## NABP 1996 MATHIDNATICS

## Report Card for the Nation and the States



THE NATION'S
U.S. DEPARTMENT OF EDUCATION OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Findings from the Notional Assessment of Educotional Progress

## What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

## The National Assessment Governing Board

Honorable William T. Randall, Chair
Former Commissioner of Education
State of Colorado
Denver, Colorado
Mary R. Blanton, Vice Chair
Attorney
Salisbury, North Carolina
Patsy Cavazos
Principal
W.G. Love Accelerated Elementary School

Houston, Texas
Catherine A. Davidson
Secondary Education Director
Central Kitsap School District
Silverdale, Washington
Edward Donley
Former Chairman
Air Products \& Chemicals, Inc.
Allentown, Pennsylvania
Honorable James Edgar
Member Designate
Governor of Illinois
Springfield, Illinois
James E. Ellingson
Fourth-Grade Classroom Teacher
Probstfield Elementary School
Moorhead, Minnesota
Thomas H. Fisher
Director, Student Assessment Services Florida Department of Education
Tallahassee, Florida
Michael J. Guerra
Executive Director
Secondary Schools Department
National Catholic Educational Association Washington, DC
Edward H. Haertel
Professor of Education
Stanford University
Stanford, California

Jan B. Loveless
District Communications Specialist
Midland Public Schools
Midland, Michigan
Marilyn McConachie
Former School Board Member
Glenbrook High Schools
Glenview, Illinois
William J. Moloney
Superintendent of Schools Calvert County Public Schools Prince Frederick, Maryland
Honorable Annette Morgan
Member
Missouri House of Representatives
Jefferson City, Missouri
Mark D. Musick
President
Southern Regional Education Board Atlanta, Georgia

Mitsugi Nakashima
First Vice-Chairperson
Hawaii State Board of Education
Honolulu, Hawaii
Michael T. Nettles
Professor of Education \& Public Policy
University of Michigan
Ann Arbor, Michigan
and Director
Frederick D. Patterson Research Institute
United Negro College Fund
Honorable Norma Paulus
Superintendent of Public Instruction
Oregon State Department of Education Salem, Oregon
Honorable Roy Romer
Governor of Colorado
Denver, Colorado

## Honorable Edgar D. Ross

Judge
Territorial Court of the Virgin Islands
Christiansted, St. Croix
U.S. Virgin Islands

Fannie L. Simmons
Mathematics Coordinator
District 5 of Lexington/Richland County Ballentine, South Carolina

## Adam Urbanski

President
Rochester Teachers Association
Rochester, New York
Deborah Voltz
Assistant Professor
Department of Special Education
University of Louisville
Louisville, Kentucky
Marilyn A. Whirry
Twelfth-Grade English Teacher
Mira Costa High School
Manhattan Beach, California
Dennie Palmer Wolf
Senior Research Associate
Harvard Graduate School of Education
Cambridge, Massachusetts
Marshall S. Smith (Ex-Officio)
Acting Assistant Secretary
Office of Educational Research and Improvement
U.S. Department of Education

Washington, DC

## Roy Truby

Executive Director, NAGB
Washington, DC

# NAEP 1996 Mathematics Report Card for the Nation and the States 

Findings from the<br>National Assessment of Educational Progress

Clyde M. Reese Karen E. Miller<br>John Mazzeo<br>John A. Dossey

## February 1997

Office of Educational Research and Improvement
U.S. Department of Education

Prepared by Educational Testing Service under contract with the National Center for Education Statistics

# U.S. Department of Education 

Richard W. Riley
Secretary

Office of Educational Research and Improvement<br>Marshall S. Smith<br>Acting Assistant Secretary<br>National Center for Education Statistics<br>Pascal D. Forgione, Jr.<br>Commissioner<br>\section*{Education Assessment Group}<br>Gary W. Phillips<br>Associate Commissioner

February 1997

## SUGGESTED CITATION

Reese, C.M., Miller, K.E., Mazzeo, J., and Dossey, J.A.
NAEP 1996 Mathematics Report Card for the Nation and the States, Washington, DC: National Center for Education Statistics, 1997.

