DOCUMENT RESUME

ED 309 991 SE 050 833

TITLE	On Reporting Student Achievement at the State Level by the National Assessment of Educational Progress. Recommendations.
INSTITUTION	Council of Chief State School Officers, Washington, D.C.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Mar 88
GRANT	NSF-SPA-1549
NOTE	44p.; Report prepared by the State Education Assessment Center's National Assessment Planning Project. For a related document, see SE 050 832.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Elementary School Mathematics; Elementary Secondary Education; *Mathematics Achievement; Mathematics Education; *Mathematics Skills; *Mathematics Tests; National Surveys; Secondary School Mathematics; State Programs; *State Surveys; *Test Construction
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

Interest in state level assessment compared with national assessment is increasing in many states. This report makes 12 recommendations to the U.S. Department of Education and to the National Assessment of Educational Progress (NAEP) grantee on how state achievement data from 1990 mathematics assessments should be measured and reported by NAEP, how comparisons should be reported, and what important variables should be analyzed. Several policy issues related to conducting state-level assessments are addressed, including special education students, limited English proficient students, administrators, and private school students. Many worksheets and state information are provided. Appendices included are: (1) "Steering Committee Members"; (2) "Analysis and Reports Committee Members"; (3) "State Education Agency Consultants"; (4) "Other Consultants"; and (5) "Current NAEP Procedures for Excluding Students." (YP)

SE

ON REPORTING STUDENT ACHIEVEMENT AT THE STATE LEVEL BY THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

RECOMMENDATIONS FROM THE
NATIONAL ASSESSMENT PLANNING PROJECT
COUNCIL OF CHIEF STATE SCHOOL OFFICERS
379 HALL OF THE STATES
400 NORTH CAPITOL STREET, N.W.
WASHINGTON, D.C. 20001

MARCH, 1988

The Council of Chief State School Officers (CCSSO) is a nationwide non-profit organization of the 56 public officials who head departments of public education in every state, U.S. Territory, and the District of Columbia. CCSSO seeks its members' consensus on major education issues and expresses their views to civic and professional organizations, to federal agencies, to Congress, and to the public. Through its structure of standing and special committees, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents the chief education administrator in each state and territory, it has access to the educational and governmental establishments in each state, and the national influence that accompanies this unique position. CCSSO forms coalitions with many other educational organizations, and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

The State Education Assessment Center was founded by CCSSO in 1985 to provide a locus for leadership by the states to improve the monitoring and assessment of education. This is a report of the Assessment Center's National Assessment Planning Project.

Council of Chief State School Officers

Gordon M. Ambach, Executive Director

Verne S. Duncan (Oregon), President

Richard A. Boyd (Mississippi), Chair
Committee on Coordinating Educational Information and Research

---

Ramsay Selden, Director
State Education Assessment Center

Wilmer S. Cody, Director
Marlene C. Holayter, Project Associate
National Assessment Planning Project

Council of Chief State School Officers
379 Hall of the States
400 North Capitol Street, N.W.
Washington, D.C. 20001
(202) 624-7700

The National Assessment Planning Project conducted under the auspices of the Council of Chief State School Officers (CCSSO) was created by a consortium of eighteen national organizations interested in education and in exploring the feasibility of expanding the National Assessment of Educational Progress (NAEP) in order to produce state-by-state comparisons of student achievement.

The project is governed by a Steering Committee (Appendix A), each member appointed on the recommendation of an organization in the consortium. This report was conveyed to the Department of Education (ED) and to the National Assessment of Educational Progress on the review and approval of the Steering Committee. The publication, however, does not necessarily reflect the views of each of the associations in the consortium.

The project was supported by Grant No. SPA-1549 from the National Science Foundation (NSF) with funds partly from NSF and partly from the Department of Education through an inter-governmental transfer from ED to NSF. The mutual interest of the two agencies in this project and their willingness to provide joint support made the project possible. This publication, however, does not necessarily reflect the views of either agency.

TABLE OF CONTENTS

I. INTRODUCTION

PURPOSE/HISTORICAL PERSPECTIVE

For almost twenty years, student achievement in the United States has been measured by the National Assessment of Educational Progress (NAEP) and reported in recent years as "The Nation's Report Card." Based on a sample size of 25,000 students at each of three grade levels, the test has provided the only data that reflect in a comprehensive way what all students in the U.S. know and can do in various subject areas. It has provided descriptive information about student strengths and weaknesses in basic and higher order skills; it has provided data comparing groups of students by race and ethnicity, gender, type of community and region; it has provided trend data reflecting the ups and downs of performance over the years; it has reported positive correlations between achievement and some student experience variables. In 1990, a large number of states will be able to participate in a state-level assessment that will provide "state report cards" which make possible state-by-state or state-to-nation comparisons of student achievement.

Interest in state level comparisons is rising in many quarters in the belief that better state-level information and better state-comparative information about student achievement than is currently available would facilitate the improvement of education in the individual states. In 1984 a majority of CCSSO members supported the development of a system of student assessment that would provide state comparisons, and in 1985, the members endorsed the expansion of NAEP as the most feasible way of providing these comparisons. During the 1986 national assessment, two individual states, Wyoming and Georgia, contracted with NAEP to conduct an in-state concurrent assessment and to provide state-to-nation comparisons. Also in 1986, and 1987, groups of southern states, in a project coordinated by the Southern Region Education Board, contracted with NAEP to conduct state level assessments. They were provided reports comparing achievement among those states. Some of the rising interest can be attributed to governors. A 1987 report from the National Governors' Association entitled Results in Education presented a number of comparative education indicators and displayed a blank column for achievement, clearly expressing the intent to fill in that column in future years with achievement data. In a 1987 report, a national group appointed by Secretary of Education William Bennett and chaired by former Governor of Tennessee, Lamar Alexander, made a series of recommendations on the future of NAEP. A major recommendation was that NAEP should be expanded to provide state-by-state comparisons.

This rising interest is not without its critics. Some are worried that Federal, state and local policymakers may misuse the data, making inappropriate inferences and drawing unwarranted cause and effect conclusions. Fears are expressed that the test will be very influential, and with that influence, foster a national curriculum. Still others fear that the compromises that might be made on objectives will result in an assessment that measures the least common denominator and discourages badly needed curriculum reform.

Designing a national assessment that would not only be constructive but also minimize potential disadvantages is the purpose of this National Assessment Planning Project. The project is to make recommendations that answer two questions of major interest to state and local educators and policymakers, who at some time prior to 1990, will be asked whether they want a report card for their state.

They will likely want to know what mathematics objectives (knowledge, skills) the assessment will measure and whether the objectives are more or less compatible

with what they believe the schools in their states are trying to teach or believe should be taught. That is the topic of a separate publication on Mathematics Objectives for the 1990 Math Assessment. That report describes the basis for the development of test item specifications and for items on the 1990 math assessment.

They will also want to know how the achievement data on students in a state will be presented and how state-to-state comparative data will be reported. However popular with some, a simple number showing only that the average achievement of students in one state is higher or lower than that of students in another state is hardly enlightening. This report makes recommendations to the U.S. Department of Education and to the NAEP grantee on how state achievement data from the 1990 math assessment should be measured and reported by NAEP, how comparisons should be reported, and on several policy issues related to conducting state-level assessments.

PROCESS FOR DEVELOPING RECOMMENDATIONS

The process which resulted in this report was designed principally to result in a broad, field-based consensus on the issues of expanding NAEP to provide state assessments. In part, the approach is reflected in the consensus process developed over the years by NAEP to select the objectives or knowledge and skills that an assessment would measure. That consensus process is responsive to the Federal law creating the National Assessment which requires that "each learning area assessment shall have goal statements devised through a national consensus approval ...."

Since the recommendations in this report address analysis and reporting matters and several assessment policy issues, a national consensus process is not technically required. The law vests authority for setting policies and approving reports with NAEP's Assessment Policy Committee. Nevertheless, an early determination was made to engage many policymakers, practitioners and scholars from across the country in an extensive process of consultation and collaboration. This determination was based on the belief that such a process was likely to identify significant problems with designing "state report cards" as well as constructive ideas. While this is a report drafted under the direction of a working committee and reviewed, modified and approved by the project's Steering Committee, many others were involved, particularly the directors of assessment programs in each of the states. Many of these directors made recommendations early in the process; they also responded to drafts with helpful suggestions.

