perform algebraic manipulations?
- F. How well can students estimate the answers to computations and measurements?

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?

IV. Mathematical application

- A. How well can students solve typical textbook problems?
- B. How well can students solve nonroutine problems?
- C. How well can students estimate the answers to problems?
- D. How well can students use mathematics in reasoning and making judgments?

While National Assessment cannot supply definitive answers to such questions (determining how "well" students perform is a subjective judgment), it can provide data to assist individual readers in making their own judgments about these questions.

CHAPTER 2

DEVELOPMENT OF EXERCISES

Exercises were developed to provide information about achievement levels for the various cells of the objective matrix and to provide data that could be used in answering the questions shown in Exhibit 2. Each exercise was designed so that its results could either be used alone, as an indicator of performance on a specific task, or used in conjunction with results from other exercises to give a more general picture of achievement levels.

Exercise Development

In developing the objectives, groups of consultants and the advisory board weighted the relative importance of the various cells in the objectives matrix. These weights were used to determine relative amounts of time to be devoted to assessing each cell and were utilized in the final selection of exercises.

The 1972—73 exercises that were reassessed in 1977—78 to measure change were reclassified in the cells of the 1977—78 objectives matrix. This allowed the same reporting scheme — cognitive-process levels — to be used for both change and 1977—78 status reporting. In addition, this meant that 1977—78 results for the items used to measure change could also be used as a part of the 1977—78 reporting.

Most item writing was done by groups of people active in the fields of mathematics and mathematics education. Approximately 150 people served as item developers. Rather than addressing specific cells of the objectives matrix, the groups worked primarily with the process dimension of the matrix, with each group developing a set of items that addressed one or two of the questions from the list of cognitive-process questions shown in Exhibit 2. Individuals generated items, which were then reviewed and revised by the members of the group. Reviewers considered age-level appropriateness, accuracy of content, how well the item measured a question or objective, and readability. Exercises passing the group review were edited by the National Assessment staff to fit NAEP format and technical requirements.

Field Tryouts

The exercises produced by the writing groups for the 1977—78 mathematics assessment were field tested in schools across the country to discover potential problems in wording, directions or administration procedures and to collect item statistics, timing information and scoring information. "Tryout" schools were selected to represent high- and low-income communities as well as more typical communities. The tryouts were administered to students in at least four classrooms (approximately 100 students) at each of the ages assessed. So that the tryouts would closely simulate actual assessment field procedures, the students recorded their answers in the test booklets; directions and questions were read to students from an audio tape; and National Assessment staff members, rather than classroom teachers, administered the test. The completed tryout packages were then scored and submitted to item analysis. The item-analysis statistics, as well as the administrators' reports of any field problems,

helped both staff and consultants to evaluate and revise the exercises. Revised exercises were generally field tested again.

After the initial exercise pool was developed by the writing groups working with the process dimension of the matrix, three writing teams reviewed the exercises and accompanying tryout data (organized by content categories) to insure that content areas were adequately covered. These teams edited existing exercises and generated new ones, which again, were field tested.

Exercise Reviews and Final Selection

Each item selected to be considered for inclusion in the final 1977—78 assessment pool of items was reviewed in a series of conferences by at least 25 different consultants: mathematicians, mathematics educators, classroom teachers and interested lay citizens. Exercises for each age group were reviewed for appropriateness by teachers who taught students at that age. Lay citizens, representing a variety of occupations and interests, also reviewed the exercises, checking for sex or racial bias and considering the general importance of each exercise.

The final exercise selection for the 1977—78 assessment was approved by the advisory board, which used assigned cell weights to decide how many exercises to include in the assessment of each objective. The number of exercises that could be assessed was limited by the amount of testing time available at each age: 245 minutes at age 9 (7 booklets), 385 minutes at age 13 (11 booklets) and 420 minutes at age 17 (12 booklets). Since there was not enough testing time to include all of the exercises selected, National Assessment staff, with the concurrance of the advisory board, deleted a small number of exercises.

An item could have one or more parts. For the mathematics assessments, each item part was analyzed as a separate unit of data. Table 1 shows the approximate number of item parts included in each content area and in each cognitive-process level by age.

In developing the mathematics assessment, National Assessment included items that

TABLE 1. Number of Item Parts Assessed for Each Cognitive-Process Level and Content Category, by Age

	Age 9	Age 13	Age 17
Cognitive-Process Level			
Knowledge	161	147	140
Skill	137	272	273
Understanding	44	108	105
Applications	44	106	136
Content Category			
Numbers and numeration	226	327	300
Variables and relationships	24	60	108
Geometry	44	100	106
Measurement	49	64	61
Other topics	43	82	79
Attitudes	118	207	208
Required use of calculator	37	41	48
Working with computers	-*	4	4

**No item parts assessed at this age.*

depart somewhat from the traditional mathematics "paper and pencil" test item. A number of items at each age required the use of apparatus such as a compass, an English ruler or a metric ruler, which were supplied to the students. One booklet at each age required the use of a simple four-function electronic hand calculator, which was also supplied. Some of the exercises included in the calculator booklets also appeared in booklets for which the calculator was not used, thus enabling comparison of performance with and without a calculator.

The second assessment of mathematics contained fewer open-ended items than the first assessment. While open-ended exercises can provide valuable diagnostic information on errors commonly made by the students, they are considerably more expensive and more complicated to score than multiple-choice items. In many cases, carefully designed multiple-choice questions can elicit sufficient information. All open-ended exercises from the first assessment used to measure change were readministered in their original format.

CHAPTER 3

PREPARATION OF ASSESSMENT MATERIALS

Overview of Packaging

Following the selection of mathematics exercises to be included in the assessment, National Assessment staff grouped and sequenced items into exercise booklets. Since students at different ages received somewhat different sets of exercises, booklets were constructed separately for each age level. Thus, exercises for 9-year-olds were not sequenced in the same order as those for 13-year-olds, and so forth.

The following constraints were observed in preparing exercise booklets:

1. Each booklet contained exercises of varying difficulty so that students would not become bored by many easy exercises or discouraged by many difficult exercises.

2. Each booklet contained exercises measuring a variety of cognitive-process levels and content categories.

3. Exercises could not cue other exercises. In other words, the answer to one exercise could not be contained in another exercise in the same booklet.

4. Each booklet was timed so that it would take no more than 45 minutes — the length of a typical class period — of a student's time. Booklets contained approximately 30—35 minutes of exercise time and an additional 10—15 minutes of introductory material, instructions and background questions.

5. Booklets were designed to be, insofar as possible, parallel with respect to the number of different objectives measured and difficulty levels. Items measuring a particular objective were scattered throughout the booklets so that many different students would respond to questions related to a particular objective.

National Assessment makes every effort to minimize difficulties connected with the testing situation so that results will be, as nearly as possible, an accurate reflection of what students know and can do. For example, students marked their answers directly in the assessment booklets, not on separate answer sheets. It was felt that this procedure would reduce possibilities for errors in marking answer sheets, especially for the younger students. To minimize guessing, students were encouraged to write "I don't know" on the answer line for open-ended questions or to select the "I don't know" response option included with each multiple-choice item if they felt they did not know the answer to a question.

Paced audio tapes were prepared for each exercise booklet. Instructions, most of the written portions of an exercise stimulus and response options were read aloud to minimize the effect of any reading difficulties and to insure that all students moved through the booklets at the same speed. In addition, the use of tapes helped to insure uniform assessment conditions across the country. During the field testing of the exercises, administrators had determined the time needed for

most of the students to respond to an item.

Differences in Packaging: 1972–73 and 1977–78

National Assessment makes every effort to make assessment conditions for items measuring change identical from assessment to assessment so that any changes observed will be attributable to changes in achievement rather than a response to an altered testing condition. Exercises that were packaged together in a booklet in the first assessment were packaged together in the second, and they were sequenced in the same relative order in both assessments. However, in 1972–73, mathematics and science were assessed together, while in 1977–78 each booklet assessing mathematics included only mathematics items. In addition, about half the 1972–73 items were released for publication following the assessment and not reassessed. Thus, although exercises were in the same relative sequence for each assessment, exercises that separated some of the exercises in the first assessment were no longer present in the second.

All exercises used to measure change were located together either at the beginning or the end of a package. Changes in the scripts for the paced audio tapes occurred between 1972–73 and 1977–78;[1] tapes for exercises measuring change followed the 1972–73 scripts. It was felt that switching between methods of reading exercises might prove confusing for students if items measuring change and new 1977–78 items were intermixed.

Item booklets contained more items and/or item parts in the 1977–78 assessment than in the 1972–73 assessment. More multipart items and more multiple-choice items, which generally required less time to administer, were included in the second assessment than the first.

---

[1] In 1972–73, the "I don't know" response choice was read aloud for each exercise; for 1977–78 exercises that did not measure change, it was not.

CHAPTER 4

SAMPLING

This chapter describes procedures used in the sample design and sample selection for the 1972–73 and 1977–78 assessments of mathematics exercises. The sample design and sample selection for both assessments of mathematics were done by the staff of the Research Triangle Institute, Raleigh, North Carolina.

The target populations for each of the assessments consisted of 9-, 13- and 17-year-olds[1] enrolled in either public or private schools at the time of the assessment who were not functionally handicapped to the extent that they could not participate in an assessment. Specific groups excluded were: non-English-speaking persons, those identified as nonreaders, persons physically or mentally unable to respond, and persons in institutions or attending schools established for the physically handicapped or mentally retarded.

National Assessment does not follow up specific individuals from one assessment to the next. In other words, the students who participated in the 1972–73 assessment are not the same ones who participated in 1977–78. However, in each assessment year, participants were carefully selected to represent each age level. For example, National Assessment assessed one set of probability samples of 9-year-olds in 1973 and a totally different set of probability samples of 9-year-olds in 1978. Each was a sample of the population of students who were 9 years old during that assessment year. Thus, when we say that 9-year-olds' achievement declined between 1973 and 1978, we mean that students who were 9 years old in 1973 correctly answered the same questions more often than those who were 9 years old in 1978.

The definitions of the target populations were identical in each assessment. However, the sample design used to obtain representative samples of the target populations was modified somewhat between assessments. An overview of the general sample design approach is given below, followed by a description of differences between the 1972–73 and 1977–78 procedures.

Overview of the National Assessment Sample Design

National Assessment used a deeply stratified, three-stage national probability sample design with oversampling of low-income and rural areas. In the first stage, the United States was divided into geographical units of counties or groups of contiguous counties meeting a minimum population size requirement. These units, called primary sampling units (PSUs), were stratified by region and size of community. From the list of PSUs, a sample of PSUs was drawn (without replacement) with probability proportional to population size measures, representing all regions and sizes of communities. Oversampling of low-income and extreme-rural areas was first performed at this stage by adjusting the estimated population size measures of those areas to increase sampling rates. Within PSUs,

[1] Definitions of 1977–78 assessment age groups are: 9-year-olds — born during calendar year 1968; 13-year-olds — born during calendar year 1964; and 17-year-olds — born October 1, 1960, through September 30, 1961.

Census Employment Survey Data were used to further delineate and oversample low-income areas. Counties with high proportions of rural families were also oversampled.

In the second stage, all public and private schools within each PSU selected in the first stage were listed. Schools within each PSU were selected (without replacement) with probabilities proportional to the number of age-eligibles in the school.

The third stage of sampling occurred during the data collection period. A list of all age-eligible students within each selected school was made. A simple random selection of eligible students without replacement, was obtained, and item booklets were administered to selected students. Specially trained personnel selected the sample and administered the booklets.

Each respondent in the sample did not have the same probability of selection because some subpopulations were oversampled, and adjustments were made to compensate for some schools' refusal to participate and for student nonresponse. The selection probability for each individual was computed, and its reciprocal was used to weight each response in any statistical calculation to compensate for unequal rates of sampling and to insure proper representation in the population structure.