## FOR MORE INFORMATION

Contact:
Arnold A. Goldstein
202-219-1741

For ordering information on this report, write:

```
National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641
```

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).
This report also is available on the World Wide Web: http://www.ed.gov/NCES/naep.
The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Educational Testing Service is an equal opportunity, affirmative action employer.
Educational Testing Service, ETS, and the ETS logo are registered trademarks of Educational Testing Service.

## Table of Contents

Executive Summary ..... i
Chapter 1. NAEP 1996 Mathematics Assessment ..... 1
NAEP's Mission ..... 1
NAEP 1996 Mathematics Framework ..... 2
Student Samples ..... 5
Reporting NAEP Results ..... 7
NAEP Mathematics Scale ..... 7
NAEP Mathematics Achievement Levels ..... 7
Item Maps ..... 9
Sample Questions from the NAEP 1996 Assessment in Mathematics ..... 13
Overview of the Remaining Report ..... 23
Cautions in Interpretations ..... 24
Chapter 2. Mathematics Scale Score Results:
National and State Trends and Comparisons ..... 25
National and State Results ..... 25
Performance of Selected Subgroups ..... 31
Gender ..... 31
Race/Ethnicity ..... 32
Parents' Highest Level of Education ..... 34
Type of School ..... 36
Participation in Title I Programs ..... 38
Eligibility for the Free/Reduced-Price Lunch Program ..... 39
Chapter 3. Achievement Level Results ..... 41
Achievement Level Descriptions ..... 41
National and State Results ..... 46
Performance of Selected Subgroups ..... 52
Gender ..... 52
Race/Ethnicity ..... 54
Parents' Highest Level of Education ..... 56
Type of School ..... 58
Participation in Title I Programs ..... 60
Eligibility for the Free/Reduced-Price Lunch Program ..... 61
Chapter 4. Exploring a More Inclusive NAEP ..... 63
The Three NAEP 1996 Mathematics Samples ..... 66
National and State Percentages of Students with Disabilities and LEP Students ..... 67
Effect of Inclusion Criteria and Provision of Accommodation on the Participation Rates ..... 68
Concluding Comments ..... 74

## TABLES

Table 1.1 Percentage Distribution of Items by Content Strand and Grade ..... 4
Table 2.1 Average Mathematics Scale Scores by Region ..... 26
Table 2.2 Average Mathematics Scale Scores - Grade 4 Public Schools ..... 28
Table 2.3 Average Mathematics Scale Scores - Grade 8 Public Schools ..... 30
Table 2.4 Average Mathematics Scale Scores by Gender ..... 31
Table 2.5 Average Mathematics Scale Scores by Race/Ethnicity ..... 33
Table 2.6 Average Mathematics Scale Scores by Parents' Highest Education Level ..... 35
Table 2.7 Average Mathematics Scale Scores by Type of School ..... 37
Table 2.8 Average Mathematics Scale Scores by Title I Participation ..... 38
Table 2.9 Average Mathematics Scale Scores by Free/Reduced Price Lunch Program ..... 39
Table 3.1 Percentage Attaining Mathematics Achievement Levels by Region ..... 47
Table 3.2 Percentage Attaining Mathematics Achievement Levels Grade 4 Public Schools ..... 49
Table 3.3 Percentage Attaining Mathematics Achievement Levels Grade 8 Public Schools ..... 51
Table 3.4 Percentage Attaining Mathematics Achievement Levels by Gender ..... 53
Table 3.5 Percentage Attaining Mathematics Achievement Levels by Race/Ethnicity ..... 55
Table 3.6 Percentage Attaining Mathematics Achievement Levels by Parents' Highest Education Level ..... 56
Table 3.7 Percentage Attaining Mathematics Achievement Levels by Type of School ..... 59
Table 3.8 Percentage Attaining Mathematics Achievement Levels by Title I Participation ..... 60
Table 3.9 Percentage Attaining Mathematics Achievement Levels by Free/Reduced-Price Lunch Program Eligibility ..... 61
Table 4.1 Percentage of National Population Identified as SD, LEP, or Both Grades 4, 8, and 12, Public and Nonpublic Schools ..... 67
Table 4.2 Percentage of National Population Excluded From the Assessment, Grades 4, 8, and 12, Public and Nonpublic Schools ..... 70
Table 4.3 Percentage of Students with Disabilities and Limited English Proficient Students in the National Population Included in the Assessment, Grades 4, 8, and 12, Public Schools Only ..... 72