The process, carried out between August 1987 and March 1988, has the following features:

A Steering Committee (Appendix A) with members recommended by each of the 18 national organizations representing policymakers, practitioners and citizens met, modified and approved the overall plan of work.

An Analysis and Reports Committee (Appendix B) was created to draft this report. Its membership consisted of policymakers, state and local district assessment specialists and scholars in the field of assessment. The Committee met once for preliminary planning to consider its charge and determine what information it wished to have prior to its major work-session in December. For that major work-session, the Committee was given many documents, including the following:

Copies of correspondence from state directors of assessment raising issues and suggesting ways that comparative data should be reported.

An options paper entitled "Alternative Ways of Reporting State-by-State Comparisons" prepared for the Committee under a contract.

Several recent publications on education indicators from CCSSO, the RAND Corporation, and others; lists of common student background questions from a variety of studies, including the National Education Longitudinal Study of 1988, Schools and Staffing Survey, Common Core of Data, High School and Beyond; and the publication The Underachieving Curriculum (International Study of Math Achievement).

A background memorandum from the Southern Regional Education Board reviewing the procedures followed in eight southern states, in which school district employees administered NAEP tests.

A report provided by the National Association of State Directors of Special Education on the inclusion of special students in NAEP.

During its December meeting, the Committee made tentative decisions about the recommendations it wished to make. In subsequent weeks, a report was drafted based on those decisions and copies were sent to state education agencies (Appendix C), and selected state policymakers, local district educators and scholars (Appendix D). Project staff discussed this report with the assessment subcommittee of the Committee on Evaluation and Information Systems (CEIS) of the Council of Chief State School Officers in a special meeting in January.

All comments received on the draft were provided to the Analysis and Reports Committee which met again in late February. The Committee reviewed the comments and completed its report.

In mid-March, the Analysis and Reports Committee's report was reviewed, modified and adopted by the Steering Committee.

It is important to acknowledge that this report contains general recommendations and for any of them to be implemented, further refinement of detail will be needed. In some instances, there are alternative ways of accomplishing the recommendation which must be explored and evaluated for technical, practical and economic feasibility. The development process, therefore, will need to continue. As NAEP continues the process of designing the form and content of the reports it will present to each state which participates in the 1990 mathematics assessment, the interest of all parties and the very success of the process itself will be best served by continuing the involvement of state education policymakers, practitioners and scholars.

GUIDING PRINCIPLES

During the process described above, a series of guiding principles emerged that influenced the recommendations made in this report.

At its initial meeting in March, the Project's Steering Committee adopted a policy statement on the purpose of state comparisons and the conditions that should be met as follows:

The purpose of State Level Student Achievement Comparison is to provide data on student performance to assist policymakers and educators to work toward the improvement of education. Such data can be useful by encouraging and

contributing to a discussion of the quality of education and the conditions that determine it.

State comparative achievement data can be useful if they:

- Represent performance on a consensus of what is important to learn;
- Use sound testing and psychometric practices;
- Use procedures that minimize intrusion into instructional time;
- Take into account different circumstances and needs that the states face; and
- Are associated with features of the school systems that can be improved by policymakers and educators.

In pursuing its task, the Analysis and Reports Committee followed the position of the Steering Committee, and during its discussions several major principles became criteria for developing options and making choices about reporting state achievement data.

- State comparative assessment can help citizens and state and local leaders identify and select educational improvement objectives if the data are sufficiently descriptive of what students know and do not know.
- Far more is needed than just an average score of student performance.
- State-level trend data are important as state-by-state comparisons. They provide encouragement to reform efforts that seem to be making a difference. They can indicate whether more effort is needed or "course corrections" are warranted.
- In addition to achievement data, assessments need to provide information on what educational variables seem to make a difference that might be affected by policies, programs and practice.
- Since comparative reports will be judgmental, they need to acknowledge that states differ in the economic and demographic settings in which schools operate and in the social and economic background of the students entering the schools. These are variables that over a short period of time may account for some part of the different outcomes among the states, and over a long period of time, may determine the magnitude of the task to achieve equal results.

II. ON REPORTING STUDENT ACHIEVEMENT DATA FOR EACH STATE

One of the basic questions that data must answer is, "How well do students in each state perform on the assessment?" The most frequently used reporting rubric, particularly that used by the popular media, is to report the mean or average numerical score of the population tested. An example is annual reporting of the average SAT or ACT scores of students in each state who took those tests. One of the ways NAEP has reported proficiency in a subject is by presenting averages on a scale of 0 - 500. To know that the average score of eighth grade students in a state is 270 on such a scale, provides, by itself, very little information to the public, policy-makers or practitioners. With two or more assessments, trends can be shown by reporting averages, but such reporting provides no information about how the performance of students is distributed or what are student strengths or weaknesses in various topics or areas of the subject assessed. Therefore, such assessments provide little guidance for those trying to decide where to focus instructional improvement efforts.

REPORTING PROFICIENCY LEVELS

Reporting the distribution of student scores in a state between quartiles is more informative than average scores and can reveal where changes occur if reported as trend data. Even more meaningful, however, is the distribution of scores against a standard or set of standards. Of the various ways that state achievement data might be reported using a readily defined standard, NAEP's recent development and application of proficiency scales seems to be an informative method that can also be easily understood.

Recommendation #1

State achievement data will be reported by scales similar to NAEP's proficiency scales, i.e., the percent of the students in each grade who score in or above each of the defined proficiency levels and descriptions of the performance of average students nationally at grades four, eight and twelve will be provided.

The definitions for the five levels used in analyzing and reporting the results of the 1986 NAEP mathematics assessment are as follows:

Level 150 - Simple Arithmetic Facts

> Learners at this level know some basic addition and subtraction facts and can add two-digit numbers without regrouping. They recognize simple situations in which addition and subtraction apply. They also are developing rudimentary classification skills.

Level 200 - Beginning Skills and Understanding

> Learners at this level have considerable understanding of two-digit numbers. They can add two-digit numbers, but are still developing an ability to regroup in subtraction. They know relations among coins, can read information from charts and graphs, and use simple measurement instruments. They are developing some reasoning skills.

Level 250 - Basic Operations and Problem Solving

Learners at this level have an initial understanding of the four basic operations. They are able to add and subtract whole numbers and apply these skills to one-step word problems and money situations. In multiplication, they can find the product of a two-digit and a one-digit number. They can also compare information from graphs and charts and are developing an ability to analyze simple logical relations.

Level 300 - Moderately Complex Procedures and Reasoning

Learners at this level are developing an understanding of number systems. They can compute with decimals, simple fractions, and commonly encountered percents. They can identify geometric figures, measure lengths and angles, and calculate areas of rectangles. These students are also able to interpret simple inequalities, evaluate formulas, and solve simple linear equations. They can find averages, make decisions on information drawn from graphs, and use logical reasoning to solve problems. They are developing the skills to operate with sign numbers, exponents, and square roots.

Level 350 - Multi-step Problem Solving and Algebra

Learners at this level can apply a range of reasoning skills to solve multi-step problems. They can solve routine problems involving fractions and percents, recognize properties of basic geometric figures, and work with exponents and square roots. They can solve a variety of two-step problems using variables, identify equivalent algebraic expressions, and solve linear equations and inequalities. They are developing an understanding of functions and coordinate systems.

This is a developmental scale; the same scale is used by NAEP for reporting the proficiency of fourth, eighth and twelfth graders. This has the advantage of revealing the progress that students attain between grades. The scale was empirically derived; test items in the 1986 assessment were identified that were good discriminators between proficiency levels. In "anchoring" the scale, items were selected whereby students at a given level would have at least 65-80 percent probability of success with those questions while students at a lower level would be much less successful.

What would also be desirable is some reference to a standard of what is considered appropriate for students at each grade level to achieve. Of necessity, such a standard would be a level (or band) on the scale accompanied by descriptions of what mathematics objectives are appropriate to be learned at each grade. Setting standards has traditionally been a state and local function. It is doubtful that a consensus could be achieved among the states on some grade level standards so that they could be applied in reporting 1990 mathematics achievement; consensus might not be possible at all. It is appropriate, however, for NAEP to explore the feasibility of the states establishing such standards through consensus.