The number of PSUs, schools within PSUs and students within schools was determined by optimum sampling principles. That is, a sample design was selected that would achieve the maximum precision for a given level of resources.

Table 2 displays the number of PSUs and schools within PSUs selected in 1972—73 and 1977—78 by age.

TABLE 2. Number of PSUs and Schools Within PSUs Selected in 1972—73 and 1977—78

	1972—73 Assessment		1977—78 Assessment	
	No. of PSUs	No. of Schools	No. of PSUs	No. of Schools
Age 9	116	971	78	538
Age 13	116	979	78	571
Age 17	116	798	78	540

Differences in Sample Design: 1972—73 and 1977—78

The 1977—78 sample was drawn according to the following procedures. Two types of PSUs were identified: (1) large-size population areas defined by the U.S. Bureau of the Census as Standard Metropolitan Statistical Areas (SMSAs) and (2) other contiguous non-SMSA counties grouped together to meet certain minimum-size requirements. The first stratification of PSUs was by geographic region, as defined by the Office of Business Economics, U.S. Department of Commerce. (See Appendix B, definition of regional subgroups.)

Within regions, PSUs were classified into five size-of-community (SOC) categories:

SOC 1 — PSUs corresponding to the 13 largest SMSAs after adjusting the population size to compensate for oversampling low-income metropolitan areas. These PSUs have selection probabilities so large that under our allocation procedures they are certain to be included in our sample each year. These PSUs are designated as *self-representing*.

SOC 2 — PSUs corresponding to the remaining 57 SMSAs, with over 500,000 population.

SOC 3 — PSUs corresponding to the remaining 162 SMSAs.

SOCs 4 and 5 — PSUs made up of non-SMSA counties. SOCs 4 and 5 are determined so that half of the remaining population (after adjustment for

oversampling of rural areas) falls into each category. SOC 4 contains PSUs in which less than 60% of the residents are classified as rural.

The self-representing PSUs represent additional stratification, making an effective total of 17 (13 + 4) size-of-community strata. Each self-representing SMSA was divided further into geographical substrata or nonoverlapping replicates that constitute multiples of convenient work units for item administration. These multiple work units were included with the rest of the nonself-representing PSUs to form the pool from which first-stage sampling units were selected. To insure adequate representation, National Assessment doubled the sampling rate of low-income and rural areas.

In 1975—76, first-stage units were selected simultaneously for four consecutive assessment years (1975—76 through 1978—79), as were schools in the self-representing PSUs. The sample design required that every four years we assess at least once in every state and not more than once in any school. There are 1,101 primary sampling units in the primary sampling frame for the four-year period, from which about 75 first-stage sampling units are selected each year.

Within the primary strata, public and private schools were listed and further stratified by the estimated number of youngsters eligible at each age. Small schools were clustered until they were large enough to respond to the same number of packages as the larger schools in a stratum. Schools or school clusters were selected without replacement, with probability proportional to the number of age-eligibles in the school or cluster of schools. Once schools were identified, districts were contacted to check for changes in grade range and for the existence of new schools. This information was used to revise probabilities of schools' selection.

In the third stage, students were selected with equal probability and without replacement within each sample school. The number of students selected is proportional to the number of age-eligibles, with oversampling in low-income and rural areas.

During data collection, allowing for variable group sizes for each package administration within schools enabled National Assessment to obtain desired sample sizes in schools having characteristically low response rates. This feature also permitted last-minute modifications and adjustments to selection probabilities necessitated by enrollment changes.

The sampling procedures used in 1972—73 differed somewhat from those used in the 1977—78 assessment (*Final Report ... Sampling and Weighting Activities, ...* 1979; Moore et al., 1974). In 1972—73, the PSUs were stratified by region, size of community and socioeconomic status (SES). In addition, the sample was constrained to include all states. The sampling of PSUs in 1972—73 was accomplished by using a controlled selection procedure (Moore et al., 1974). In 1977—78, PSUs were stratified by region and size of community, with the constraint that each state must appear in the sample once every four years. Controlled selection of PSUs was no longer required.

The size-of-community stratifications in 1972—73 were different from those in 1977—78, which were described earlier in this chapter. There were only four SOC stratifications in the first assessment of mathematics. The first SOC category in 1972—73 consisted of all central cities with overall population greater than 180,000. The second SOC category consisted of the remainder of the SMSA containing the central city in SOC 1. The SOC 3 category in 1972—73 consisted of the remainder of the SMSAs and all counties not in SOCs 1 and 2, containing at least one city with a minimum population of 25,000. In 1972—73, the SOC 4 category consisted of all the remaining counties not in SOCs 1, 2 or 3.

In 1977—78, oversampling of low-income metropolitan areas and extreme-rural areas

was accomplished at the primary stage by increasing the estimated population size measures of PSUs containing these areas and then sampling with probabilities proportional to those adjusted size measures. In 1972–73, a poverty index was used to stratify PSUs into high- and low-SES stratifications. The sampling rates within these strata were then increased in order to achieve the desired oversampling.

CHAPTER 5

DATA COLLECTION

Professional data collection staff from the Research Triangle Institute, Raleigh, North Carolina, administered the assessment booklets. This staff was used to minimize the burden on participating schools and to ensure, insofar as possible, uniform administration conditions across the country (*Final Report . . . In-School Field Operations...*, 1978).

Participation in the National Assessment is voluntary. NAEP makes every effort to encourage the schools selected in the sample to participate in the assessment, and National Assessment and Research Triangle Institute staff have obtained high rates of school cooperation, as shown in Table 3. Student cooperation rates were also high, especially for 9- and 13-year-olds (*Final Report . . . Field Operations...*, 1978, p. 35, Table 22). The effect of student nonresponse is discussed in Appendix D. Table 4 shows the actual number of students that responded to a particular exercise booklet at each age level in the 1977–78 assessment.

Since National Assessment reports data for *groups* of students, not *individuals*, it is not necessary for each student to respond to every item. In 1977–78, booklets were administered to groups of 10–25 students; each group responded to only one booklet. The groups varied in size depending on an estimate of the rate of nonresponse for a particular school. In 1972–73, the planned session sizes were fixed at 12 students at each age.

National Assessment takes steps to guarantee the anonymity of each respondent. Students' names were listed with their booklet identification number so that scoring and processing personnel could go back to the school lists for data verification — for instance, on background information — if necessary. These lists did not leave the schools and were destroyed six months following the assessment in a school.

School officials were asked to respond to a "principal's questionnaire." This questionnaire asked about the enrollment in various grades, the types of communities in which the students lived and the general occupational levels of the people in the community. Principals of 9-year-olds' schools

TABLE 3. School Cooperation Rates, 1977–78 Assessment

Age	Percent of Eligible Schools Participating in 1977–78 Assessment
9	91.5
13	91.3
17	89.5
Overall	90.8

TABLE 4. Number of Students Responding to Each Item Booklet in 1977–78 Assessment, by Age

Age 9		Age 13		Age 17	
Booklet	Number Responding	Booklet	Number Responding	Booklet	Number Responding
1	2,496	1	2,423	1	2,295
2	2,463	2	2,433	2	2,274
3	2,438	3	2,469	3	2,273
4	2,445	4	2,437	4	2,165
5	2,429	5	2,462	5	2,264
6	2,482	6	2,440	6	2,233
7	2,437	7	2,316	7	2,254
		8	2,431	8	2,221
		9	2,368	9	2,238
		10	2,431	10	2,216
		11	2,451	11	2,199
				12	2,124
Total	17,190		26,661		26,756

were asked to indicate the number of minutes per week devoted to mathematics instruction and whether the time devoted to mathematics computation had increased, decreased or remained the same over the past five years. Examples of forms used to collect background information from students and school officials are located in Appendix C.

The assessment administrator coded each student's birth date, sex, grade, racial/ethnic classification and identification number on his or her booklet. Administrators made a visual racial/ethnic identification at the time each student turned in his or her booklet. During the 1977–78 assessment, six different racial classifications were used: white, black, Spanish heritage, American Indian or Alaskan native, Pacific Islander or Asian, and unclassified. If an administrator was unsure of a student's racial/ethnic group, he or she referred to the student's name or listened to the student talk to make the identification. Students were not verbally asked to give a racial identification for themselves by the assessment administrator; however, 17-year-old students were asked to provide this information in one of the background questions included in the exercise booklet.

Sample sizes of the two classifications American Indian or Alaskan native and Pacific Islander or Asian are too small to permit reporting for these groups. Results for the group classified as Spanish heritage cannot be reported for separate exercises but can be reported for aggregate results across a number of exercises. Change results between 1972–73 and 1977–78 are reported for whites, blacks, and Hispanos or Spanish surnamed.

Each age group was assessed in approximately the same time of the school year in each assessment. As noted previously, 13-year-olds were assessed in October–December, 9-year-olds in January–February and 17-year-olds in March–May.

Following data collection, assessment administrators sent completed booklets to the scoring contractor, Measurement Research Center, Iowa City, Iowa.[1] Booklets were quality-checked to verify that correct administrative procedures were being

[1]Now Westinghouse DataScore Systems, Iowa City, Iowa.

followed by the field staff and that all booklets were accounted for. Coded identification information was also checked for accuracy; inconsistencies that could not be reconciled were sent back to the assessment administrator to be checked against the list of student names and identification numbers retained by the school for six months following the assessment.

In 1972—73, 17-year-olds who were not currently attending school and young adults aged 26—35 were included in the mathematics assessment. Out-of-school 17-year-olds and adults could each answer up to four booklets of assessment materials; they were paid $5.00 for each booklet that they completed. Unpaced audio tapes were used for these groups. Assessment of out-of-school 17-year-olds and adults is quite expensive, and in 1977—78, funds to assess these groups were not available. Since out-of-school 17-year-olds were not assessed in 1977—78, only results for 17-year-olds attending school in 1972—73 and 17-year-olds attending school in 1977—78 were used in calculating changes in 17-year-olds' performance.

CHAPTER 6

SCORING

The data that National Assessment collects in the field must be converted into a form suitable for computer processing and analysis. This conversion was done by the Measurement Research Center, Iowa City, Iowa,[1] the contractor in charge of printing the assessment item booklets, receiving, scoring and machine-processing the data.

In the 1977—78 assessment, approximately half the items were multiple-choice and half were open-ended. Responses to multiple-choice exercises were read directly by optical scanning machines. The scoring contractor employed a special staff to hand score open-ended exercises. Scorers were responsible for categorizing responses, using the scoring guides for open-ended items that defined categories of acceptable and unacceptable responses, and were responsible for coding the category of responses for an item into ovals that could be read by the optical scanning machine.

Scoring guides for open-ended exercises were developed using field tryout data. Scoring categories included likely errors and thus were useful in identifying frequently made mistakes.

For changes in performance to be measured accurately, scoring had to be the same in each assessment year. For multiple-choice items, the same responses were scored correct in each year. Scoring guides for open-ended 1972—73 items used to measure change were revised in 1977—78 to be consistent with the guides for items first administered in 1977—78. All responses to open-ended items made in 1972—73 were rescored at the same time that the 1977—78 responses were scored, using the revised scoring guides, to ensure that scoring of the two sets of data was identical.

Scorers were trained in the use of the scoring guides by scoring sample responses taken from arriving assessment data; National Assessment staff was involved in the training process. Scorers initially worked as a group and discussed the appropriate categorization of each example response. They then worked individually on another set of responses. Discrepancies were resolved and explained. Once the group felt comfortable using the guides, they started categorizing the actual data. Supervisory personnel checked all work done for the first few days of a scoring effort to be sure that scoring was consistent.

To further ensure the quality and consistency of scoring open-ended exercises, quality-control checks were conducted during the scoring of these exercises. At regular intervals, randomly selected responses were drawn from the total pool of responses for an item and read by randomly selected scorers. Both the responses and the scorers were selected without replacement; approximately 10% of the responses were included in the quality-control check. Scores for the quality-control readings were recorded, and the responses selected for quality control were then put back into the total pool of responses to be scored during the regular course of scoring. Following scoring of all responses, the two scores for quality-control responses were compared. If discrepancies in scoring became apparent, scorers were

[1] Now Westinghouse DataScore Systems, Iowa City, Iowa.

retrained and, on some occasions, work was rescored.