## FIGURES

Figure 1.1 Mathematical Framework for the 1996 Assessment ..... 3
Figure 1.2 Participating Jurisdictions in the NAEP 1996 State Assessment Program in Mathematics ..... 6
Figure 1.3 Policy Definitions of NAEP Achievement Levels ..... 8
Figure 1.4 Map of Selected Questions on the NAEP Mathematics Scale for Grade 4 ..... 10
Figure 1.5 Map of Selected Questions on the NAEP Mathematics Scale for Grade 8 ..... 11
Figure 1.6 Map of Selected Questions on the NAEP Mathematics Scale for Grade 12 ..... 12
Figure 1.7 NAEP 1996 Mathematics Sample Questions for Grade 4 ..... 14
Figure 1.8 NAEP 1996 Mathematics Sample Questions for Grade 8 ..... 16
Figure 1.9 NAEP 1996 Mathematics Sample Questions for Grade 12 ..... 20
Figure 2.1 Average Mathematics Scale Scores ..... 25
Figure 3.1 NAEP Mathematics Achievement Levels - Grade 4 ..... 43
Figure 3.2 NAEP Mathematics Achievement Levels - Grade 8 ..... 44
Figure 3.3 NAEP Mathematics Achievement Levels - Grade 12 ..... 45
APPENDICES
Appendix A Overview of Procedures Used for the NAEP 1996 Mathematics Assessment ..... 75
Appendix B 1996 State-Level Results for Selected Subgroups ..... 109
Appendix C State-Level Contextual Variables ..... 137
Appendix D State-Level SD/LEP Information ..... 141
Appendix E Discussion of the Grade 8 Asian/Pacific Islander Sample ..... 147
Appendix F Standard Errors ..... 151
Acknowledgments

## Dxecutive Summary ${ }^{1}$

For more than a quarter of a century, the National Assessment of Educational Progress (NAEP) has reported to policy makers, educators, and the general public on the educational achievement of students in the United States. As the nation's only ongoing survey of students' educational progress, NAEP has become an important resource for obtaining information on what students know and can do.

The NAEP 1996 mathematics assessment continues the mandate to evaluate and report the educational progress of students at grades 4,8 , and 12 . The national results provided herein describe students' mathematics achievement at each grade and within various subgroups of the general population. State-level results for grades 4 and 8 are presented for individual states and jurisdictions that chose to participate in the 1996 state assessment. In addition, trends in performance since 1990 are reported for the nation and for states and jurisdictions that participated in the 1990, 1992, and 1996 assessments. NAEP national and state data assess the performance of students in both public and nonpublic schools.

## The NAEP 1996 Mathematies Frameworls

The NAEP 1996 mathematics assessment, like previous mathematics assessments in 1992 and 1990, uses a framework influenced by the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics. The 1996 framework was updated to more adequately reflect recent curricular emphases and objectives.

The framework measures a mathematics domain containing five mathematics strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions). In addition to the five content strands, the assessment examined mathematical abilities (conceptual understanding, procedural knowledge, and problem solving) and mathematical power (reasoning, connections, and communication). Since 1990, the NAEP mathematics assessments have placed increasing emphasis on mathematical power. The 1996 assessment deliberately focused on reasoning and communication by requiring students to connect their learning across mathematical strands.