In the meantime, a better understanding of what the average fourth, eighth and twelfth grader knows and can do in mathematics would be well served by NAEP including a lengthy description or list of the average students' proficiencies, descriptions that are far more detailed than the scale descriptions presented above. Such descriptions would give more meaning to an average grade-level number or band. Such grade-level descriptions of what the average student in the nation does know would also help each state and district deliberate the goals and standards it may want to establish.

REPORTING BY CONTENT AREA

A second basic question that reports should answer is, "What are student strengths and weaknesses in the various areas of a subject?" NAEP reports have addressed the question in the past. The NAEP Reading Report Card reported trends from 1971 - 1984 in reading. Selected questions were categorized into three reading skill areas: literal comprehension, inferential comprehension and reference skills. Changes were reported in the average percent correct in each category for each grade level between 1980 and 1984. The reports revealed which reading skills were improving and which were declining. This potentially useful information can be used by states to decide whether some changes in curriculum emphases are desirable.

Recommendation #2

State achievement data be reported by various mathematics ability and content areas to reveal the strengths and weaknesses of students in a state in each area.

In a separate report for this project, a matrix for the 1990 math assessment is proposed. It contains three mathematical abilities:

- conceptual understanding
- procedural knowledge
- problem solving

and five content areas:

- numbers and operations
- measurement
- geometry
- data analysis, statistics and probability
- algebra and functions

The state reports on student achievement should include data on how well students in each grade level tested perform in each of those ability and content areas. If the number of items is large enough, then achievement in each mathematical ability area within each content area should be reported.

REPORTING TRENDS

A third basic question that reports must answer is, "How does student achievement change over time?" In 1987, eight states contracted with NAEP to conduct a state math assessment using NAEP items. For those states which may also participate in a 1990 state level assessment, a trend report will be possible. In other states, trend reports will only be possible after two or more successive assessments. It will be in each state's interest to participate in successive assessments so that NAEP can be a source of information about whether improvement is occurring.

Recommendation #3

Cross-sectional trend data, as they are available, be reported to each state by changes in the proportion of students scoring at each proficiency level as well as changes in average achievement scores. In addition, NAEP will report trend data for student achievement in each ability and content area.

One issue concerns the basis for trends. NAEP trend data has been cross-sectional, i.e., performance of 1986 ninth graders was compared to that of 1982 ninth graders. Other studies are based on longitudinal data. Students are assessed at one level, then the same students are assessed several years later. The Alexander-James study report, The Nation's Report Card, recommended that this method be considered. While the research potential in longitudinal studies is considerable, the burden on school systems to keep track of individual students over two to four years and the cost of administering the assessment to those students, wherever they are, seems excessive. Longitudinal studies are valuable, but are best carried out through vehicles other than a state-level national assessment. Furthermore, the focus of NAEP is on the performance of the educational system not individual students. How well the system has done with successive groups of students can be answered with cross-sectional data. NAEP trend data should continue to be cross-sectional.

Another issue is whether trend data should be reported only as change in average achievement or also on changes in percent correct in the content areas and by change in proficiency scale distribution. Changes in the percentages of students in a state on different proficiency scales are far more informative than changes in an over-all average score. There is a difference in whether an increase in over-all average achievement is primarily the result of more students performing in an advanced category, or fewer scoring in a rudimentary category, or due to increases in both. At which level improvements are occurring or not occurring has important implications for directing where to focus improvement efforts. Whereas reporting trends in over-all average achievement may be desirable to reveal the magnitude of a trend, the analytic advantage of reporting proficiency scale trends is also considerable.

Likewise, it can be very valuable to know what trends occur in student achievement in various mathematical process and content areas. Changes in over-all average math scores do not reveal in which content or ability area the change has occurred, information that would be useful in planning where to target improvement efforts.

While data like those described above can be useful, it is important to acknowledge that the size of a state sample of students assessed and the number of items in each ability and content area may limit the level of detail that can be reliably presented. For example, reporting assessment results by ability and topic could call for reporting by a matrix of 5x3 or 15 cells. This is desirable, but NAEP will need to explore whether such a presentation for each state is feasible given the sample size and budget constraints to which it must adjust.

REPORTING BY SUB-GROUPS

A fourth basic question that reports to the states should answer is "What are the achievement levels of population sub-groups, reported by race, ethnicity, gender, type of community or other sub-grouping?"

Recommendation #4

Achievement data will be reported to each state for population sub-group, by race/ethnicity, gender, and type of community in which the student lives as long as the numbers of students in each sub-group in the sample tested are sufficient to provide stable and reliable estimates.

It can be helpful to policy-makers and practitioners to know whether improvement is occurring in all types of communities: urban, suburban, and rural, or only in one or two types of communities. It can be important for leaders to know the magnitude of the levels of achievement of black, white or Hispanic students and to know whether improvement is occurring within each racial or ethnic group in the student population. Monitoring the achievement of boys and girls in conjunction with enrollment in various courses can help state leaders decide whether there is a problem that needs attention. To begin to explore the possibility that problems involving the education of sub-groups of the student population requires that state achievement data be provided on various sub-groups. A note of caution: sub-groups of some populations are very small in some states. For example, Idaho had, as of 1984, a black student population of only 0.4%. Some states may not have enough urban students or enough rural students in the sample to provide valid representation. Whatever sample size in each state turns out to be economically feasible, it is likely that the number of students in some groups in a few states will not be large enough to provide valid results without "over-sampling." Over-sampling should be a decision that a state makes jointly with NAEP. When the size of the sample of a sub-group is too small to provide stable and reliable estimates, then the report should not include achievement data on that sub-group.

Traditionally, the national assessment has reported the race/ethnicity of students in three groups: black, Hispanic and white. The number of other groups of students in the national sample has not been large enough to provide valid generalizations. In some individual states, however, there may be enough students or an interest in over-sampling to provide achievement data on native American Indians, Asian-Pacific students or others. If requested by a state, NAEP should provide reports on groups other than black, Hispanic and white students if the sample size is large enough.

ILLUSTRATIVE TABLES

In following the above four recommendations, the analysis and presentation of state data could take many forms: tables, bar charts, or line graphs. The forms should vary, depending on the audience for the reports. Such decisions should be made by NAEP as it prepares its reports. The main data elements on which students mathematics achievement would be reported for each state should include at least the following:

- Change in achievement over time
- Achievement of all students and of sub-groups identified by race/ethnicity, gender, and type of community
- Achievement displayed as the percent of students scoring at or above each of the levels of the proficiency scales
- Achievement data reported by the mean percent of selected items correct in each mathematics ability and content area

Below are tables that illustrate one way of following the recommendations. The following tables have columns for 1992, as well as 1990 data to illustrate that the measurement of trends is very important.

In presenting graphic displays of data, NAEP should not overlook the importance of making narrative observations of what the analysis of data reveals. NAEP has always made narrative observations in the past and should continue to do so for state-level assessments.

State Mathematics Proficiency For All Students And By Gender

(State) Grade 4

Scale	All Students '90	'92	Chg.	Female '90	'92	Chg.	Male '90	'92	Chg.
150	—	—	—	—	—	—	—	—	—
200	—	—	—	—	—	—	—	—	—
250	—	—	—	—	—	—	—	—	—
300	—	—	—	—	—	—	—	—	—
350	—	—	—	—	—	—	—	—	—
Average	—	—	—	—	—	—	—	—	—

GRADE 8
(The Same)

GRADE 12
(The Same)

State Mathematics Proficiency By Race/Ethnicity

(State) Grade 4

Scale	Black Students '90	'92	Chg.	Hispanic Students '90	'92	Chg.	White Students '90	'92	Chg.
150	—	—	—	—	—	—	—	—	—
200	—	—	—	—	—	—	—	—	—
250	—	—	—	—	—	—	—	—	—
300	—	—	—	—	—	—	—	—	—
350	—	—	—	—	—	—	—	—	—
Average	—	—	—	—	—	—	—	—	—

GRADE 8
(The Same)

GRADE 12
(The Same)

State Mathematics Proficiency By Type of Community

(State) Grade 4

Scale	Rural '90	'92	Chg.	Suburban '90	'92	Chg.	Urban '90	'92	Chg.
150	—	—	—	—	—	—	—	—	—
200	—	—	—	—	—	—	—	—	—
250	—	—	—	—	—	—	—	—	—
300	—	—	—	—	—	—	—	—	—
350	—	—	—	—	—	—	—	—	—
Average	—	—	—	—	—	—	—	—	—