Percentages of agreement between quality-control and regular scoring were computed for each open-ended exercise. These data are summarized in Table 5. There is a greater range and a slightly lower average percent of agreement as age increases. Since the age 13 responses were scored first, the age 9 responses second and the age 17 responses last, the decline in average percent of agreement and the increase in range were not solely due to "staleness" of the scorers over time. A more likely hypothesis is that the exercises (thus, their associated scoring guides) became increasingly more complex and difficult for the older age levels.

Scoring for each age group began during the administration of the assessment to that age group. Scoring for each age group took from 6 to 12 weeks. Scorers were divided into two teams, and each team learned the scoring guides and scored about half the open-ended items at each age level. Approximately 10 to 20 scorers were working on scoring the mathematics data at any given time.

TABLE 5. Quality Control Summaries for the 1977–78 Mathematics Assessment Open-Ended Exercises, by Age

	Age 9	Age 13	Age 17
Total number of responses sampled	70,947	84,470	106,732
Range of average percent of score agreement for exercises	94–100	92.3–100	86.4–100
Average percent of score agreement across exercises	98.9	98.5	97.7

CHAPTER 7

DATA ANALYSIS

Measures of Achievement

As stated previously, National Assessment reports the performance of groups of students, not individuals. The basic measure of achievement reported by the assessment is the percentage responding acceptably to a given item. This percentage is an estimate of the percentage of 9-, 13- or 17-year-olds who would respond acceptably to a given item if every 9-, 13- or 17-year-old in the country were assessed.

Percentages of correct responses are used because each item is designed as a separate measure of some aspect of an objective or subobjective. The purpose of National Assessment is to discover if more or fewer people are able to answer these items correctly — and thus meet the objectives — over time. Some items — such as attitudinal questions — did not have "correct" answers. For these items, National Assessment reported the percentages responding in various ways.

In addition to providing national results, National Assessment reports on the achievement of various subpopulations of interest. Groups are defined by region of the country, sex, race, size and type of community lived in and level of parents' education. For the 1977—78 mathematics assessment, results were also analyzed by different levels of mathematics course-taking reported by 17-year-olds. Results for some additional variables were also analyzed. Information on these variables is found in *Mathematics Technical Report: Summary Volume* (1976). Definitions of the reporting groups are found in Appendix B.

Procedures for estimating percentages of acceptable responses to exercises are dependent on the sample design. Each response by an individual is weighted and multiplied by an adjustment factor for nonresponse.[1] An estimate of the percentage of a particular age group that would have responded to an exercise acceptably if the entire age group were assessed is defined as the weighted number of acceptable responses divided by the weighted number of all the responses. A similar ratio of weights is used to estimate percentages of acceptable responses for reporting groups or subpopulations of interest.[2]

The difference between the percentage of acceptable responses for a reporting group and that of the entire age group (nation) on an exercise describes the performance of any reporting group relative to the entire age group. This difference is a positive number if the group achieves a higher percentage than the entire age group and is a negative number if the group achieves a lower percentage. For example, a group performance of +1.8 indicates that the percentage of acceptable responses for the group is 1.8 percentage points higher than the national percentage of

---

[1] Appendix D discusses nonresponse in assessment samples.

[2] Following the 1976—77 assessment, a weighting-class adjustment procedure was used to smooth estimated population proportions across the eight assessments conducted between 1969—70 and 1976—77. Documentation of this procedure and estimated population proportions are included in Appendix F.

acceptable responses for a particular age level.

Increases or decreases in the percentage of acceptable responses between two assessments are estimated by finding the difference between percentages obtained from each assessment. A positive difference indicates an increase, and a negative difference indicates a decrease in the percentage of students who responded in a particular way from one assessment to the next. These differences, or change measures, are used to indicate trends in achievement, or performance, for an age level or subpopulation of interest. Changes in group differences from the national performance between two assessments are used to indicate the relative trend of a group compared with the national trend of the age group.

To present a general picture of changes in achievement, National Assessment summarizes the gains or losses on each exercise (either for the entire learning area or for some integral set of exercises) by using the mean, or arithmetic average, of the changes in percentages of acceptable responses to the exercises. During the first years of National Assessment, the median was used as the principal summary measure. However, the mean was chosen as the principal summary measure of change after extensive investigation showed empirically that it was more suitable for National Assessment change data than alternative measures.[3] In addition, the mean is an easily understood and well-known statistic and has simple arithmetic properties useful for the analysis of differences or change measures — in particular, the difference between means is the same as the mean difference. This property allows us to accurately describe the mean change as the difference between mean percentages of acceptable responses from one assessment to the next. Mean percentages are used to simplify descriptions of change and descriptions of differences in group performances. Differences in mean performance between reporting groups and the nation are used to summarize the performance of various groups relative to the nation. We do not intend that the mean percentage should be construed as an average test score.

Unless the items summarized in the mean percentages of acceptable responses are identical, the means of one age group should not be compared to the means of another, since their values reflect the choice of exercises in addition to the performance of the students. When only a few exercises are summarized by a mean, we should be especially cautious in interpreting results, since a small set of exercises might not adequately cover the wide range of potential behaviors included under a given objective or subobjective. The mean should be interpreted literally as the arithmetic average of the percentage of acceptable responses obtained from National Assessment samples on a specific set of exercises.

In considering National Assessment's achievement measures, it is the differences in performance between assessment years, among groups and among ages that are the most useful. By maintaining the same item or set of items in making these comparisons, we have a reasonable indicator of whether more or fewer people know or can do something judged important.

---

[3] Twenty-two empirical distributions of change measures from the 1969—70 and 1972—73 science assessments were used to generate Monte Carlo simulations of sampling distributions for several measures of central location. In addition to the mean and median, other measures of central location that were considered in the simulation studies included the average of the extremes, two forms of biweighted estimates and three forms of weight-matching estimators described by John W. Tukey in the research paper "Some Considerations on Locators Apt for Some Squeezed-Tail (and Stretched-Tail) Parents" (1975). In almost every case, the sampling stability of the mean change was as good as or better than that of the other measures studied.

Estimating Variability in Achievement Measures

National Assessment uses a national probability sample at each age level to estimate the proportion of people who would successfully complete an exercise. The particular sample selected is one of a large number of all possible samples of the same size that could have been selected with the same sample design. Since an achievement measure computed from each of the possible samples would differ from one sample to another, the standard error of this statistic is used as a measure of the sampling variability among achievement measures from all possible samples. A standard error, based on one particular sample, serves to estimate that sampling variability.

In the interest of sampling and cost efficiencies, National Assessment uses a complex, stratified, multistage probability sample design. Typically, complex designs do not provide for unbiased or simple computation of sampling errors. A reasonably good approximation of standard error estimates of acceptable response percentages is obtained by applying the jackknife procedure (Miller, 1964, pp. 1594–1705; Miller, 1968, pp. 567–82; Mosteller, 1968) to first-stage sampling units within strata. Standard errors for achievement measures such as group differences, mean percentages or mean group differences for a particular assessment year are estimated directly, taking advantage of features of the jackknife procedure that are generic to all of these statistics.[4] Since samples for different assessments are independent, the standard errors of the differences in achievement measures between assessments can be estimated simply by the square root of the sum of squared standard errors from each of the assessments.

The standard error provides an estimate of sampling reliability for the achievement measures used by National Assessment. It is comprised of sampling error and other random error associated with the assessment of a specific item or set of items. Random error includes all possible nonsystematic error associated with administering specific exercises to specific students in specific situations. Random differences among scorers for open-ended items are also included in the standard errors.

Differences in performance between assessment years or between a reporting group and the nation are reported as differences *only* if the differences are at least twice as large as their standard errors. Differences this large would occur by chance in fewer than 5% of all possible replications of our sampling and data collection procedures for any particular reporting group or national estimates.

Controlling Nonrandom Errors

Systematic errors can be introduced at any stage of an assessment — exercise development, preparation of exercise booklets, design or administration procedures, field administration, scoring or analysis. These nonsampling, nonrandom errors rarely can be quantified, nor can the magnitude of the bias they introduce into our estimates be evaluated directly.

Systematic errors can be controlled in large part by employing uniform administration and scoring procedures and by requiring rigorous quality control in all phases of an assessment. If the systematic errors are the same from age to age or group to group, then the differences in percentages or mean percentages are measured with reduced bias because subtraction tends to cancel the effect of the systematic errors.

Similarly, the effect of systematic errors in different assessment years can be controlled by carefully replicating in the second assessment the procedures carried out in the first. Differences in achievement across assessment years will also be measures with reduced bias since subtraction will again tend to cancel

[4] See Appendix E for a more detailed description of National Assessment's computation of standard errors.

systematic errors.

Although it is not possible for every condition or procedure to remain exactly the same between assessments conducted several years apart, National Assessment has made every effort to keep conditions as nearly the same as possible. Changes in procedures described in this report were judged to have a relatively minor impact.

CHAPTER 8

REPORTS ABOUT THE MATHEMATICS ASSESSMENTS

Each assessment generates a tremendous amount of data. To make these data as useful as possible to a variety of audiences, National Assessment provides several types of data reports.

Summary Reports

Since it is difficult and time consuming to synthesize many discrete bits of data, National Assessment prepares reports for the general public — including parents, classroom teachers, school administrators and legislators — that summarize and highlight assessment results. Summary reports for the mathematics assessment are organized by cognitive-process level.

Although National Assessment does not interpret assessment results, it recognizes that data presented alone are often difficult to consider in perspective. Accordingly, NAEP asked a group of mathematics educators to review and comment upon the results; their comments are included in the reports.

The following reports, describing 1977–78 status in mathematics achievement and changes between 1972–73 and 1977–78, are available: *Changes in Mathematical Achievement, 1973–78* (1979); *Mathematical Knowledge and Skills* (1979); *Mathematical Applications* (1979); and *Mathematical Understanding* (1979).

Objectives

A description of the 1977–78 mathematics objectives and the procedures used in developing the objectives and items for the 1977–78 assessment are available in: *Mathematics Objectives, Second Assessment* (1978).

Exercise Level and Summary Data

For those wishing to use specific National Assessment items, NAEP provides copies of released items, exercise documentation (including exercise timing, objective being measured, and so forth) and exercise-level data. The materials available are:

The Second Assessment of Mathematics, 1977–78, Released Exercise Set (1979) — this loose-leaf set provides copies of all released mathematics exercises from the 1977–78 mathematics assessment, exercise documentation (including timing, objective measured and exercise administration mode) and percentages of correct responses for the nation, regional groups, males and females and community-size groups.

Data Appendix for the 1977–78 Mathematics Released Exercise Set (1980) — this microfiched appendix provides additional data on the exercises included in the *Released Exercise Set*. It presents percentages of respondents giving each response alternative for an exercise at an age level for the nation and for 10 reporting variables: region, sex, race/ethnicity, level of parental education, type of community, size of community, grade level, modal

grade by region, modal grade by community size and modal grade by sex.

In addition to exercise-level data, National Assessment computes means for a learning area and for selected sets of exercises within the learning area. In mathematics, for example, means were computed for the various cognitive-process levels and content areas. These means are useful in comparing performance between assessment years and between reporting groups. Means are computed for a number of reporting variables, including region, sex, race/ethnicity, size and type of community and level of parental education, as well as additional variables on course-taking for 17-year-olds. These summary data are found in: *Mathematics Technical Report: Summary Volume* (forthcoming).