## Student Achievement

Students' mathematics performance is summarized on the NAEP mathematics scale, which ranges from 0 to 500 . In addition, results for each grade are reported according to three achievement levels: Basic, Proficient, and Advanced. The National Assessment Governing Board (NAGB) developed and adopted the mathematics achievement levels, based on collective judgments about what students should know and be able to do in mathematics. The Basic level

[^0]denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade. The Proficient level represents solid academic performance, while the Advanced level signifies superior performance. These achievement levels are still developmental, and the process for setting them remains in transition.

## Major Findings for the Nation, Regions, and States ${ }^{2}$

National data from the NAEP 1996 mathematics assessment showed progress in the mathematics performance by students on a broad front, compared with both the 1990 and 1992 assessments.

- Students' scores on the NAEP mathematics scale increased for all three grades. Scores were higher in 1996 than in 1992 for all three grades, and higher in 1992 than in 1990. The national average scale score for fourth graders in 1996 was 224, an increase of 11 points over the national average for 1990; the average for eighth graders in 1996 was 272 , an increase of 9 points; and the average score for twelfth graders was 304 , also an increase of 10 points.
- Student performance also increased as measured by the three mathematics achievement levels set by NAGB. The percentage of students at or above the Basic level increased for all three grades. The percentage of fourth-grade students at or above the Proficient level increased from 1990 to 1992, and from 1992 to 1996, while the percentage of eighth- and twelfth-grade students at or above the Proficient level increased over the period 1990 to 1996. However, only eighth-grade students showed an increase in the percentage at the Advanced level, and this increase was for the period 1990 to 1996.
- For fourth-grade students, the percentage performing at or above the Basic level was 64 percent in 1996, as compared to 50 percent in 1990; for eighth-grade students, 62 percent as compared to 52 percent; and for twelfth-grade students, 69 percent as compared to 58 percent.

Regional results showed positive trends similar to the national results for some but not all regions. NAEP divides the United States into four regions: the Northeast, Southeast, Central, and West.

- The Southeast and Central regions recorded increases in the average NAEP mathematics scale scores over the period 1990 to 1996 for all three grades. The Northeast recorded an increase for fourth graders only, while the West showed an increase for twelfth grades only.
- For fourth-grade students, average scale scores were higher in the Northeast and Central than the Southeast and West; for eighth-grade students, scores in the Central were higher than the West; and for the twelfth-grade students, scores in the Southeast were lower than the other three regions.

[^1]State data for the NAEP 1996 mathematics assessment covered fourth graders in 47 states, territories, and other jurisdictions and eighth graders in 44 states and jurisdictions. Many but not all states and jurisdictions showed increases in mathematics performance for the 1996 assessment.

- Fourth graders in 15 of the 39 states and jurisdictions that participated in both the 1992 and 1996 assessments showed an increase in their average scale scores for 1996.
- Eighth graders in 13 of the 37 states and jurisdictions that participated in both the 1992 and 1996 assessments showed an increase in their average scale scores.
- Eighth graders in 27 of the 32 jurisdictions that participated in both the 1990 and 1996 assessments showed an increase in their average scale scores.
- Colorado, Connecticut, Indiana, North Carolina, Tennessee, Texas, and West Virginia reported increases in the percentages of fourth graders who scored at or above the Basic and Proficient achievement levels over the period 1992 to 1996.
- Maryland, Michigan, Minnesota, Nebraska, North Carolina, and Wisconsin reported increases in the percentages of eighth graders who scored at or above all three achievement levels over the period 1990 to 1996.


## Major Findings for Student Subgroups

The NAEP 1996 mathematics assessment reports national results on the basis of demographic subgroups, type of school attended, participation in Title I programs, and eligibility for the free/reduced-price lunch component of the National School Lunch Program.