GRADE 8
(The Same)

GRADE 12
(The Same)

State Mean Sub-Scale Score By Type Of Mathematics Question By All Students And By Gender

(State) Grade 4

Ability	All Students			Female			Male		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

Topic	All Students			Female			Male		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Numbers and Operations	—	—	—	—	—	—	—	—	—
Measurement	—	—	—	—	—	—	—	—	—
Geometry	—	—	—	—	—	—	—	—	—
Data Analysis Statistics & Probability	—	—	—	—	—	—	—	—	—
Algebra & Functions	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

GRADE 8
(The Same)

GRADE 12
(The Same)

State Mean Sub-Scale Score By Type Of Mathematics Question By Race/Ethnicity

(State) Grade 4

Ability	Black Students			Hispanic Students			White Students		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

Topic	Black Students			Hispanic Students			White Students		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Numbers and Operations	—	—	—	—	—	—	—	—	—
Measurement	—	—	—	—	—	—	—	—	—
Geometry	—	—	—	—	—	—	—	—	—
Data Analysis Statistics & Probability	—	—	—	—	—	—	—	—	—
Algebra & Functions	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

GRADE 8
(The Same)

GRADE 12
(The Same)

State Mean Sub-Scale Score By Type Of Mathematics Question By Type Of Community

(State) Grade 4

Ability	Rural			Suburban			Urban		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

Topic	Rural			Suburban			Urban		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Numbers and Operations	—	—	—	—	—	—	—	—	—
Measurement	—	—	—	—	—	—	—	—	—
Geometry	—	—	—	—	—	—	—	—	—
Data Analysis Statistics & Probability	—	—	—	—	—	—	—	—	—
Algebra & Functions	—	—	—	—	—	—	—	—	—
Combined (# Items)	—	—	—	—	—	—	—	—	—

GRADE 8

(The Same)

GRADE 12

(The Same)

If feasible, a more desirable alternative to the forgoing tables would be as follows:

(State) Grade 4

	All Students			Female			Male		
	'90	'92	Chg	'90	'92	Chg	'90	'92	Chg
Numbers and Operations									
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Measurement									
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Geometry									
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Data Analysis Statistics and Probability									
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—
Algebra and Functions									
Conceptual Understanding	—	—	—	—	—	—	—	—	—
Procedural Knowledge	—	—	—	—	—	—	—	—	—
Problem Solving	—	—	—	—	—	—	—	—	—

III. ON REPORTING STATE-BY-STATE COMPARISONS

In addition to providing useful information about the proficiency of students in each state, assessments need to report how well students in one state perform in comparison to the performance of students in other states. If the data are detailed enough about levels of proficiency and about each mathematical ability and content area, then comparisons can serve a purpose of helping state leaders to select and target desired improvements. In addition, the comparisons need to acknowledge that states differ in the education needs of the students they serve. During the deliberations of the working committee and in responses from the field to proposals in a draft, interest in and support for three kinds of comparison reports emerged. One would take into account the differences in those demographics in each state that are likely to relate to educating students such as per capita income, educational level of the adult population and percent of students in poverty. A second would take into account the background variables of the students who are assessed. Both of the above would result in assessment reports which acknowledge that states differ in the nature and magnitude of their educational task by characteristics of the general population they serve and the children they teach. In addition, considerable interest was expressed in receiving reports which make it easy for leaders in a state to examine its students' achievement in comparison to that of students in nearby states.

Three forms of state-comparative reports are, therefore, proposed.

RANKING BY STATE DEMOGRAPHIC VARIABLES

Recommendation #5A

The achievement of students in each state will be displayed in a ranking of the states based on a composite of state demographic variables.

The main task is to choose the most appropriate demographic indicators for ranking the states. An in-depth examination of various alternatives was not possible within the time available to prepare this report. NAEP itself, with the advice of technical experts, and in consultation with state representatives, should select a basis for demographic ranking.

In another project of the Council of Chief State School Officers, a series of reliable and readily available set of education related indicators has been identified and presented in a report, Volume I: State Education Indicators 1987. Three of the indicators seem to give the most appropriate reflection of differences in the states that relate to the background of the students in the schools and are used here for the purpose of illustration. For each state, these are: per capita income, the percent of adults who have completed four years of high school and the percent of school age children who live in poverty. The per capita income indicator is 1986 data from the U.S. Bureau of Economic Analysis. Information on education attainment level and percent of school age population in poverty is from the U.S. Bureau of the Census for 1980. Hopefully, by the time state comparative reports for a 1990 mathematics NAEP are issued, 1990 census data will be available. From one perspective, the 1980 census data are appropriate for use. Many of the children who will be assessed in 1990 were born in 1980. They, as well as those born earlier, have lived a substantial period of their lives during which 1980 census data closely approximated the demographics of their environment. In any event, other measures of income, educational attainment and school-age poverty might be considered. For example, state income per pupil might be more desirable than per capita. Education attainment of adults age 25-55 might be

better than that of all adults. In addition, factors such as percent of high income families and percent of college graduates might be included. The specific types and sources of data need to be further explored.

The state demographic ranking, presented as an illustration, was created through the following procedure which started with the data on the three variables shown in the next table.

The distribution of each indicator was standardized to a mean of zero and a standard deviation of one. Then, the composite index was computed for each state giving equal weighting to each of the three components and adding their values. Because percent poverty is a negative concept, its sign was reversed. States were then ranked by the composite index.

The composite and the standardized scores for each of the three variables is shown on the second table. The third table represents one way the achievement data could be presented. The states are listed by composite rank, the base data on which the composite was derived are displayed and the achievement data presented.

One advantage of this form of presentation is that any state can create its own comparisons by identifying and clustering a group of states with similar demographic characteristics.

STATE DEMOGRAPHIC VARIABLES

STATE	1986 Per-Capita Income	1980 Percent Adults 4 years H.S.	1980 Percent School Age Population in Poverty
Alabama	11,336	56.5	22.7
Alaska	17,796	82.5	11.0
Arizona	13,474	72.4	15.4
Arkansas	11,073	55.5	22.3
California	16,904	73.5	13.8
Colorado	15,234	78.6	10.5
Connecticut	19,600	70.3	10.2
Delaware	15,010	68.6	14.4
Florida	14,646	66.7	17.2
Georgia	13,446	56.4	20.1
Hawaii	14,886	73.8	11.4
Idaho	11,223	73.7	13.1
Illinois	15,586	66.5	13.9
Indiana	13,136	66.4	10.8
Iowa	13,348	71.5	10.6
Kansas	14,650	73.3	10.5
Kentucky	11,238	53.1	20.7
Louisiana	11,193	57.7	22.6
Maine	12,790	68.7	14.8
Maryland	16,864	67.4	11.6
Massachusetts	17,772	72.2	12.1
Michigan	14,775	68.0	12.2
Minnesota	14,994	73.1	9.3
Mississippi	9,716	54.8	29.8
Missouri	13,789	63.5	13.7
Montana	11,803	74.4	12.5
Nebraska	13,742	73.4	11.4
Nevada	15,437	75.5	9.0
New Hampshire	15,911	72.3	8.7
New Jersey	18,626	67.4	13.2
New Mexico	11,422	68.9	21.2
New York	17,111	66.3	17.5
North Carolina	12,438	54.8	17.5
North Dakota	12,472	66.4	13.7
Ohio	13,933	67.0	12.0
Oklahoma	12,283	66.0	14.7
Oregon	13,328	75.6	10.4
Pennsylvania	14,249	64.7	13.0
Rhode Island	14,579	61.1	12.4
South Carolina	11,299	54.0	20.3
South Dakota	11,814	67.9	19.0
Tennessee	12,002	56.2	19.8
Texas	13,478	62.6	18.1
Utah	10,981	80.0	9.6
Vermont	13,348	71.0	12.7
Virginia	15,408	62.4	14.1
Washington	15,009	77.6	10.0
West Virginia	10,576	56.0	17.9
Wisconsin	13,909	69.6	9.5
Wyoming	12,781	77.9	7.4