Public Use Data Tape

For those who wish to perform their own analyses of National Assessment data, NAEP makes available a data tape of respondent-level data for the 1977—78 mathematics assessment. To protect the confidentiality of the respondents, all identifying information (school, district, state) has been deleted. The tape also includes documentation of exercises and classification schemes used by NAEP to summarize exercises. The tape is organized and documented to be easy to use with standard statistical packages.

User Services

National Assessment provides some assistance to those wishing to use assessment items or to replicate assessment methodology. Those interested in receiving assistance should contact the National Assessment office.

APPENDIX A

MEMBERS OF THE MATHEMATICS ADVISORY BOARD

Sarah T. Herriot, Henry M. Guhn Sr. High School, Palo Alto, Calif.
Prince Jackson, Savannah State College, Savannah, Ga.
Jeremy Kilpatrick, University of Georgia, Athens, Ga.
Donald L. Kreider, Dartmouth College, Hanover, N.H.
John LeBlanc, Indiana University, Bloomington, Ind.
James W. Wilson, University of Georgia, Athens, Ga.

APPENDIX B

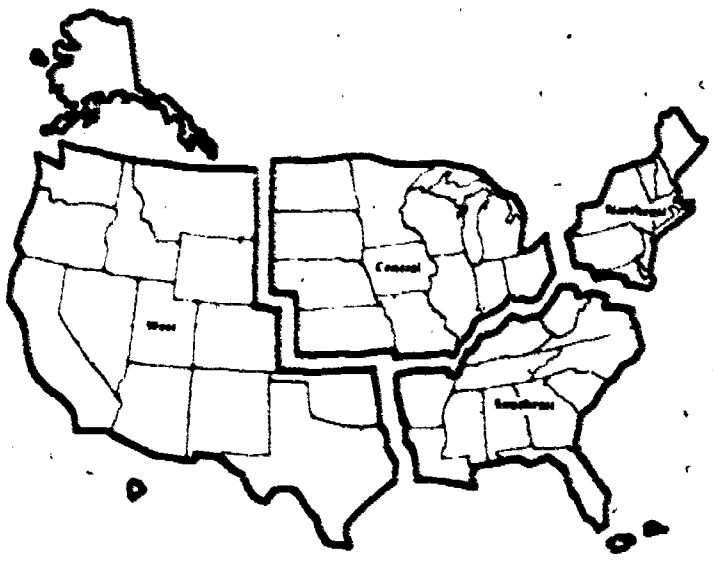
DEFINITIONS OF NATIONAL ASSESSMENT REPORTING GROUPS

In addition to reporting results for all 9-, 13- and 17-year-old students in the United States, National Assessment reports results for a number of population subgroups. Most of these subgroups are defined for both the 1972–73 and the 1977–78 mathematics assessments. Data on performance for different levels of course-taking by 17-year-olds are only available for the 1977–78 assessment.

Definitions of the subgroups follow:

Region

The country has been divided into four regions: Northeast, Southeast, Central and West. States included in each region are shown on the following map.



Sex

Results are reported for males and females.

Race

Results are presented for blacks, whites and Hispanos.[1]

Level of Parental Education

Three categories of parental-education levels are defined by National Assessment, based on students' reports. These categories are: (1) those whose parents did not graduate from high school, (2) those who have at least one parent who graduated from high school and (3) those who have at least one parent who has had some post-high-school education.

Type of Community

Communities in this category are defined by an occupational profile of the area served by a school as well as by the size of the community in which the school is located.

Advantaged-urban communities. Students in this group attend schools in or around cities having a population greater than 200,000 where a high proportion of the residents are in professional or managerial positions.

Disadvantaged-urban communities. Students in this group attend schools in or around cities having a population greater than 200,000 where a relatively high proportion of

[1] Data for Hispanos are reported only for sets of exercises and not for individual items because of small sample sizes.

the residents are on welfare or are not regularly employed.

Extreme-rural communities. Students in this group attend schools in areas with a population under 10,000 where many of the residents are farmers or farm workers.

Size of Community

Big cities. Students in this group attend schools within the city limits of cities having a 1970 census population over 200,000.

Fringes around big cities. Students in this group attend schools within metropolitan areas (1970 U.S. Bureau of the Census urbanized areas) served by cities having a population greater than 200,000 but outside the city limits.

Medium cities. Students in this group attend schools in cities having a population between 25,000 and 200,000, not classified in the fringes-around-big-cities category.

Small places. Students in this group attend schools in communities having a population less than 25,000, not classified in the fringes-around-big-cities category.

Grade in School

Results are categorized for 9-year-olds in the 3rd or 4th grade, 13-year-olds in the 7th or 8th grade, and 17-year-olds in the 10th, 11th or 12th grade.

Modal Grade by Region

Results are categorized for 9-, 13- and 17-year-old respondents in grades 4, 8 and 11, respectively, who live in the Northeastern, Southeastern, Central or Western regions of the country.

Modal Grade by Community Size

Results are categorized for 9-, 13- and 17-year-old respondents in grades 4, 8 and 11, respectively, who live in big cities, fringes around big cities, medium cities and small places.

Modal Grade by Sex

Results are categorized for 9-, 13- and 17-year-old males and females in grades 4, 8 and 11, respectively.

Mathematics Courses Taken by 17-Year-Olds

Seventeen-year-olds were asked about their mathematics course-taking experiences. A student was said to have taken a course if he or she had taken it for one-half year or longer. Five levels of mathematics-course experience are defined: (1) those who had not taken Algebra or any more advanced mathematics courses; (2) those who had taken Algebra I but no more advanced courses; (3) those who had taken Geometry but not Algebra II or any more advanced courses; (4) those who had taken Algebra II but no more advanced courses (Trigonometry, Probability and Statistics, or Pre-Calculus); and (5) those who had taken Algebra II and at least one more advanced course (Trigonometry, Probability and Statistics, or Pre-Calculus).

APPENDIX C

FORMS USED TO OBTAIN BACKGROUND INFORMATION

This appendix includes the forms used by National Assessment to collect background information from school officials and respondents for the 1977–78 assessment. Following are a listing and a brief description of the forms included.

School Principal's Questionnaire

This report is authorized by law (20 U.S.C. 1221 c-1). While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate, and timely.

Primary Sampling Unit ☐☐ School Number ☐☐☐☐☐☐☐☐☐

Age Group(s) 9 13 17

PLEASE PRINT

Name of School _____

Address of School _____
(Street)

_____
(City) (State) (Zip Code)

Name of School Principal _____

Name and title of person completing the form if other than school principal

Name _____ Title _____

1. What is your best estimate of the current enrollment and the average daily attendance by grade of your school (1977-78 school year)? (Enter zeros for grades not served by your school.)

Grade	K	1	2	3	4	5	6	7	8	9	10	11	12
Enrollment													
Average Daily Attendance													

2. Approximately what percentage of the students attending your school live in each of the following areas?

_____% A In a rural area (less than 2,500)

_____% B In a town of 2,500 to 10,000

_____% C In a town of 10,000 or more

_____
100% (Items A-C should add to 100%)

41

3. Approximately what percentage of the students attending your school are children of

____% A Professional or managerial personnel

____% B Sales, clerical, technical or skilled workers

____% C Factory or other blue collar workers

____% D Farm workers

____% E Persons not regularly employed

____% F Persons on welfare

____
100% (Items A-F should add to 100%)

4. Approximately what percentage of the students attending your school are

____% A American Indian or Alaskan Native

____% B Asian or Pacific Islander

____% C Hispanic, regardless of race (Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin)

____% D Black and not Hispanic

____% E White and not Hispanic

____
100% (Items A-E should add to 100%)

5. Does your school qualify for ESEA Title I assistance?

____ Yes - If Yes, approximately what number of students qualify for and what number of students are receiving ESEA Title I assistance?

____ Approximate number of students qualifying for ESEA Title I assistance

____ Approximate number of students receiving ESEA Title I assistance.

____ No

THANK YOU FOR YOUR COOPERATION

The purpose of this questionnaire is to provide additional information which will be used in the analysis of the mathematics data. Please answer the following questions for each school grade listed. Darken the appropriate ovals with a soft lead pencil. If you have questions about any of the following items, please contact the National Assessment District Supervisor. Thank you for your cooperation.

1. For each grade level in your school, how many minutes per week, on the average, does each student spend in a specifically scheduled mathematics class?

Minutes Per Week

	No such level	0	1-50	51-100	101-150	151-200	201-250	251-300	300 or more
A. Kindergarten	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
B. Grade 1	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
C. Grade 2	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
D. Grade 3	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
E. Grade 4	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
F. Grade 5	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
G. Grade 6	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

2. How has the time specifically scheduled for mathematics instruction changed in your school during the past five years?

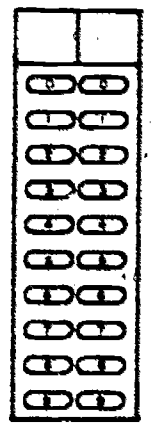
Increased Substantially	Increased Moderately	Remained Constant	Decreased Moderately	Decreased Substantially
⬭	⬭	⬭	⬭	⬭

3. How has the amount of instructional time spent on computational skills in grades K-6 changed in your school during the past five years?

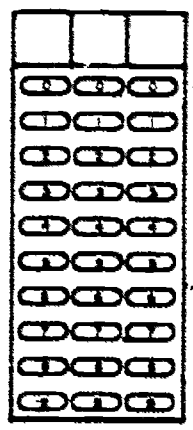
Increased Substantially	Increased Moderately	Remained Constant	Decreased Moderately	Decreased Substantially
⬭	⬭	⬭	⬭	⬭

School Name: _____

P.S.U. Number



School Number



Standard Background Information Form for 9-Year-Olds

1. Does your family get a newspaper regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

2. Does your family get any magazines regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

3. Are there more than 25 books in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

4. Is there an encyclopedia in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

5. How much school did your father complete?
   (FILL IN THE ONE OVAL which best shows how much school your father completed.)

   ⬭ Did not complete the 8th grade

   ⬭ Completed the 8th grade, but did not go to high school

   ⬭ Went to high school, but did not graduate from high school

   ⬭ Graduated from high school

   ⬭ Some education after graduation from high school

   ⬭ I don't know.

6. Did your father graduate from a college or university?

   ⬭ Yes ⬭ No ⬭ I don't know.

7. How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school
- ⬭ I don't know.

8. Did your mother graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

Standard Background Information Form for 13-Year-Olds

1. Does your family get a newspaper regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

2. Does your family get any magazines regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

3. Are there more than 25 books in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

4. Is there an encyclopedia in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

5. How much school did your father complete?
   (FILL IN THE ONE OVAL which best shows how much school your father completed.)
   - ⬭ Did not complete the 8th grade
   - ⬭ Completed the 8th grade, but did not go to high school
   - ⬭ Went to high school, but did not graduate from high school
   - ⬭ Graduated from high school
   - ⬭ Some education after graduation from high school
   - ⬭ I don't know.

6. Did your father graduate from a college or university?

   ⬭ Yes ⬭ No ⬭ I don't know.

7. How much school did your mother complete?
   (FILL IN THE ONE OVAL which best shows how much school your mother completed.)
   - ⬭ Did not complete the 8th grade
   - ⬭ Completed the 8th grade, but did not go to high school
   - ⬭ Went to high school, but did not graduate from high school
   - ⬭ Graduated from high school
   - ⬭ Some education after graduation from high school
   - ⬭ I don't know.

8. Did your mother graduate from a college or university?

   ⬭ Yes ⬭ No ⬭ I don't know.

9. Where did you live on your ninth birthday?

- ⬭ In the United States (Please specify the state or territory.)

  _____

- ⬭ Outside the United States (Please specify the country.)

  _____

- ⬭ I don't know.

A. How much school did your father complete?
(FILL IN THE ONE OVAL which best shows how much school your father completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school
- ⬭ I don't know.

B. Did your father graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

C. How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school
- ⬭ I don't know.

D. Did your mother graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

E. Where did you live on your ninth birthday?

⬭ In the United States (Please specify the state or territory.)