- Average scale scores for eighth- and twelfth-grade males and females showed no significant differences in 1996. Scores for fourth-grade males were higher than scores for fourth-grade females.
- White students recorded increases in their average mathematics scale scores for all three grades over the period 1990 to 1996.
- Black and Hispanic students recorded increases in their average mathematics scale scores for grades 4 and 12 over the period 1990 to 1996.
- Scores for Black, Hispanic, and American Indian students remained below scale scores for White students. The gaps between scores for these subgroups did not change in 1996.
- Generally, students with higher scale scores reported higher levels of parental education. The more education parents had, the higher the scores of their children.
- Both public and nonpublic schools showed increased scale scores for fourth- and eighth-grade students. Public schools showed increased scores for twelfth-grade students as well. Students attending nonpublic schools continued to outperform their peers attending public schools.
- Fourth- and eighth-grade students receiving services supplied by Title I programs had lower scale scores than those who did not participate in Title I. (The sample for twelfth graders who participated was not large enough to permit a comparison.) Title I of the Elementary and Secondary Education Act provides funding to local educational agencies to meet the needs of children who are failing or most at risk of failing. For this reason, the difference between the scores cannot be taken as an indication that Title I programs fail to benefit students. The NAEP 1996 mathematics assessment was the first mathematics assessment to collect data on these students.
- Students eligible for the free/reduced-price lunch program administered by the U.S. Department of Agriculture (USDA) scored lower than those not eligible, for all three grades. Eligibility for free/reduced-price lunches is determined by the USDA's Income Eligibility Guidelines. Information on eligibility was lacking for 16 percent of fourth graders, 17 percent of eighth graders, and 27 percent of twelfth graders. The NAEP 1996 mathematics assessment was the first mathematics assessment to collect data on these students.


## Exploring a More Inclusive NAEP

NAEP has always attempted to report results that reflect the achievement of all students at a given grade or age. Logistical difficulties prevent the sampling of certain students, for example, students who receive home schooling, who are in ungraded schools, who attend special schools for the deaf and blind, or who are incarcerated. Some students who are enrolled in regular schools also present special considerations with respect to sampling - those with disabilities and those who are limited English proficient (LEP). NAEP 1996 results indicate that 15 percent of the nation's fourth graders, 11 percent of the eighth graders, and 8 percent of twelfth graders are classified as students with disabilities or as LEP students. Previous NAEP assessments sampled more than half of these students.

The NAEP 1996 assessments investigated the feasibility of increasing the participation of students with disabilities and LEP students. Revised inclusion criteria, in combination with accommodations to remove barriers to participation, were examined to determine their impact of participation rate, for students with disabilities and LEP students.

The analysis of inclusion issues featured in this report is only a first step in an ongoing research and development effort. A comprehensive research report on inclusion issues will be published later in 1997.

## Chapter 1

## NAEP 1996 Mathematics Assessment

## NAEP's Mission

The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what students in the United States know and can do in various academic subjects. NAEP is authorized by Congress and directed by the National Center for Education Statistics (NCES) of the U.S. Department of Education. The National Assessment Governing Board (NAGB), an independent body, provides policy guidance for NAEP.

Since its inception in 1969, NAEP's mission has been to collect, analyze, and produce valid and reliable information about the academic performance of students in the United States in various learning areas. In 1990, the mission of NAEP was expanded to provide state-by-state results on academic achievement. Participation in the state-by-state NAEP is voluntary and has grown from 40 states and territories in 1990 to 48 in 1996.

NAEP has also become a valuable tool in tracking progress toward the National Education Goals. The subjects assessed by NAEP are those highlighted at the 1989 Education Summit and in later legislation. ${ }^{1}$ The NAEP 1996 assessment in mathematics marks the third time the subject has been assessed with the new framework in the 1990s, enabling policy makers and educators to track mathematics achievement since the release of the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics ${ }^{2}$ in 1989.

The following report is the first release of results from the NAEP 1996 assessment in mathematics. National results at grades 4,8 , and 12 and state-by-state results at grades 4 and 8 are presented. The focus of this report, and the mission of NAEP, is to inform policy makers and the public about student achievement.