STATES RANKED ON COMPOSITE OF STANDARDIZED DEMOGRAPHIC VARIABLES

STATE	Composite	Percent 4 Yrs. H.S.	Per Capita Income	Percent Above Pov.
Alaska	4.4762	2.2433	1.6094	0.6094
Nevada	3.6135	1.1177	0.8722	1.6235
Connecticut	3.5502	0.2779	2.2433	1.0290
New Hampshire	3.3697	0.4932	1.0290	1.8475
Wyoming	3.2658	1.4600	-0.4376	2.2433
Colorado	3.1948	1.6235	0.7344	0.8368
Washington	3.0554	1.3283	0.6094	1.1177
Minnesota	2.6197	0.6094	0.5503	1.4600
Massachusetts	2.2278	0.4376	1.4600	0.3301
Hawaii	1.9626	0.9477	0.4932	0.5218
Kansas	1.8907	0.6706	0.3833	0.8368
Oregon	1.8861	1.2163	-0.2779	0.9477
California	1.8422	0.8014	1.2163	-0.1755
New Jersey	1.7226	-0.1000	1.8475	-0.0249
Wisconsin	1.6797	0.2264	0.1250	1.3283
Maryland	1.4553	-0.1000	1.1177	0.4376
Utah	1.4402	1.8475	-1.6235	1.2163
Nebraska	1.2811	0.7344	0.0249	0.5218
Iowa	0.9168	0.3833	-0.2009	0.7344
Michigan	0.7404	0.0249	0.4376	0.2779
Illinois	0.4434	-0.2779	0.9477	-0.2264
Delaware	0.4153	0.0749	0.6706	-0.3301
Montana	0.4031	1.0290	-0.8014	0.1755
Ohio	0.3833	-0.1755	0.1755	0.3833
Vermont	0.2542	0.3301	-0.2009	0.1250
New York	0.2508	-0.4376	1.3283	-0.6400
Indiana	-0.0163	-0.3567	-0.3301	0.6705
Arizona	-0.0177	0.5503	-0.0749	-0.4932
Virginia	-0.2109	-0.7344	0.8014	-0.2779
Pennsylvania	-0.2490	-0.5503	0.2264	0.0749
Rhode Island	-0.2971	-0.8014	0.2779	0.2264
Idaho	-0.3191	0.8722	-1.2163	0.0249
Florida	-0.4466	-0.2264	0.3301	-0.5503
Missouri	-0.6344	-0.6094	0.0749	-0.1000
Maine	-0.6959	0.1250	-0.3833	-0.4376
North Dakota	-0.9499	-0.3567	-0.4932	-0.1000
Oklahoma	-1.2858	-0.4932	-0.6094	-0.3833
Texas	-1.4970	-0.6706	-0.0249	-0.8014
South Dakota	-1.6316	-0.0249	-0.7344	-0.8722
New Mexico	-2.0250	0.1755	-0.8722	-1.3283
Georgia	-2.1830	-1.0290	-0.1250	-1.0290
North Carolina	-2.7321	-1.5418	-0.5503	-0.6400
Tennessee	-2.7360	-1.1177	-0.6706	-0.9744
Louisiana	-3.2824	-0.8722	-1.3282	-1.6235
Alabama	-3.7429	-0.9477	-0.9477	-1.8475
West Virginia	-3.7982	-1.2163	-1.8475	-0.7344
South Carolina	-3.9942	-1.8475	-1.0290	-1.1177

(Cont'd)

STATE	Composite	Percent 4 Yrs. H.S.	Per Capita Income	Percent Above Pov.
Arkansas	-4.2484	-1.3283	-1.4600	-1.4600
Kentucky	-4.5773	-2.2433	1.1177	-1.2163
Mississippi	-6.0284	-1.5418	-2.2433	-2.2433

Based on illustration.
Prepared by Program Evaluation and Research Division,
California State Department of Education.

STATES RANKED ON COMPOSITE INDEX OF DEMOGRAPHIC VARIABLES

STATES	1986 Per Capita Income	1980 % Adults 4 yrs H.S.	1980 % School Age Pop. in Poverty	Achievement Grade 4 Proficiency Scales					
				150	200	250	300	350	Aver.
Alaska	17,796	82.5	11.0	—	—	—	—	—	—
Nevada	15,437	75.5	9.0	—	—	—	—	—	—
Connecticut	19,600	70.3	10.2	—	—	—	—	—	—
New Hampshire	15,911	72.3	8.7	—	—	—	—	—	—
Wyoming	12,781	77.9	7.4	—	—	—	—	—	—
Colorado	15,234	78.6	10.5	—	—	—	—	—	—
Washington	15,009	77.6	10.0	—	—	—	—	—	—
Minnesota	14,994	73.1	9.3	—	—	—	—	—	—
Massachusetts	17,772	72.2	12.1	—	—	—	—	—	—
Hawaii	14,886	73.8	11.4	—	—	—	—	—	—
Kansas	14,650	73.3	10.5	—	—	—	—	—	—
Oregon	13,328	75.6	10.4	—	—	—	—	—	—
California	16,904	73.5	13.8	—	—	—	—	—	—
New Jersey	18,626	67.4	13.2	—	—	—	—	—	—
Wisconsin	13,909	69.6	9.5	—	—	—	—	—	—
Maryland	16,864	67.4	11.6	—	—	—	—	—	—
Utah	10,981	80.0	9.6	—	—	—	—	—	—
Nebraska	13,742	73.4	11.4	—	—	—	—	—	—
Iowa	13,348	71.5	10.6	—	—	—	—	—	—
Michigan	14,775	68.0	12.2	—	—	—	—	—	—
Illinois	15,586	66.5	13.9	—	—	—	—	—	—
Delaware	15,010	68.6	14.4	—	—	—	—	—	—
Montana	11,803	74.4	12.5	—	—	—	—	—	—
Ohio	13,933	67.0	12.0	—	—	—	—	—	—
Vermont	13,348	71.0	12.7	—	—	—	—	—	—
New York	17,111	66.3	17.5	—	—	—	—	—	—
Indiana	13,136	66.4	10.8	—	—	—	—	—	—
Arizona	13,474	72.4	15.4	—	—	—	—	—	—
Virginia	15,408	62.4	14.1	—	—	—	—	—	—
Pennsylvania	14,249	64.7	13.0	—	—	—	—	—	—
Rhode Island	14,579	61.1	12.4	—	—	—	—	—	—
Idaho	11,223	73.7	13.1	—	—	—	—	—	—
Florida	14,646	66.7	17.2	—	—	—	—	—	—
Missouri	13,789	63.5	13.7	—	—	—	—	—	—
Maine	12,790	68.7	14.8	—	—	—	—	—	—
North Dakota	12,472	66.4	13.7	—	—	—	—	—	—
Oklahoma	12,283	66.0	14.7	—	—	—	—	—	—
Texas	13,478	62.6	18.1	—	—	—	—	—	—
South Dakota	11,814	67.9	19.0	—	—	—	—	—	—
New Mexico	11,422	68.9	21.2	—	—	—	—	—	—
Georgia	13,446	56.4	20.1	—	—	—	—	—	—
North Carolina	12,438	54.8	17.5	—	—	—	—	—	—

(Cont'd)

STATES	1986 Per Capita Income	1980 % Adults 4 yrs H.S.	1980 % School Age Pop. in Poverty	Achievement Grade 4 Proficiency Scales 150	200	250	300	350	Aver.
Tennessee	12,002	56.2	19.8	—	—	—	—	—	—
Louisiana	11,193	57.7	22.6	—	—	—	—	—	—
West Virginia	10,576	56.0	17.9	—	—	—	—	—	—
Alabama	11,336	56.5	22.7	—	—	—	—	—	—
South Carolina	11,299	54.0	20.3	—	—	—	—	—	—
Arkansas	11,073	55.5	22.3	—	—	—	—	—	—
Kentucky	11,238	53.1	20.7	—	—	—	—	—	—
Mississippi	9,716	54.8	29.8	—	—	—	—	—	—
National Average	—	—	—	—	—	—	—	—	—

COMPARING THE ACHIEVEMENT OF SIMILAR STUDENTS

Whereas the above method of comparison is based on differences in the demographic characteristics of each state's general population, another way is to base the comparisons on the social and economic differences in the background of the students who are being assessed. States differ considerably in the percentage of students tested who are advantaged and disadvantaged in terms of their pre-school and out-of-school experiences. All states have substantial numbers of advantaged and disadvantaged students. It is, however, the differences in the percentages of such students in each state that make the nature and magnitude of the state's educational task different. This leads to the second of the two recommendations on the presentation of state comparisons.