_____

⬭ Outside of the United States (Please specify the country.)

_____

⬭ I don't know.

F. Where did you live on your thirteenth birthday?

⬭ In the United States (Please specify the state or territory.)

_____

⬭ Outside of the United States (Please specify the country.)

_____

⬭ I don't know.

1. Which of the following does your family have at home? (Fill in one oval on each line.)

	Have	Do not have
A. Newspaper received regularly	⬭	⬭
B. Magazines received regularly	⬭	⬭
C. More than 25 books	⬭	⬭
D. Encyclopedia	⬭	⬭
E. Dictionary	⬭	⬭
F. Record player	⬭	⬭
G. Tape recorder or cassette player	⬭	⬭
H. Typewriter	⬭	⬭
I. Vacuum cleaner	⬭	⬭
J. Electric dishwasher	⬭	⬭
K. Two or more cars or trucks that run	⬭	⬭

2. How much time did you spend on homework yesterday?

⬭ No homework was assigned
⬭ I had homework but didn't do it
⬭ Less than one hour
⬭ Between 1 and 2 hours
⬭ More than 2 hours

3. How many different schools have you attended since you started the first grade?

- ⬭ 1 to 3 schools
- ⬭ 4 to 6 schools
- ⬭ 7 to 9 schools
- ⬭ 10 or more schools

4. How long have you lived in the community in which you now live?

- ⬭ All my life
- ⬭ 10 or more years but not all my life
- ⬭ 5 to 9 years
- ⬭ 2 to 4 years
- ⬭ 1 year
- ⬭ Less than 1 year

5. How much television did you watch yesterday?

⬭ None	⬭ 2 hours	⬭ 5 hours
⬭ 1 hour or less	⬭ 3 hours	⬭ 6 hours or more
⬭ 1 hour	⬭ 4 hours	

6. Is English the language spoken most often in your home?

⬭ Yes ⬭ No

7. Is a language other than English spoken in your home?

⬭ Often ⬭ Sometimes ⬭ Never

8. How many brothers or sisters do you have who are older than you?

None	1	2	3	4	5	6 or more
⬭	⬭	⬭	⬭	⬭	⬭	⬭

9. How many brothers or sisters do you have who are younger than you?

None	1	2	3	4	5	6 or more
⬭	⬭	⬭	⬭	⬭	⬭	⬭

10. A. What is your racial background?

- ⬭ American Indian or Alaskan Native
- ⬭ Asian or Pacific Islander
- ⬭ Black
- ⬭ White
- ⬭ Other

B. Is your ethnic heritage Hispanic (such as Mexican, Puerto Rican, Cuban, Central or South American or other Spanish Culture or origin)?

⬭ Yes ⬭ No

11. How often has each of the following been used in the courses you are taking this year? (Fill in one oval on each line.)

		Never	Seldom	Fairly Often	Frequently
A.	Listening to the teacher's lecture	⬭	⬭	⬭	⬭
B.	Participating in student-centered discussions	⬭	⬭	⬭	⬭
C.	Working on a project or in a laboratory	⬭	⬭	⬭	⬭
D.	Writing essays, themes, poetry, stories	⬭	⬭	⬭	⬭
E.	Going on field trips	⬭	⬭	⬭	⬭
F.	Having individualized instruction (small groups or one-to-one with a teacher)	⬭	⬭	⬭	⬭
G.	Using teaching machines or computer-assisted instruction	⬭	⬭	⬭	⬭
H.	Watching television lectures	⬭	⬭	⬭	⬭
I.	Studying from textbooks	⬭	⬭	⬭	⬭
J.	Library or media center assignments	⬭	⬭	⬭	⬭

For each of the following questions, fill in one oval in each box.

A. The metric system uses units like centimeters, liters and kilograms. Have you used the metric system of measurement?

Yes ⬭ No ⬭ I don't know. ⬭

B. How often have you used the metric system in mathematics?

Often ⬭ Sometimes ⬭ Never ⬭ I don't know. ⬭

C. How often have you used a hand calculator?

Often ⬭ Sometimes ⬭ Never ⬭ I don't know. ⬭

D. Do you or your family own a hand calculator?

Yes ⬭ No ⬭ I don't know. ⬭

E. Does your school have hand calculators that you can use in mathematics class?

Yes ⬭ No ⬭ I don't know. ⬭

For each of the following questions, fill in one oval in each box.

A. The metric system of measurement uses units like centimeters, liters, and kilograms. How often have you used the metric system?

Often ⬭ Seldom ⬭ Never ⬭ I don't know. ⬭

B. How often do you use a hand calculator?

Almost Daily ⬭ A few times a week ⬭ Less than once a week ⬭ Once a month ⬭ Never ⬭ I don't know. ⬭

C. Do you or your family own a hand calculator?

Yes ⬭ No ⬭ I don't know. ⬭

D. Does your school provide hand calculators for use in mathematics class?

Yes ⬭ No ⬭ I don't know. ⬭

For each of the following questions, fill in one oval in each box.

A. The metric system of measurement uses units like centimeters, liters, and kilograms. How often have you used the metric system of measurement?

Often ⬭ Seldom ⬭ Never ⬭ I don't know. ⬭

B. How often do you use a hand calculator?

Almost Daily ⬭ A few times a week ⬭ Less than once a week ⬭ Once a month ⬭ Never ⬭ I don't know. ⬭

C. Do you or your family own a hand calculator?

Yes ⬭ No ⬭ I don't know. ⬭

D. Does your school provide hand calculators for use in mathematics classes?

Yes ⬭ No ⬭ I don't know. ⬭

E. Does your school provide hand calculators for use in other classes?

Yes ⬭ No ⬭ I don't know. ⬭

Which of the following mathematics courses have you studied? Fill in one oval on each line. (If you have not studied a particular course, fill in the oval under "Not Studied".)

		Studied 1 school year	Studied ½ school year	Studied less than ½ year	Not studied	I don't know.
A.	General, Business or Consumer Mathematics	⬭	⬭	⬭	⬭	⬭
B.	Introduction to Algebra (Pre-Algebra)	⬭	⬭	⬭	⬭	⬭
C.	1st year Algebra	⬭	⬭	⬭	⬭	⬭
D.	2nd year Algebra	⬭	⬭	⬭	⬭	⬭
E.	Geometry	⬭	⬭	⬭	⬭	⬭
F.	Trigonometry	⬭	⬭	⬭	⬭	⬭
G.	Probability & Statistics	⬭	⬭	⬭	⬭	⬭
H.	Computer Programming	⬭	⬭	⬭	⬭	⬭
I.	Pre-Calculus/ Calculus	⬭	⬭	⬭	⬭	⬭

APPENDIX D

NONRESPONSE IN ASSESSMENT SAMPLES

In addition to sampling variability, estimates of population values computed from sample surveys might be subject to random error and systematic bias. Systematic bias, or nonrandom error, might result from estimation procedures, errors inherent in measurement and data collection procedures, and nonresponse. This appendix examines nonresponse in the 1972–73 and 1977–78 assessments. Since nonresponse rates at ages 9 and 13 are relatively small, the following discussion concerns 17-year-olds' response rates only.

Bias due to nonresponse is present in virtually every sample survey but is frequently ignored since it is difficult to estimate its size. A variety of factors contribute to nonresponse. Nonrespondents might be either difficult to notify or reluctant to participate once they are notified; some might be absent from school during the entire contact period with item administrators. However, these nonrespondents can be important, since, if they respond differently than the people actually included in the sample, estimates of percentage based solely on the sample are biased and not properly representative of the age population being assessed.

To provide some information about the size of the bias due to nonresponse in National Assessment surveys, the Research Triangle Institute, Raleigh, North Carolina, was asked to conduct a special study of nonrespondents during the 1972–73 assessment of science and mathematics. The study was conducted on the age population of eligible 17-year-olds who, at the time of the assessment, were listed as enrolled in school. Some of these students, in fact, were no longer attending school at the time of the assessment. Eligibles had to be English-speaking, had to be physically and emotionally able to respond to exercises as administered, and could not reside in an institution.

The results of the nonresponse study (Kalsbeek et al., 1975; Rogers et al., 1977) indicate that 17-year-olds listed as enrolled in schools but not appearing at the designated time of assessment can be divided into two different groups. The first group of nonrespondents, which comprises about 80% of the total nonrespondent group, did not appear for the assessment because of conflicting school activities or illness. The performance of this group was not very different from the performance of students assessed during the regularly scheduled sessions. The second group of nonrespondents, which comprises about 20% of the nonrespondents, do not appear to be available in the schools at any time. They attend infrequently if they attend at all (for practical purposes, they have dropped out of school), or they have moved out of the school attendance area. In either case, these students should probably not have been listed in the in-school population of eligibles. This group, in contrast to the group of nonrespondents who were, in fact, attending school, performed more poorly on assessment questions than students assessed during the scheduled sessions.

The weights used by National Assessment to estimate the percentage of acceptable responses are adjusted for nonresponse. The adjustment assumes that the nonrespondents would perform, on the average, in a manner similar to those who did respond. However, the nonresponse study showed that the

second group of nonrespondents, those enrolled in but not actually attending school, typically performed at a lower level than either those who did respond or the first group of nonrespondents. If the second group is included in the population of eligibles, the nonresponse adjustment procedure used by National Assessment would result in overestimates of the true percentages of acceptable responses.

Because the second group of students is effectively no longer attending school, it does not seem appropriate to include them in estimates for 17-year-olds in school. Thus, these students are not considered part of the population of eligibles and are excluded from the computations of percentage of the sample covered for 17-year-olds, as shown in Table D-1.

Including the second group of students and then reducing bias due to their nonresponse would require the location and testing of some of these individuals. The difficulty and costs associated with supplementary data collection of the nonrespondents not actually attending school are so great that this has not been a feasible alternative in recent years.

Table D-1 shows the average sample coverage per package (booklet) of exercises administered in 1972–73 and 1977–78. The rate of coverage is based on an estimated total eligible age population of students who are available in school — for 17-year-olds, those enrolled minus the 20% estimated to be enrolled but unavailable in school.

TABLE D-1. Number of Students Assessed and Percent of Sample Covered, by Age and Assessment Year

Year	Age	Number of Packages	Total Number of Students Assessed	Average Number Assessed Per Package	Average Sample Coverage in Percent
1972–73	9	7	18,638	2,663	91.0
	13	9	23,507	2,612	84.6
	17	11	25,865	2,351	77.7
1977–78	9	7	17,190	2,456	87.2
	13	11	26,661	2,424	85.2
	17	12	26,757	2,230	73.2

Note: Computations for 17-year-olds do not include those enrolled in school but not actually attending.

APPENDIX E

NATIONAL ASSESSMENT ESTIMATION OF STANDARD ERRORS

Several measures of achievement that National Assessment uses in its reports are described in Chapter 7. The sample design described in Chapter 4 is a complex, deeply stratified, multistage probability sample design. A reasonably good approximation of standard error estimates of these achievement measures can be obtained by applying the jackknife procedure to first-stage sampling units within strata, using the method of successive differences and accumulating across strata.

In this section, the measures of achievement are first defined in algebraic form, followed by a description of the jackknife method used by National Assessment to estimate their standard errors.

Measures of Achievement

Based on the sample design, a weight is assigned to every individual who responds to an exercise administered in an assessment. The weight is the reciprocal of the probability of selecting a particular individual to take a particular exercise with adjustment for nonresponse. Since the probabilities of selection are based on an estimated number of people in the target age population, the weight for an individual estimates the number of similar people that that individual represents in the age population. As explained in Appendix F the weights were adjusted to reflect information from previous assessments on population distributions.

A sum of the weights for all individuals at an age level responding to an exercise is an estimate of the total number of people in that age population. A sum of weights for all individuals at an age responding correctly to an exercise is an estimate of the number of people who would be able to respond correctly in the age population if the entire population were assessed. These concepts also apply to any reporting group (e.g., defined by region, sex, etc.) and category of response (e.g., correct, incorrect and "I don't know").