[^2]
## NAEP 1996 Mathematies Frameworlk

The NAEP assessment measures a mathematics domain containing five mathematics strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions). Questions involving content from one or more of the strands are also categorized according to the domains of mathematical abilities and mathematical power. The first of these, mathematical abilities, describes the nature of the knowledge or processes involved in successfully handling the task presented by the question. It may reflect conceptual understanding, procedural knowledge, or a combination of both in problem solving. The second domain, mathematical power, reflects processes stressed as major goals of the mathematical curriculum. Mathematical power refers to the students' ability to reason, to communicate, and to make connections of concepts and skills across mathematical strands, or from mathematics to other curricular areas. Figure 1.1 summarizes the structure of the NAEP mathematics assessment.

The mathematics framework for the NAEP 1996 assessment is a revision of that used in the 1990 and 1992 assessments. Changes were made to the earlier framework in light of the NCTM Standards and changes taking place in school mathematics programs. The previous NAEP mathematics framework was refined and sharpened so that the 1996 assessment would: (1) more adequately reflect recent curricular emphases and objects and yet (2) maintain a connection with the 1990 and 1992 assessments to measure trends in student performance. Prior to the 1996 assessment, investigations were conducted to ensure that results from the assessment could be reported on the existing NAEP mathematics scale. The conclusion drawn from these investigations was that results from the 1990, 1992, and 1996 assessments could be reported on a common scale and trends in mathematics performance since 1990 examined. Appendix A briefly highlights selected changes in the current NAEP mathematics framework.

The conception of mathematical power as reasoning, connections, and communication has played an increasingly important role in measuring student achievement. In 1990, the NAEP assessment included short constructed-response questions as a way to begin addressing mathematical communication. In 1992, the extended constructed-response questions included on the assessment required students not only to communicate their ideas but also to demonstrate the reasoning they used to solve problems. The 1996 assessment continued to emphasize mathematical power by including constructed-response questions focusing on reasoning and communication and by requiring students to connect their learning across mathematical content strands. These connections were addressed within individual questions reaching across content strands and by families of questions contained within a single content strand.

## Figure 1.1

Mathematical Framework for the 1996 Assessment



SOURCE: National Assessment Governing Board, Mathematics Framework for the 1996 National Assessment of Educational Progress.

In real life, few mathematical situations can be clearly classified as belonging to one content strand or another, and few situations require only one facet of mathematics thinking. Therefore, many of the questions are classified in a number of ways. In addition to being classified by all applicable content strands, each question was classified by its assessment of applicable mathematical abilities (procedural knowledge, conceptual understanding, and problem solving) and mathematical powers (reasoning, communication, and connections). As displayed in Figure 1.1, the content strands, mathematical abilities, and mathematical power combine to form the framework for the NAEP assessment. (A brief description of the five content strands is presented in Appendix A.)

| Table 1.1 | Percentage Distribution of Items by Content Strand and Grade |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Grade 4 |  |  | Grade 8 |  |  | Grade 12 |  |  |
|  | 1990 | 1992 | 1996 | 1990 | 1992 | 1996 | 1990 | 1992 | 1996 |
| Content Area |  |  |  |  |  |  |  |  |  |
| Number Sense, Properties and Operations ${ }^{1}$ | 45 | 45 | 40 | 30 | 30 | 25 | 25 | 25 | 20 |
| Measurement | 20 | 20 | 20 | 15 | 15 | 15 | 15 | 15 | 15 |
| Geometry and Spatial Sense ${ }^{2}$ | 15 | 15 | 15 | 20 | 20 | 20 | 20 | 20 | 20 |
| Data Analysis, Statistics, and Probability | 10 | 10 | 10 | 15 | 15 | 15 | 15 | 15 | 20 |
| Algebra and Functions | 10 | 10 | 15 | 20 | 20 | 25 | 25 | 25 | 25 |

${ }^{1}$ Approximately half the questions in 1996 at each grade level involved some aspect of estimation.
${ }^{2}$ At grade 12 in 1996, approximately 25 percent of the geometry questions involved topics in coordinate geometry.