Recommendation #5B

State comparisons will be presented on the achievement of groups of students tested who have similar backgrounds so that similar students are compared.

If students can be grouped reliably by similar background characteristics, comparisons of similar students in different states can be made, even though the number or percent of students in a particular grouping will vary from state-to-state.

Obtaining reliable student data for establishing groupings is the challenge. It would seem desirable, for example, to establish at least three comparison groups such as disadvantaged, advantaged and highly advantaged. Historically, NAEP has asked students a series of background questions that reasonably provide for the creation of groupings and the classification of students who are assessed. These questions relate to parent education, numbers of books and periodicals in the home, and other pertinent factors. The answers about parent education from students in the fourth grade are not as reliable as those received from eighth and twelfth graders. The objective, however, is to create reliably consistent clusters of students and not to determine the precisely accurate assignment of individual students. NAEP would need to construct and validate the use of student questions for this purpose.

One alternative is to secure student social and economic background data by asking the students' parents. A telephone survey of the parents of students who are assessed may be a feasible procedure. A telephone survey of a sample of parents is used successfully in a number of states and local districts. It could provide a wealth of educationally relevant opinions and facts as well as information about parent education level, income range, and other SES data.

A second alternative way of creating groupings of students and assigning individuals that may be feasible is to use data from the United States Census that are aggregated by zip code or postal carrier routes. Zip codes or carrier routes would be used to establish a grouping of disadvantaged, advantaged and highly advantaged based on such census data as average parent income and education level as well as occupation. While the data are not about the specific parents of the individual students assessed, they may be as reliable as student self-reports. NAEP should explore the feasibility of creating groupings of similar students with census data by zip code or postal route.

These and other alternatives described above should be evaluated by NAEP and the most feasible one selected.

One way to present state comparison data on similar students would be as follows:

MATHEMATICS PROFICIENCY OF DISADVANTAGED STUDENTS

Grade 4

	Percent of Sample Tested	150	200	250	300	350	Scale Mean
Alabama							
Alaska							
Arizona							
Arkansas							
California							
Colorado							
Connecticut							
Delaware							
Florida							
Georgia							
Hawaii							
Idaho							
Illinois							
Indiana							
Iowa							
Kansas							
Kentucky							
Louisiana							
Maine							
Maryland							
Massachusetts							
Michigan							
Minnesota							
Mississippi							
Missouri							
Montana							
Nebraska							
Nevada							
New Hampshire							
New Jersey							
New Mexico							
New York							
North Carolina							
North Dakota							
Ohio							
Oklahoma							
Oregon							
Pennsylvania							
Rhode Island							
South Carolina							
South Dakota							
Tennessee							
Texas							
Utah							

(Cont'd)

	Percent of Sample Tested	150	200	250	300	350	Mean
Vermont							
Virginia							
Washington							
West Virginia							
Wisconsin							
Wyoming							

The same presentation would be made of data for advantaged and highly advantaged students in all three grades tested.

In addition, achievement data would be presented for each group of similar students in each grade level by each mathematics ability and content area.

COMPARING ACHIEVEMENT IN NEAR-BY STATES

There is a desire expressed by a number of state leaders to be able to conveniently compare the achievement of students in their states with that of students in adjoining or near-by states. Several respondents to an earlier draft of this report recommended that such a presentation be provided. A recent report from the Council of Chief State School Officers entitled Volume I: State Education Indicators 1987 presented regional groupings as a useful way of displaying various kinds of state data including student achievement. The same regional groupings were used in a report for the National Governors' Association, Time For Results: 1987.

Recommendation #5C

The achievement of students will be presented in regional clusters of states.

While NAEP should explore further with state leaders whether these particular groupings best satisfy the interest in regional comparisons, they do seem reasonable to those involved in preparing this report. Achievement data would be presented by proficiency scales and by performance level in mathematics ability and topic areas by regional groupings like the following:

REGIONAL CLUSTERS FOR COMPARING ACHIEVEMENT

SOUTH ATLANTIC

Florida
Georgia
North Carolina
South Carolina
Virginia
West Virginia

WEST SOUTH CENTRAL

Arkansas
Louisiana
Oklahoma
Texas

MOUNTAIN

Arizona
Colorado
Idaho
Montana
Nevada
New Mexico
Utah
Wyoming

PACIFIC

Alaska
California
Hawaii
Oregon
Washington

NEW ENGLAND

Connecticut
Maine
Massachusetts
New Hampshire
Rhode Island
Vermont

MID-ATLANTIC

Delaware
Maryland
New Jersey
New York
Pennsylvania

MIDWEST

Illinois
Indiana
Michigan
Minnesota
Ohio
Wisconsin

WEST NORTH CENTRAL

Iowa
Kansas
Missouri
Nebraska
North Dakota
South Dakota

EAST SOUTH CENTRAL

Alabama
Kentucky
Mississippi
Tennessee

Source: National Governors' Association. Time for Results: 1987 Washington, D.C.: National Governors' Association

IV. ON RELATING ACHIEVEMENT TO EDUCATION VARIABLES

The purpose of a state-level assessment should not be to just measure achievement, it should also be one source of information about what education variables seem to make a difference in the achievement of students, variables that can be influenced by policies, programs and administrative practices. The value of state assessment data in the improvement of education will be greatly enhanced if information about what seems to make a difference is secured, as well as how well students perform.

The major question is, "Which educational context data should be collected and analyzed? It should be limited by how much is reasonable to ask of teachers, principals and others. It must be limited by a concern for the quality of the data. If the only feasible way to secure particular information results in data that are weak, then the data should not be collected. NAEP cannot answer all research questions. Some can be better and more economically answered through other kinds of research and assessment projects. Nevertheless, NAEP can and should be one source of information on what seems to make a difference so that state leaders, in examining NAEP results and information from other sources, can make reasonable decisions about how education might be improved.

During the process of considering what education experience data should be collected and examined, many variables were considered. One consensus that emerged was that the topics should be limited not only for the reasons stated above, but also limited to those that might be influenced, at least in many states, by state policy or programs or practices. In addition, a consensus emerged on the view that the first attempt to collect education experience data at the state level should be a cautious one.

Recommendation #6

Data will be secured, analyzed and reported on relationships between achievement and educational variables that can be affected by educational policies, programs or practices. Examples of such variables are:

- The number of mathematics courses the student has completed in high school. (Grade 12)
- The types of mathematics courses the student has completed or the advanced level courses the student has completed such as Algebra 2, Geometry, Statistics. (Grade 12)
- The amount of time the student spends on mathematics homework. (Grades 4, 8 & 12)
- The actual time (minutes per week) that students receive mathematics instruction. (Grades 4, 8 & 12)
- Whether students have had the opportunity to learn through class instruction specific topics within a process or content area of mathematics. Have certain topics been part of the scheduled and presented instructional program? (Grades 4, 8 & 12)

- Whether teachers of mathematics have a state certificate to teach mathematics. (Grades 4, 8 & 12)
- Years of experience teaching mathematics. (Grades 4, 8 & 12)
- Number of courses or credits teachers of mathematics classes received in college and graduate school in mathematics and in mathematics pedagogy. (Grades 4 & 8)
- Whether teachers of mathematics have college majors or minors in mathematics. (Grades 4, 8 & 12)
- The extent of the mathematics teachers' recent professional development in mathematics or mathematics education (graduate courses, inservice workshops, conferences, meetings, etc.). (Grades 4, 8 & 12)
- Per pupil expenditures.

The specific method of collecting data on the above variables is beyond the purview of this particular report. NAEP has access to technical experts on its staff and elsewhere to design the appropriate questions or methods to secure useful data.

V. ON SELECTED PROCEDURES

THE INCLUSION AND EXCLUSION OF SPECIAL EDUCATION STUDENTS

Over the years, the decision to exclude from testing a special education student who was "drawn" in the sample has been made by local school staff members responsible for coordinating NAEP. The coordinator was guided by definitions and procedures provided by NAEP, and individual reports that explained the basis for the exclusion were provided to NAEP for each child excluded. In the past, this process has worked relatively well for the national assessment. With state comparisons on the horizon, the process warrants a review. Scores are very sensitive to exclusion or inclusion of students with limitations.