Let W^e_{ihk} = sum of weights for respondents to exercise e who are in reporting subgroup i and who are in the kth PSU of the hth sampling stratum, and

C^{ej}_{ihk} = sum of weights for respondents to exercise e who are in subgroup i, who are in the kth PSU of stratum h and who selected response category j (e.g., correct foil) for the exercise.

Note that $W^e_{ihk} = \sum_j C^{ej}_{ihk}$

Then, summing K over the n_h, sample PSUs in the stratum h, and summing over the H sampling strata, $W^e_{i++} = \sum_{h=1}^{H} \sum_{k=1}^{n_h} W^e_{ihk}$ estimates the number of eligibles in the population who are in subgroup i.

Similarly, $C^{ej}_{i++} = \sum_{h=1}^{H} \sum_{k=1}^{n_h} C^{ej}_{ihk}$ estimates the number of eligibles in the population who are in subgroup i and who would se-

lect response category j for exercise e.

An estimate of the proportion of the eligibles in the age population in group i who would select response category j on exercise e is:

$$(1) \quad P_{i\bullet}^{ej} = C_{i++}^{ej} \; / \; W_{i++}^{e}.$$

In the special case where the proportion of all age eligibles who would select response category j on exercise e is estimated, the index A (for ALL) will be used in place of i as follows:

$$(2) \quad P_{A}^{ej} = C_{A++}^{ej} \; / \; W_{A++}^{e}$$

In National Assessment reports, the proportion in (1) multiplied by 100 is called the group percentage, and the proportion in (2) multiplied by 100 is called the national percentage. The difference between the proportion in subgroup i who would select category j on exercise e and the proportion in the nation is denoted by:

$$(3) \quad \Delta P_{i}^{ej} = P_{i}^{ej} - P_{A}^{ej}.$$

National Assessment also reports the arithmetic mean of the percentage of correct responses over sets of exercises corresponding to the measures in (1), (2) and (3). These means are taken over the set of all exercises or a subset of exercises classified by a reporting topic or content objective. The mean percentage of correct responses taken over m exercises in some set of exercises corresponding to measures (1), (2) and (3) are, respectively:

$$(4) \quad \overline{P}_{i} = \frac{1}{m} \sum_{e} C_{i++}^{e} \; / \; W_{i++}^{e},$$

$$(5) \quad \overline{P}_{A} = \frac{1}{m} \sum_{e} C_{A++}^{e} \; / \; W_{i++}^{e} \text{ and}$$

$$(6) \quad \overline{\Delta P}_{i} = \overline{P}_{i} - \overline{P}_{A}.$$

Note that the response category subscript j has been suppressed since the means are understood to be taken over the correct response category for each exercise.

Each of these six achievement measures is computed and routinely used in reports describing achievement data for any assessment. The simple difference in these measures between two assessments of the same exercise (or sets of exercises) provides six measures of change in achievement that are routinely used in National Assessment's change reports. The next section describes how standard errors are estimated for the 12 statistics routinely used in NAEP reports.

Computation of Standard Errors

In order to obtain an approximate measure of the sampling variability in the statistics (1) through (6), a jackknife replication procedure for estimating the sampling variance of nonlinear statistics from complex, multistage samples was tailored to National Assessment's sample design. References (4), (5) and (7) provide information about the jackknife technique, while reference (3) describes how the procedure is used in estimating standard errors for National Assessment's sample designs.

To demonstrate the computational aspects of this technique, consider estimating the variance of the statistic in (1) — the proportion of age-eligibles in subgroup i who would select response category j on exercise e.

This statistic is based on the data from all the n_h PSUs in the H strata. Let p_{i-hk}^{ej} be defined as a replication estimate of p_{i}^{ej} and constructed from all the PSUs, excluding the data from PSU k in stratum h. These replication estimates are computed as if the excluded PSU had not responded and a reasonable nonresponse adjustment is used to replace the data in PSU hk in estimating p_{i}^{ej}. Several choices for replacing the data in PSU hk are available. In order to obtain a convenient and computationally efficient algorithm for approximating standard errors,

National Assessment replaces C^{ej}_{ihk} and W^{e}_{ihk} from the hkth PSU with corresponding sums from another paired PSU in the same stratum. The replicate estimate is then computed. The replicate estimates to be used in the calculations are determined by arranging all the PSUs in each stratum into successive pairs. That is, PSU 1 is paired with PSU 2, PSU 2 with PSU 3, 3 with 4, . . . (n_h-1) with n_h and PSU n_h with PSU 1.

The contribution to the variance of p^{ej}_i by each pair of PSUs is the change in the value of the statistic incurred by replacing the data from each PSU in the pair with the data from the other PSU in the pair and recomputing p^{ej}_i in the usual way. This produces two replicate estimates. Squaring the difference between these replicate estimates and then dividing by eight measures the contribution of this pair of PSUs to the total variance. The sum of these contributions over all n_h successive pairs in the stratum is the contribution by stratum h to the total variance. The square root of the sum of the H stratum contributions is the estimate of the standard error of p^{ej}_i

Algebraically, the two replicate estimates for the pair k, $k+1$ (where $k=1, \ldots n_h$ and $n_h+1=1$) are:

$$(7)\quad p^{ej}_{i-hk} = \frac{C^{ej}_{i++} - C^{ej}_{ihk} + C^{ej}_{ih(k+1)}}{W^{ej}_{i++} - W^{ej}_{ihk} + W^{ej}_{ih(k+1)}}$$

and

$$(8)\quad p^{ej}_{i-h(k+1)} = \frac{C^{ej}_{i++} - C^{ej}_{ih(k+1)} + C^{ej}_{ihk}}{W^{ej}_{i++} - W^{ej}_{ih(k+1)} + W^{ej}_{ihk}}$$

The contribution to the total variance from stratum h is:

$$(9)\quad \text{var}\ (P^{ej}_{ih}) = \frac{1}{8} \sum_{k}^{n_h} \left(p^{ej}_{i-hk} - p^{ej}_{i-h(k+1)} \right)^2.$$

And, finally, an estimate of the standard error of p^{ej}_i is:

$$(10)\quad SE\ (p^{ej}_i) = \left(\sum_{h}^{H} \text{var}\ p^{ej}_{ih} \right)^{1/2}.$$

Multiplying p^{ej}_i by 100 yields the percentage of response to category j. Multiplying $SE\ (p^{ej}_i)$ by 100 yields the corresponding estimated standard error of the percentage.

In general, the jackknifed standard errors of the proportion estimates will be larger than the simple random sampling formula $(pq/n)^{1/2}$, where $p=p^{ej}_i$, $q=1-p$ and n is the number of sampled respondents in subgroup i who took the exercise. The larger the size of $SE\ (p^{ej}_i)$ reflects mainly the loss of precision due to cluster-sampling of schools and students.

The standard errors for the achievement measures (2) through (6) are computed through a series of steps analogous to those followed in computing $SE\ (p^{ej}_i)$. The most complicated step in computing standard errors occurs in forming the paired replicate estimates analogous to (7) and (8) for each successive pair of PSUs. Once this bookkeeping chore is done, the computations (9) and (10) follow in a straightforward manner.

The standard errors for the differences between two assessments for any of the achievement measures (1) through (6) are computed as the square root of the sum of the squared standard errors from each of the separate assessments.

The size of the standard errors depends largely not only on the number of PSUs and schools included in the sample (Table 2, Chapter 4), but also on the number of respondents in each of the reporting groups. Table E-1 shows the average number of students responding to an exercise package for each of the reporting groups for each age and for each of the two mathematics assessments.

The size of the standard errors of the means of the achievement measures for sets of exercises is also influenced by the number of exercises in the exercise set and the number of packages over which the items in the set

are spread. Table E-2 shows the number of exercises and packages included in the mean achievement measures used to measure changes between mathematics assessments for each of the cognitive-process levels.

REFERENCES

Folsom, R.E. Jr. *National Assessment Approach to Sampling Estimation*, Sampling Error Monograph 25U-796-5. Research Triangle Park, N.C.: Research Triangle Institute, 1977.

Miller, R.G. Jr. "A Trustworthy Jackknife," *Annals of Mathematical Statistics*, no. 39 (1968), pp. 567–82.

Miller, R.G. Jr. "The Jackknife – A Review," *Biometrika*, no. 61 (1974), pp. 1–15.

Mosteller, F. and J.W. Tukey. *Data Analysis and Regression* (Chapter 8). Reading, Mass.: Addison-Wesley, 1977.

TABLE E-1. Average Number of Respondents in Reporting Groups Taking an Item Booklet, by Age and Assessment Year

	Age 9		Age 13		Age 17	
	1972–73	1977–78	1972–73	1977–78	1972–73	1977–78
Nation	2,663	2,456	2,612	2,424	2,351	2,230
Region						
Northeast	656	552	651	547	573	516
Southeast	669	591	667	546	596	509
Central	672	764	649	770	596	731
West	665	549	645	561	586	474
Sex						
Male	1,328	1,234	1,294	1,210	1,126	1,085
Female	1,335	1,221	1,318	1,214	1,225	1,145
Race						
White	1,997	1,843	1,977	1,862	1,852	1,796
Black	466	433	436	396	358	306
Parental education						
Not graduated high school	271	204	417	302	455	334
Graduated high school	564	579	792	797	720	763
Post high school	787	772	994	942	1,028	1,020
Type of community						
Extreme rural	265	240	263	242	230	221
Disadvantaged urban	266	245	264	241	239	214
Advantaged urban	267	247	260	243	234	227
Size of community						
Big cities	619	640	583	685	439	612
Fringes around big cities	515	524	531	519	493	483
Medium cities	372	170	365	170	326	126
Small places	1,157	1,122	1,133	1,049	1,094	1,008
Grade						
3, 7, 10	646	634	693	652	305	303
4, 8, 11	1,946	1,755	1,809	1,704	1,688	1,675
12	--	--	--	--	304	220

TABLE E-2. Number of Exercises in Various Cognitive-Process Levels Used To Measure Change Between 1972–73 and 1977–78 and Number of Packages in Which Those Exercises Appeared in Each Assessment, by Age

	Age 9			Age 13			Age 17		
	Number of Exercises	Number of Packages		Number of Exercises	Number of Packages		Number of Exercises	Number of Packages	
		1972–73	1977–78		1972–73	1977–78		1972–73	1977–7
All exercises	55	7	6	77	9	9	102	11	11
Type of mathematics									
Knowledge	17	4	4	16	4	4	18	8	8
Skills	21	7	6	37	9	9	46	11	11
Understanding	8	5	5	12	8	8	13	9	9
Applications	9	4	4	12	7	7	25	10	10

APPENDIX F

ADJUSTMENT OF RESPONDENT WEIGHTS AND ESTIMATED POPULATION PROPORTIONS OF REPORTING GROUPS BASED ON NATIONAL ASSESSMENT SAMPLES

As noted in Appendix D, a weight is assigned to every individual who responds to an exercise administered in an assessment. Because the weight is the reciprocal of the probability of selection of the individual with adjustment for nonresponse, and since the probabilities of selection are based on the estimated number of people in the target age populations, the weight for an individual estimates the number of similar people that that individual represents in the age population.

The estimated population proportions for reporting groups shown in this appendix are based on weights derived from the sampling process used in the two assessments of 9-, 13- and in-school 17-year-olds. These proportions vary from year to year because of random sampling variability or systematic differences in sampling procedures. A better estimate of population proportions for any single year can be obtained by smoothing[1] the proportions over several assessment years. Smoothing does not make the estimated proportions identical but does reduce variability. The estimated population proportions shown in this appendix and used in estimating performance were obtained after smoothing proportions from the first nine years of assessment. The procedures used to obtain the smoothed proportions are detailed below.