Table 1.1 shows the percentage of questions prescribed by the frameworks in each of the five content strands by grade level. The 1996 percentages reflect changes from those in 1990 and 1992. Overall, the percentages of questions by content strand reflect the refinement of the NAEP mathematics assessment to conform with recommendations from the NCTM Standards. Questions could be classified under more than one content strand. For example, the number sense, properties, and operations strand may underlie concepts in other strands.

The framework incorporated the use of calculators (four-function at grade 4 and scientific at grades 8 and 12), rulers (at all grades) and protractors (at grades 8 and 12), and manipulatives such as geometric shapes and spinners.

Also, the framework continued the shift from multiple-choice questions to questions that required students to construct responses. In 1996, more than 50 percent of student assessment time was devoted to constructed-response questions. Two types of constructedresponse questions were included - (1) short constructed-response questions that required students to provide answers to computation problems or to describe solutions in one or two sentences, and (2) extended constructed-response questions that required students to provide longer responses when answering the questions.

There were additional types of blocks of questions created for each of the grade levels. Each grade level had two estimation blocks. These blocks employ a paced-audiotape format to measure students' estimation skills and to move students through some word problems. The pacing method curtails time for computations thus requiring students to estimate their answers. Each grade level also had two 30-minute theme blocks consisting of a mixture of multiplechoice and constructed-response questions. All of the questions in these theme blocks related to some aspect of rich problem setting that served as a theme uniting the entire block of questions. The estimation and theme blocks were not included in the mathematics scale presented in this report. A future report on the NAEP 1996 mathematics results will highlight findings from the estimation and theme block components of the assessment.

At each grade level assessed, the NAEP 1996 mathematics assessment consisted of a set of booklets, each containing student background questions and cognitive mathematics questions. The background sections asked students to provide information about themselves, classroom instruction, and motivation to complete the assessment. The cognitive sections included multiple-choice and constructed-response questions designed to assess students' mathematical knowledge and skills. Additional discussion of the content of the assessment and the various student, teacher, and school instruments is presented in Appendix A.

## Student Samples

The NAEP 1996 mathematics assessment was conducted nationally at grades 4,8 , and 12 , and state-by-state at grades 4 and 8 . For both the national and state-by-state assessments, representative samples of public and nonpublic school students were assessed. (For many of the states participating in the 1996 assessment, the sample of nonpublic school students was not adequate for reporting separate nonpublic school results.) Appendix A contains information on sample sizes and participation rates for the national and state-by-state assessments.

Forty-four states, the District of Columbia, Guam, the Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS), and the overseas Department of Defense Dependent Schools (DoDDS) participated in the 1996 state-by-state assessment. (Throughout this report, participants in the state-by-state assessment are referred to as "jurisdictions.") To ensure comparability across jurisdictions, NCES has established guidelines for school and student participation rates. Appendix A highlights these guidelines and jurisdictions failing to meet these guidelines are noted in tables and figures presenting state-by-state results. For jurisdictions failing to meet the initial school participation rate of 70 percent, results are not reported.

Figure 1.2 lists the jurisdictions that participated in the 1996 mathematics assessment and notes those jurisdictions failing to meet one or more NCES-established participation rate guidelines for public schools. (Information on nonpublic school participation rates is presented in Appendix A.)