To examine the issue of inclusion or exclusion of special education students, a report was provided by the National Association of State Directors of Special Education (NASDE). It contains many constructive recommendations that go beyond the scope of this report and so is being transmitted to NAEP and the Department of Education as a supplemental resource paper. Copies are available from NASDE. It was a principle source of ideas that led to the following recommendations:

Recommendation #7

- All students and all schools will be considered as potentially eligible for inclusion in the NAEP sample to be tested. Initially, include schools that are solely for special students and do not remove a special school from the list unless there is a determination at the school that none of the students should be tested. Include all categories of special students in the lists from which the sample will be drawn.

- NAEP will continue the practice that the decision to include or exclude a special student be made at the local school level, but specify that the decision be made on a team basis consisting of the parent(s), teacher(s) and others working with the student (i.e., the IEP team). If the student's current IEP contains recommendations regarding group standardized testing, then they should be applied making a special meeting unnecessary.

- New criteria will be provided to NAEP site coordinators and IEP teams regarding inclusion/exclusion as follows:

  - when in doubt, include the student
  - include all students placed in general education settings unless the IEP team specifically decides to exclude them
  - do not exclude any students only on the basis of category of handicap

- exclude students where experience and program plans do not include and never have included the test content, e.g., when the focus has been on life skills rather than on academics

- exclude students who cannot respond to test items even with accommodations.

o The use of the excluded student questionnaire will continue and NAEP will report for each state, the number and percent of students in the population who were excluded by reasons for exclusion or type of handicap.

o Confidential codes will be included on the test forms that identify students who take the test and do have handicaps (as determined by an IEP). Report for each state the number and percent of special students who were tested, by type of handicap

o NAEP should seek the resources that would be required and develop standard "accommodated" tests that do not invalidate test results, e.g., large print, braille, signed directions, mechanical aides, etc. so that students who, with accommodations, can take the test are included in the assessment.

THE EXCLUSION OF LIMITED ENGLISH PROFICIENT STUDENTS

In the national assessment, the decision of whether to exclude a student who has a limited proficiency in English has been handled by the same procedure used to make the decision about a special education student. (See Appendix E for a summary of the process.) The local school coordinator, with the advice of other staff members, may exclude "Non-English speaking students - those who do not read or speak English and would be unable to overcome the language barrier in the test situation."

For each student excluded, an individual report explaining the decision is provided to NAEP by the school.

While appropriate for national achievement data, the criteria for exclusion is far less appropriate when state comparisons are to be made. The achievement results in some states will be significantly affected because they have a high number of new students from foreign countries for whom English is a second language, but who would not be excluded because they do have some proficiency and can respond to some NAEP exercises. They are limited in English, but not incapacitated.

Distinguishing the effects of limited English proficiency (LEP) from mathematics ability in a reliably consistent manner is very difficult. Securing consistent application of some criterion across the United States is even more difficult. In addition, some states have laws that prohibit the use of tests in English with LEP students. A more objective and consistent criterion would be the amount of time a student for whom English is a second language has spent in a school in the United States. Most states and districts expect that students who enter school in the U.S. and who have little or no proficiency in English will have overcome all or most of their limitations after two or three years. It is common for states to allow the exclusion of an LEP foreign-born student from a state testing program for a limited period of time. On

the other hand, some students from other countries overcome the English limitation more rapidly than others. The following is therefore recommended:

Recommendation #8

NAEP criteria for the exclusion of non-English speaking students will be replaced by a guideline for the exclusion of Limited English Proficient students as follows: Any student who entered school in the U.S. with limited English proficiency as determined by the school will be excluded from the NAEP test if the student has been in U.S. schools for less than three full school years, unless in the judgement of the school staff, the student's English proficiency is equal to that of his or her proficiency in the native language. All students with three or more full school years in U.S. schools are to be tested.

WHO SHOULD ADMINISTER THE TESTS TO STUDENTS?

In the past, employees of a NAEP sub-contractor have administered the tests to students. This minimized the data collection burden on local school staff. That practice is partly the basis for widespread confidence in the conditions under which NAEP data have been collected, but it is an expensive part of the process. In 1986 and 1987, a number of states carried out state assessments that permitted comparisons to the national NAEP. School district employees (not teachers) administered the tests. Opinions differ, but there is no clear evidence that one method is more or less reliable than the other or whether differences in the methods bring different results.

Research is needed on the possible effects of different procedures. Only the results of such a study can reasonably address this issue. It would be tragic to discover after-the-fact that a change to local staff administration adversely affects the results of state-to-state or state-to-nation comparisons. Until that question is answered, the credibility of state-to-state comparison will be at risk if past practice is changed.

Recommendation #9

The administration of state-level assessments in 1990 will be carried out by employees of NAEP or its contractor and federal funds should be budgeted accordingly. A project should be conducted by NAEP whereby a small but sufficient number of tests are administered by local district employees to determine whether who administers the tests makes a difference. In addition, NAEP, in cooperation with interested states, shall develop procedures whereby local district employees can, in a reliable way, administer the tests being used for state comparisons or being used to collect information at the school or district level.

REPORTING THE ACHIEVEMENT OF PRIVATE SCHOOL STUDENTS

The national assessment has always been an assessment of all students in the United States, private school as well as public. It seems reasonable to continue that practice.

In most states, the percent of private school students is too small to provide a reliable measure without over-sampling, so the comparison of public and private school students within a state is not feasible in any but a few states.

Who is responsible for the effectiveness of education in a state? For the most part, it is those elected and appointed government officials and educators who are responsible for the public schools in their states. In most states, such officials have little jurisdiction over private school education. It is appropriate, therefore, to make the following recommendation:

Recommendation #10

NAEP will continue to include private as well as public school students in its national sample, but a state sample and the subsequent reports about state-level achievement will be of public school students only.

VI. CONCLUSION/FOLLOW-UP

Clearly, there is much yet to be done to adequately plan for conducting state-level assessment in 1990 and for reporting the results of that assessment. Some of the recommendations in this report can be implemented in a variety of ways. To implement the recommendations in the other major report of this project that defines the mathematics knowledge and skills the 1990 assessment is to measure, item specifications and exercises need to be written and reviewed. These observations lead to the final recommendation in this report.

Recommendation #11

As it pursues the development of the assessment instrument to be used in the 1990 math assessment, the way in which the assessment will be administered, and the ways in which state level assessment results will be reported to the states, NAEP will continue the consensus process started by this project and widely engage state policymakers, practitioners and scholars in its deliberations and planning.

Recommendation #12

NAEP will continue its practice of soliciting comments from the field on drafts of its reports before putting them in final form and will include a representative from the state education agency of each state that participates in a state-level assessment in that field review. In addition, the results of the assessment for each participating state will be reported to the chief state school officer of the state in which the data were collected well in advance of any other release of the state's data.

Planning for state-to-state comparison of student achievement is a complex process. There are many issues that will need to be addressed and resolved before such assessments are conducted, far more issues than this report addresses. The resolution of those complex issues should involve education and public policy leaders across the United States. State level NAEP's will be voluntary and the decision to participate will be made by the appropriate decision-makers in each state, one state at a time. Their acceptance of final plans is crucial. It is also true, however, that there is a strong belief among the hundreds of people who have participated, and are willing to continue to participate in this process, that such widespread involvement will result in decisions that will make state comparisons of student achievement a useful tool for improving the education of young people in America.