The purpose of smoothing estimated population proportions is to reduce sampling fluctuations that can affect estimates of the change over time in the percentage of acceptable responses to an exercise. For example, the percentage of acceptable responses for an age group is a function of the relative proportions of high-performing and low-performing groups. If the relative proportions of these groups are very different in different assessments because of sampling variability, then a portion of the change in percentage of acceptable responses for an age group is directly attributable to the yearly sampling difference in the relative proportions of high- and low-achieving groups. Smoothing estimates of population proportions reduces a large portion of the sampling variability while preserving, as far as possible, actual trends occurring in the age population.

Besides reporting performance estimates for an age group as a whole, National Assessment also reports performance for various subpopulations, such as whites or blacks. Because variability of subgroups within these subpopulations (such as white males and females within the white subpopulation) influences the performance estimates for the subpopulations, it is desirable that fluctuations of proportions of all subgroups of each subpopulation be reduced as much as possible.

The most direct way to accomplish this is first to classify people into mutually exclusive multiway cells on the basis of their membership in categories of various important

[1] The word "smoothing" is used here in the sense of drawing a "smooth" curve to fit a sequence of numbers. Proportions for each reporting group covering eight years were smoothed by the robust/resistant procedures (Tukey, 1977, Chapter 7).

variables and then to smooth the proportions within each of the resulting multiway cells across years. Unfortunately, this procedure tends to produce a large number of cells with few people and consequently, quite unstable estimates of smoothed proportions. To circumvent this difficulty, the smoothing procedure, which is basically a weighting-class adjustment applied independently to each age, is designed to control, to varying degrees, fluctuations in certain key subgroups while avoiding, as much as possible, instabilities due to small cells. The procedure involves two distinct levels of partitioning of the population.

The first step in the weight adjustment process involves partitioning the population of age class eligibles into 672 six-way cells on the basis of membership in categories of six key demographic variables: race, grade, sex, size and type of community, region and parental education. The categories of these variables that were used in the partitioning are presented in Table F-1. Then, for each of the three ages and each of the nine years, the proportion of the population in each of the six-way cells was estimated. For a given age and year, the proportion of the population in a particular six-way cell was computed as the sum of weights of all respondents assessed in the given year who were of the specified age and who belonged in the six-way cell divided by the total of the weight of all respondents of the given age assessed in that year.

The second step in the weight adjustment process is the aggregation of the six-way cells into 17 larger cells, called major cells. These major cells, which comprised the first level of partitioning for weight smoothing, were the same for all three ages and all nine years and are detailed in Table F-2. The purpose of the aggregation to major cells is to obtain subgroups of the populations that have historically exhibited substantial differences in performance and that define groupings for which the most control of the weight was desired. In order to ensure stability in the weight adjustment process, each major cell was required to contain an adequate number of respondents for each age and year.

TABLE F-1. Categories of the Variables Used in the Construction of Six-Way Cells

Variables	Categories
Race	White, black, other
Grade	Less than modal grade, modal grade or greater
Sex	Male, female
Size and type of community	Extreme rural, disadvantaged urban, advantaged urban, main big cities, urban fringes, medium cities, small places
Region	Northeast, Southeast, Central, West
Parental education	Graduated high school and less, post high school

To exert control on the six-way cells within each of the major cells, a version of the Smear-and-Sweep procedure developed in the National Halothane Study (Bunker et al., eds., 1969) was used. National Assessment's use of this procedure involved ordering the six-way cells within a major cell by predicted performance and then aggregating adjacent six-way cells into supercells. Each supercell was required to represent no less than 1% of the total population in each year. The ordering of six-way cells and aggregating into supercells was done independently for each age. There were between 1 and 80 six-way cells in a supercell and between 1 and 16 supercells in a major cell. There were a total of 66, 74 and 65 supercells created for ages 9, 13 and 17, respectively.

The partitioning of the major cells into supercells allowed some dampening of the fluctuations of proportions of the six-way cells, and hence, of all one-way margins. At the same time, instabilities of adjustment due to small cell size was reduced by using supercells instead of the six-way cells. Additionally, by aggregating six-way cells with similar performance, bias in performance estimates using

TABLE F-2. Major Cells Used in the Adjustment of Population Proportions

Major Cell	Race	Region*	Size and Type of Community†	Grade**	Sex
1	White	SE	DU	All	All
2	White	SE	AU	All	All
3	White	SE	not DU,AU	GEMG	Male
4	White	SE	not DU,AU	GEMG	Female
5	White	SE	not DU,AU	LTMG	All
6	White	not SE	DU	All	All
7	White	not SE	AU	All	All
8	White	not SE	Not DU,AU	GEMG	Male
9	White	not SE	not DU,AU	GEMG	Female
10	White	not SE	not DU,AU	LTMG	Male
11	White	not SE	not DU,AU	LTMG	Female
12	Black	SE	DU	All	All
13	Black	SE	not DU	All	All
14	Black	not SE	DU	All	All
15	Black	not SE	not DU	All	All
16	Other	All	DU	All	All
17	Other	All	not DU	All	All

*SE=Southeast.
†DU=disadvantaged urban; AU=advantaged urban.
**GEMG=modal grade or greater; LTMG=less than modal grade.

smoothed proportions was reduced.

The next step in the smoothing procedure is the actual smoothing of the supercell proportions. For each age, there was a set of nine proportions for each supercell, representing the proportion of people in the supercell for each of the nine assessment years.

Each such set of proportions was then smoothed by robust/resistant methods to give a sequence of adjusted population proportions that tended to preserve actual time trends in proportions while reducing the sampling variability of these estimates over time. The adjusted proportions were constrained by requiring that the sum of adjusted proportions across all supercells for each age and year total one.

The final step in the smoothing procedure is to adjust the respondents' weights to be consistent with the smoothed supercell proportions. Since each respondent takes only one package, the weight adjustments were done independently for each package. For a given age, year and package, population proportions using the original weights were obtained for each of the supercells appropriate for the age. Then the weights of all respondents of a given supercell were adjusted by multiplying by a factor that is the ratio of the smoothed supercell proportion divided by the proportion using the original weights. This produced the adjusted weights that were used in all analyses.

The population proportions presented in Table F-3 were computed using these adjusted weights. All performance estimates were also computed using the adjusted weights. As explained in Appendix E, the percentage of correct response is estimated by dividing the sum of weights for students responding correctly to an exercise by the sum of weights for all students exposed to the exercise. For example, in 1977, 11 packages of exercises were administered to 13-year-olds. Smoothed

population proportion estimates were based on 26,661 13-year-olds, but each exercise-specific performance estimate was based on the approximately 2,424 13-year-olds who took a particular package. Consequently, respondent weights were adjusted to dampen both between-year variability and package-to-package variability within an assessment year.

The result of the smoothing and weight-adjustment process is that both adjusted performance estimates and changes in those estimates appear to be somewhat less susceptible to sampling variability, both across and within years.

TABLE F-3. Estimated Population Proportions of National Assessment Reporting Groups for Ages 9, 13 and 17 in 1972–73 and 1977–78

Reporting Groups	1972–73 Age 9	1972–73 Age 13	1972–73 Age 17	1977–78 Age 9	1977–78 Age 13	1977–78 Age 17
Sex						
Male	.500	.498	.484	.502	.502	.484
Female	.500	.502	.516	.498	.498	.516
Race						
White	.808	.816	.849	.799	.811	.832
Black	.133	.127	.109	.133	.125	.113
Other	.059	.057	.042	.068	.064	.055
Region						
Northeast	.257	.241	.259	.239	.236	.240
Southeast	.223	.220	.199	.234	.224	.205
Central	.280	.300	.311	.289	.292	.314
West	.240	.239	.231	.238	.248	.241
Parental education						
Not graduated high school	.096	.146	.175	.089	.132	.148
Graduated high school	.222	.302	.320	.245	.321	.326
Post high school	.320	.416	.458	.328	.420	.480
Unknown	.362	.136	.047	.338	.127	.046
Type of community						
Extreme rural	.085	.092	.084	.084	.098	.085
Disadvantaged urban	.074	.074	.090	.075	.067	.085
Advantaged urban	.122	.114	.114	.106	.109	.111
Other	.719	.720	.712	.735	.726	.719
Size of community						
Big cities	.211	.194	.181	.190	.200	.173
Fringes around big cities	.227	.233	.250	.213	.213	.235
Medium cities	.141	.134	.141	.138	.135	.142
Small places	.421	.439	.428	.459	.452	.450
Grade in school						
<3,<7,<10	.010	.027	.017	.006	.021	.015
3, 7, 10	.223	.246	.126	.233	.254	.133
4, 8, 11	.755	.716	.729	.751	.716	.752
<4,<8, 12	.006	.010	.127	.006	.008	.100
Other	.006	.001	.001	.004	.001	.000

GLOSSARY OF NATIONAL ASSESSMENT TERMS

Acceptable response. Any response to an exercise that demonstrates achievement of the objective measured by that exercise.

Administration time. The total time allowed on the paced audio tape for an exercise. (Includes the time allowed for the stimulus and the response.)

Administration timetable. Time periods during the school year when the various age groups are assessed. For the mathematics assessment the timetable was as follows:

Age	First Mathematics Assessment	Second Mathematics Assessment
9	January to February 1973	January to February 1978
13	October to December 1972	October to December 1977
17	March to May 1973	March to May 1978

In the first mathematics assessment, young adults were assessed from October 1972 to May 1973.

Age group or age level. Three age groups were sampled in the 1977—78 assessment: 9-year-olds, 13-year-olds and 17-year-olds.

In 1972—73 young adults aged 26—35 were also sampled. For the 1977—78 assessment, the three age groups were defined as follows:

Age 9 — born during the calendar year 1968

Age 13 — born during the calendar year 1964

Age 17 — born from October 1, 1960, through September 30, 1961

Assessment. The documentation of the progress in knowledge, skills and attitudes of American youth. Measures are taken at periodic intervals for each learning area, with the goal of determining trends and reporting the findings to the public and to the education community.

Assessment administrator. Individual employed to administer the assessment in participating schools.

Attitudes. The mathematics assessment surveyed attitudes toward mathematics in school, mathematics in society, mathematics as a discipline and mathematics in relationship to oneself.

Background questions. Questions specific to mathematics were included in the item booklets and dealt with respondents' education and experience in mathematics. Standard background questions asked in every learning area are found on the back pages of the item booklets and include such things as level of parental education and reading materials in the home. Samples of these questions are found in Appendix C.

Booklet. Items (exercises) are presented to respondents in booklets. Booklets are designed to be scored by optical scanning machines. Each booklet contains (1) instructions on answering items and sample items, (2) assessment items and (3) background questions. Each booklet contains approximately 30—35 minutes of assessment items and 10—15 minutes of introductory material and background questions. A booklet typically includes exercises of varying difficulty from different objectives of the learning area being

Calculator booklet. An entire booklet at each age level was designed to be answered using a four-function electronic hand-held calculator. Calculators were issued to students taking these booklets.

Calculator item. An item in a calculator booklet designed to be answered using a hand-held calculator.

Category (scoring). A classification of a response to an open-ended exercise. See *scoring guide.*

Category within a variable. A subclassification within a variable. For example, male and female are categories of the variable, sex. Also see *reporting groups.*

Cognitive-process level. Different types of mental processes required for mathematics problems. These levels form one dimension of the two-dimensional matrix used to structure the mathematics assessment. The four cognitive-process levels defined for the mathematics assessment are: knowledge, skills, understanding and application.

Content category or area. Classification of the content included in the mathematics assessment. Content forms one dimension of the two-dimensional matrix used to structure the mathematics assessment. The five content categories defined were: numbers and numeration; variables and relationships; geometry (shape, size and position); measurement; and other topics (including graphs and tables, probability and statistics, and logic).

Exercise. A task designed to measure an objective. Because NAEP does not administer "tests," but instead describes education achievement over time, the term "exercise" is often used instead of the term "item" or "test item." The terms "item" and "exercise" are used synonomously in this report.