APPENDIX A

NATIONAL ASSESSMENT PLANNING PROJECT

STEERING COMMITTEE

American Association of School Administrators
James E. Morrell - Superintendent of Public Schools, Muhlenberg, Pennsylvania

American Federation of Teachers
Antonia Cortese - First-Vice-President, New York State United Teachers, New York

Association of State Assessment Programs
Thomas Fisher - State Assessment Administrator, Florida

Association for Supervision and Curriculum Development
Alice Houston - Assistant Superintendent, Seattle Public Schools, Washington

Council for American Private Education and National Association of Independent Schools
Glenn Bracht - Director, American Lutheran Church, Minnesota

Council of Chief State School Officers
Richard A. Boyd - Superintendent of State Department of Education, Mississippi

Council of the Great City Schools
Lillian Barna - Superintendent, Albuquerque, New Mexico

Directors of Research and Evaluation
Glenn Ligon - Director, Department Management/Information Service, Austin Independent School District, Texas

National Association of Elementary School Principals
C. June Knight - Principal, Hobart Middle School, Oklahoma

National Association of Secondary School Principals
Stephen Lee - Principal, Southwood High School, Indiana

National Association of State Boards of Education
Barbara Roberts Mason - President, State Board of Education, Michigan

National Association of Test Directors
Paul LeMahieu - Director, Office of Research, Testing and Evaluation, Pittsburgh Public Schools, Pennsylvania

APPENDIX A (Cont'd)

National Community on Catholic Education Association
Mary Brian Costello - Superintendent, Archdiocese of Chicago, Illinois

National Council of State Legislators
Wilhelmina Delco - Congresswoman, Texas House of Representatives, Texas

National Education Association
Robert Astrup - President, Minnesota Education Association, Minnesota

National School Board Association
William M. Soult - Director, St. Urain Valley Board of Education, Colorado

National Governors' Association
Nancy DiLaura - Education Assistant, State House, Indiana

APPENDIX B

ANALYSIS AND REPORTS COMMITTEE

Bandy, Irene - Asst. Superintendent, Columbus, Ohio

Burstein, Leigh - Professor, UCLA, Los Angeles, California

Cohen, Michael - Assoc. Director, National Governors' Association, Washington, D.C.

Ford, Valeria - Director, Student Assessment, District of Columbia Schools, Washington, D.C.

Forgione, Jr., Pascal D. - Chief, Office of Research and Evaluation, State Department of Education, Hartford, Connecticut

Hess, Anne C. - Coordinator of Student Assessment, Department of Education, Montgomery, Alabama

Hoover, H. D. - Professor/Director In Basic Skills Program, University of Iowa, Iowa City, Iowa

LeMahieu, Paul - Director, Office of Research, Testing and Evaluation, Pittsburgh, Pennsylvania

Marsden, Margaret - Member, Virginia Board of Education, Arlington, Virginia

Sandifer, Paul - Director, Office of Research, State Department of Education, Columbia, South Carolina

APPENDIX C

STATE EDUCATION AGENCY CONSULTANTS

Bernknopf, Stan - Director, State Assessment Program, Georgia

Badger, Elizabeth - Office of Planning, Research & Evaluation, Massachusetts

Blust, Ross S. - Acting Chief, Division of Education Testing & Evaluation, Pennsylvania

Brown, William J. Jr. - Assistant State Superintendent, Research and Testing, North Carolina

Chin-Chance, Selvin - Administrator, Test Development Section, Hawaii

Clausen, Thomas G. - Superintendent of Education, Louisiana

Cooperman, Saul - Commissioner of Education, New Jersey

Crowe, Kevin - Director of Planning, Research & Evaluation, Nevada

Danzig, Arnold - Director, Education Information Center, Arizona

Davis, S. John - Superintendent of Public Instruction, Virginia

Donahue, Dennis - Assistant Superintendent, Wyoming

Earhart, J. Troy - Commissioner of Education, Rhode Island

Donovan, David - Assistant Superintendent for Technical Assistance and Evaluation, Michigan

Fjelstad, Donna - Assistant to the State Superintendent, Division of Education, South Dakota

Gabrys, Robert E. - Chief, Program Assessment, Evaluation, and Instructional Support Branch, Maryland

Grover, Herbert J. - State Superintendent of Public Instruction, Wisconsin

Harrison, Ann E. - Director of Program Planning, Research, & Evaluation Section, Kansas

Hartman, Allan - Office of Planning, Research & Evaluation, Massachusetts

Hawks, Gary - Interim Superintendent of Public Instruction, Michigan

Hudson, Roger - Administrator, Curriculum Services, Nebraska

Jackson, Roberta - Minimum Competency Specialist, Kansas

Keene, William B. - Superintendent, Delaware

APPENDIX C (Cont'd)

Keister, Sally - Coordinator, Guidance/Assessment & Evaluation, Idaho

Kerins, Tom - Manager, Student Assessment Section, Illinois

Law, Alex - Director, Division of Program Evaluation & Research, California

Lepley, William - Director, Department of Education, Iowa

Lockwood, Robert E. - Coordinator, Office of Research, Alabama

Maxcy, Horace P. - Associate Director, Division of Educational Assessment, Maine

Moak, Lynn M. - Deputy Commissioner for Research and Information, Texas

Randall, William T. - Commissioner of Education, Colorado

Randall, Ruth E. - Commissioner of Education, Minnesota

Roeber, Edward - Supervisor, Educational Assessment Program, Michigan

Sandifer, Paul D. - Director, Office of Research, South Carolina

Sansted, Wayne G. - Superintendent of Public Instruction, North Dakota

Saterfiel, Thomas H. - Deputy State Superintendent, Mississippi

Sobol, Thomas - Commissioner of Education, New York

Snodgrass, H. M. - Deputy Associate Superintendent, Kentucky

Strange, William B - Senior Officer, Center of School Assessment, Indiana

Trent, E. Roger - Director, Division of Educational Services, Ohio

Van Fleet, Donald S. - Director, Division of Public Schools, Florida

APPENDIX D

OTHER CONSULTANTS

Apt, Denise C. - Chairman, House Committee on Education, Kansas House of Representatives

Barrett, Ben F. - Associate Director, Kansas Legislative Research Department

Bell, Clarence E. - State Senator, Arkansas State Senate

Faatz, Jeanne - State Representative, Colorado House of Representatives

Fuhrman, Susan - Professor, Rutgers University

Godshall, Deb - Senior Analyst, Colorado Legislative Council

Guthrie, James W. - Professor of Education/PACE Director, University of California at Berkely

Hoff, Philip H. - State Senator, Chair-Senate Education Committee, Vermont State Senate

Houston, Paul - Superintendent, Tucson Schools

Katz, Vera - Speaker of the House, Oregon House of Representatives

Keppel, Francis - Senior Lecturer on Education, Harvard University

Linn, Robert L. - Co-Director, UCLA Center for Research on Evaluations Standards and Students Testing/Professor, University of Colorado

Malone, David - State Senator, Arizona State Senate

Musick, Mark D. - Vice-President, Southern Regional Education Board

Oakes, Jeannie - Social Scientist, The RAND Corporation

Ollie, Art - State Representative, Chair House Education Committee, Iowa House of Representatives

Payzant, Thomas W. - Superintendent, San Diego Schools

Perry, Marcia D. - Assistant Director, Georgia Assessment Project, Georgia State University

Powell, Lee Etta - Superintendent, Cincinnati Schools

Silver, Edward, A. - Professor of Mathematics Education, University of Pittsburgh

Snyder, Cooper - Senate Chairman, Education Committee, Ohio State Senate

Whiteneck, William D. - Chief Consultant, Senate Education Committee, California State Senate

APPENDIX E

CURRENT NAEP PROCEDURES FOR EXCLUDING STUDENTS

Sampling. NAEP currently uses a stratified, three-stage national probability sample design. In the first stage of sampling, the United States is divided into geographical units comprised of counties or groups of contiguous counties which meet a minimum school enrollment size. These units are stratified by both region and size of community. In the second stage of sampling, all public, parochial, private, Bureau of Indian Affairs, and Department of Defense schools are listed according to three grade/age groups within each primary sampling unit. They are selected with a probability proportional to assigned measures of size. Schools that are designated as being only for students with handicaps are eliminated from the sample. In the third stage of sampling, a consolidated list of all grade and age eligible students is established for each selected school. This list includes all students regardless of English proficiency or handicap.

Exclusions. After a list of all students who are grade or age eligible is prepared, the directions from NAEP to local school coordinators of the assessment indicate that:

> the school coordinator should review the list of eligible students and, in concert with other school officials, determine whether any student should be excluded from assessment because they are non-English speaking, educable mentally retarded, or functionally disabled...These categories are defined as follows:
>
> - Non-English Speaking Students - Those who do not read or speak English and would be unable to overcome the language barrier in the test situation.
>
> - Educable Mentally Retarded (EMR) - Students who have been psychologically tested as EMR students or students who are considered EMR in the professional opinion of the principal or other qualified staff members. However, students should not be excluded because of poor academic performance or normal discipline problems. Only those students should be excluded who could not give meaningful responses to exercises at their age level.
>
> - Functionally Disabled - (temporary or permanent physical disability) - Students who are so disabled that they cannot perform in the NAEP testing situation, should also be excluded. However, functionally disabled students who can respond should be included.