Exercise booklet. See *booklet.*

Exercise part. See *item part.*

Exercise pool. The entire set of exercises prepared for a learning area. This set includes recycled exercises, exercises developed for previous assessments but not used due to exercise booklet or budgetary constraints, and newly developed exercises.

Field trials (tryouts). A pretest of exercises to obtain information regarding clarity, difficulty levels, timing, feasibility and special administration problems needed for revision and selection of exercises to be used in the assessment.

Grade. Results are reported for 9-year-olds in the 3rd and 4th grades, 13-year-olds in the 7th and 8th grades, and 17-year-olds in the 10th, 11th and 12th grades.

Group administration. Booklets were administered to groups of between 10 and 25 students in 1977—78. In 1972—73, group size was 12 respondents.

Hand scoring (scoring). The coding of responses in a format compatible with the optical scanning equipment being used. Multiple-choice exercises can be directly machine scored; however, responses to open-ended exercises must be coded in scoring ovals so that they can be machine scored. See *scoring guide.*

ID number. An identification number. Usually refers to the unique number assigned to each respondent. This number is assigned to preserve the anonymity of each respondent. NAEP does not keep records of the names of any individuals.

Item. See *exercise.*

Item booklet. See *booklet.*

Item part. Each part of an item asks a separate question. Parts may all pertain to one stimulus, such as a graph or a table, or may instead concern the same topic. In the mathematics assessment, each item part

was analyzed as a separate unit of data, so counts for various content categories and cognitive-process levels are done by item part. Same as *exercise part*.

Learning area. One of the 10 areas assessed by National Assessment: art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies and writing.

Level of parental education. Three levels of parental education are reported. These levels are described in Appendix B.

Mathematics assessment. The content domain for the second assessment of mathematics draws primarily from the current curriculum of elementary and secondary schools, although some projection of future mathematics emphases is included (for example, greater use of metric measure, earlier introduction of decimals and calculator computation). Mathematics up to, but not including, Calculus is included in the assessment.

Modal grade. The grade in which most of the students at an age level are enrolled. For 9-year-olds, the modal grade is the 4th grade; for 13-year-olds, the 8th grade; and for 17-year-olds, the 11th grade.

Objective. A desirable education goal agreed upon by scholars in the field, educators and concerned lay persons, and established through the consensus approach.

Objectives redevelopment. After the initial assessment of a learning area, one of the first steps in preparing for reassessment is a review of the learning-area objectives. This is carried out by scholars in the field, educators and concerned lay persons. These reviews may result in revision, modification or total rewriting of the learning-area objectives to reflect current curricular goals and emphases; they may also result in the endorsement of the objectives from the previous assessment as totally adequate for the next assessment.

Open-ended exercise. A non-multiple-choice exercise that requires some type of written or oral response.

Paced audio tape. A tape recording that accompanies each booklet to assure uniformity in administration. Instructions and exercises are read by the announcer onto the tape so that reading difficulties will not interfere with an individual's ability to respond. Response time is included on the tape.

Principal's questionnaire. A data collection form given to school principals. The principals respond to questions concerning enrollments, size of the community, occupational composition of the community, etc. See also *supplementary principal's questionnaire.* Samples of these questionnaires are found in Appendix C.

PSU. A primary sampling unit. This is the basic geographic sampling unit for National Assessment. A PSU is either a single county or a set of contiguous counties.

Public use data tapes. Computer tapes containing respondent-level data for exercises. These tapes are available for use by external researchers wishing to do secondary analyses of National Assessment data.

Racial/ethnic category. Racial categories for which National Assessment reports results are white, black and Hispanic.

Receipt control. Procedures implemented by scoring staff to check in and screen materials from the field. Information gained from receipt control procedures is relayed to assessment administration staff so that any errors may be corrected.

Recycled exercises. The set of exercises that is kept secure from one assessment to the next that will be used to measure changes (growth, stability or decline) in performance for the learning area.

Region. One of four geographical regions used

in gathering and reporting data: Northeast, Southeast, Central and West. States included in each region are shown in Appendix B.

Released exercise. An exercise for which results and exercise text have been reported to the public.

Released exercise set. A set of released exercises, including documentation and scoring guides, made available to state and local education agencies and the research community.

Reporting groups. Categories of variables for which National Assessment data are reported. Variable categories are defined in Appendix B.

Rescore. To insure that responses to open-ended items from both assessments of mathematics were scored in exactly the same manner, all responses to open-ended items from the 1972—73 assessment were scored again at the same time as the 1977—78 responses to these items were scored.

Respondent. A person who responds to the exercises in an assessment booklet.

Response options. Different alternatives to a multiple-choice question that can be selected by the respondent.

Review conference. A conference held to review the objectives of a learning area to assure their acceptance as measures of the objectives by scholars, educators and lay persons or to review exercises for racial, ethnic, social or regional bias.

Sample. National Assessment does not assess an entire age population but rather selects a representative sample from the age group to answer assessment items. (See Chapter 4 for a description of National Assessment sampling procedures.)

Scoring guide. A guide for hand scoring an open-ended exercise that specifies descriptive or diagnostic categories by giving definitions and example responses. Categories are usually defined as "acceptable" or "unacceptable."

Size of community. Results are reported for four size-of-community categories: big cities, fringes around big cities, medium cities and small places. These categories are defined in Appendix B.

SMSA. Standard Metropolitan Statistical Area. SMSAs are economic and social units defined by the U.S. Bureau of the Census.

Stem. The portion of an exercise that states the problem or asks the question.

Stimulus. See *stem.*

Subject area. See *learning area.*

Subpopulation or subgroup. Groups within the national population for which data are reported.

Supplementary principal's questionnaire. A data collection form given to principals. On this form, principals are asked to respond to questions concerning course offerings, materials and staffing specific to the learning area being assessed. See also *principal's questionnaire.* A sample of this questionnaire is found in Appendix C.

Tapescript. A script prepared for the announcer to use in producing the paced tape. It indicates exactly what is to be read or not read and indicates the amount of response time allowed for each exercise. See *paced audio tape.*

Type of community. Results are reported for three type-of-community categories: disadvantaged urban, advantaged urban and extreme rural. Definitions of these categories are found in Appendix B.

Variable. A classification of respondents. Standard reporting variables are: region,

sex, race, level of parental education, size of community, type of community and grade in school.

Weight. A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for assessment with adjustment for nonresponse — an estimate of the number of persons in the population represented by a respondent in the sample. Theoretically, the sum of weights for all respondents at an age level is equal to the number of persons in the country at that age level.

BIBLIOGRAPHY

Bunker, John P. et al., eds. *The National Halothane Study.* Washington, D.C.: U.S. Government Printing Office, 1969.

Changes in Mathematical Achievement, 1973—78, Report 09-MA-01, 1972—73 and 1977—78 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1979. ERIC no. ED 177 011. ISBN 0-89398-134-6.

Data Appendix for the 1977—78 Mathematics Released Exercise Set, no. 09-MA-26, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-139-7. (Available on microfiche only.)

Final Report on National Assessment of Educational Progress Sampling and Weighting Activities for Assessment Year 09. Research Triangle Park, N.C.: Research Triangle Institute, 1979.

Final Report on Year 09 Field Operations and Data Collection Activities, National Assessment of Educational Progress. Research Triangle Park, N.C.: Research Triangle Institute, 1978.

Final Report on Year 09 In-School Field Operations and Data Collection Activities, National Assessment of Educational Progress. Research Triangle Park, N.C.: Research Triangle Institute, 1978.

Kalsbeek, W.D. et al. *No Show Analysis,* final report. Raleigh, N.C.: Research Triangle Institute, 1975.

Mathematical Applications, Report 09-MA-03, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1979. ERIC no. ED 176 965. ISBN 0-89398-136-2.

Mathematical Knowledge and Skills, Report 09-MA-02, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1979. ERIC no. ED 176 964. ISBN 0-89398-135-4.

Mathematical Understanding, Report 09-MA-04, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1979. ISBN 0-89398-137-0.

Mathematics Objectives, Second Assessment, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1978. ERIC no. ED 156 439. ISBN 0-89398-133-8.

Mathematics Technical Report: Summary Volume, Report 04-MA-21, 1972—73 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1976. ERIC no. ED 129 636. ISBN 0-89398-130-3.

Mathematics Technical Report: Summary Volume, Report 09-MA-21, 1977—78 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States (forthcoming). ISBN 0-89398-142-7.

Miller, R.G. Jr. "A Trustworthy Jackknife," *Annals of Mathematical Statistics*, no. 35, 1964.

Miller, R.G. Jr. "Jackknifing Variances," *Annals of Mathematical Statistics*, no. 39, 1968.

Moore, R. et al. *The National Assessment Approach to Sampling*. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1974. ERIC no. ED 099 416.

Mosteller, F. and J.W. Tukey. "Data Analysis Including Statistics," in *Handbook of Social Psychology*. 2nd ed., edited by Aronson, E. and G. Lindzey. Reading, Mass.: Addison-Wesley, 1968.

Rogers, W.T. et al. "Assessment of Nonresponse Bias in Sample Surveys: An Example From National Assessment," *Journal of Educational Measurement*, vol. 14, no. 4, 1977.

The Second Assessment of Mathematics, 1977—78, Released Exercise Set, no. 09-MA-25. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1979. ISBN 0-89398-138-9.

Tukey, John W. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley, 1977.

Tukey, John W. "Some Considerations on Locators Apt for Some Squeezed-Tail (and Stretched-Tail) Parents" (paper prepared in connection with research at Princeton University, supported by the Army Research Office), summer 1975.

☆U.S. GOVERNMENT PRINTING OFFICE 1980—677-719/298

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Education Commission of the States

William G. Milliken, Governor of Michigan, Chairperson, Education Commission of the States
Warren G. Hill, Executive Director, Education Commission of the States
Roy H. Forbes, Director, National Assessment

All National Assessment reports and publications are available through NAEP offices at the address shown at the bottom. Some of the more recent results reports are also available at the Superintendent of Documents (SOD), usually at lower prices. To order from the SOD, write to Supt. of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Check must accompany order. Allow four to eight weeks for delivery.

Reports ordered from National Assessment should be delivered within 12 days. Reports related to this report and available from National Assessment include:

MATHEMATICS

1st Assessment (1972-73)

04-MA-01	*Math Fundamentals*, January 1975	$ 2.45
04-MA-02	*Consumer Math*, June 1975	2.00
04-MA-00	*The First National Assessment of Mathematics: An Overview*, October 1975	1.80
04-MA-20	*Mathematics Technical Report: Exercise Volume*, February 1977	25.00
04-MA-21	*Mathematics Technical Report: Summary Volume*, September 1976	19.50

2nd Assessment (1977-78)

09-MA-01	*Changes in Mathematical Achievement: 1973–78*, August 1979	1.75
09-MA-02	*Mathematical Knowledge and Skills*, August 1979	4.65
09-MA-03	*Mathematical Applications*, August 1979	4.50
09-MA-04	*Mathematical Understanding*, December 1979	3.85
09-MA-21	*Mathematics Technical Report: Summary Volume*, April 1980	12.10
09-MA-40	*Procedural Handbook: 1977–78 Mathematics Assessment*, April 1980	3.70
—	*The Second Assessment of Mathematics, 1977–78: Released Exercise Set*, May 1979	21.00

BACKGROUND REPORTS

BR-2	*Hispanic Student Achievement in Five Learning Areas: 1971–75.* Data for 9-, 13- and 17-year-olds in reading, mathematics, science, social studies and career and occupational development, May 1977	4.45
03/04-GIY	*General Information Yearbook.* A condensed description of the Assessment's methodology, December 1974	2.50

In addition to the above reports, National Assessment has produced reports in the areas of social studies, citizenship, writing, literature, reading, music, art, science and career and occupational development. A complete publications list and ordering information are available from the address below.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295