## Figure 1.2 <br> Participating Juristidions in the NAEP 1996 State Assessment Program in Mathematics

## Fridele 4

| Alabama | Indiana | Nebraska | Texas |
| :--- | :--- | :--- | :--- |
| Alaska ${ }^{2}$ | lowa $^{2}$ | Nevada $^{2}$ | Utah |
| Arizona | Kentucky | New Jersey ${ }^{2}$ | Vermont $^{2}$ |
| Arkansas $^{2}$ | Maisiana | New Mexico | Virginia |
| California | Maryland | New York | Washington |
| Colorado | Massachusetts | North Carolina | West Virginia |
| Connecticut | Michigan | North Dakota | Wisconsin |
| Delaware | Minnesota | Oregon | Wyoming |
| District of Columbia | Mississippi | Pennsylvania ${ }^{2}$ | DDESS |
| Florida | Missouri | Rhode Island | DoDDS |
| Georgia | Montana ${ }^{2}$ | South Carolina ${ }^{2}$ | Guam |
| Hawaii | Tennessee |  |  |

## Grude 8

| Alabama | Indiana | Nebraska | Texas |
| :---: | :---: | :---: | :---: |
| Alaska² | lowa ${ }^{2}$ | Nevada ${ }^{1}$ | Utah |
| Arizona | Kentucky | New Hampshire ${ }^{1}$ | Vermont ${ }^{2}$ |
| Arkansas ${ }^{2}$ | Louisiana | New Jersey ${ }^{1}$ | Virginia |
| California | Maine | New Mexico | Washington |
| Colorado | Maryland ${ }^{2}$ | New York² | West Virginia |
| Connecticut | Massachusetts | North Carolina | Wisconsin ${ }^{2}$ |
| Delaware | Michigan ${ }^{2}$ | North Dakota | Wyoming |
| District of Columbia | Minnesota | Oregon | DDESS |
| Florida | Mississippi | Rhode Island | DoDDS |
| Georgia | Missouri | South Carolina ${ }^{2}$ | Guam |
| Hawaii | Montana ${ }^{2}$ | Tennessee |  |

${ }^{1}$ Failed to meet the initial school participation rate of 70 percent for public schools; public school results not reported.
2 Failed to meet one or more participation rate guidelines for public schools; public school results reported with appropriate notation.
For more details on participation rate guidelines and nonpublic school participation rates, see Appendix A.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Reporting NAEP Results

## NAEP Mathematics Scale

The questions composing the NAEP 1996 assessment span the broad field of mathematics in each of the grades assessed. Because of the survey nature of the assessment and the breadth of the content strands, each student participating in the assessment cannot be expected to answer all the questions. Thus, each student is administered a portion of the assessment, and data across students are combined to report on the achievement of fourth-, eighth-, and twelfth-grade students and on the achievement of subgroups of students (e.g., subgroups defined by gender or parental education).

The NAEP mathematics scale, which ranges from 0 to 500 , is used to report performance across the three grade levels and is a composite of the five content strands measured in the NAEP mathematics assessment. Student responses to assessment questions are analyzed to determine the percentage of students responding correctly to each multiple-choice question and the percentage of students responding to each of the score categories for constructed-response questions. Item response theory (IRT) methods are used to produce content strand scales that summarize results for each of the five mathematics content strands. These content strand scales are linked to their corresponding 1992 mathematics content strand scales using IRT procedures. (Linking refers to the procedures used to make the scales for the reported 1990, 1992, and 1996 results comparable.)

An overall composite scale is developed by weighting the separate content strand scales. The weights correspond to the relative importance of each content strand in the NAEP 1996 mathematics framework. The 1996 columns in Table 1.1 present the relative contribution that each content strand makes to the composite scale score. As displayed, the weighting of each content strand scale changes from grade to grade to reflect the changing emphasis in mathematics curricula as students progress from elementary school through high school. The resulting scale, which is also linked to the mathematics scale used to report 1990 and 1992 results, defines the reporting metric used to present results in Chapter 2. (Details of the scaling procedures will be presented in the forthcoming NAEP 1996 Technical Report and the Technical Report for the NAEP 1996 State Assessment in Mathematics.)

## NAEP Mathematies Achievement Levels

Results for the NAEP 1996 assessment in mathematics are also reported using the mathematics achievement levels that were authorized by the NAEP legislation and adopted by the National Assessment Governing Board. The achievement levels which are based on collective judgments about what students should know and be able to do relative to the body of content reflected in the NAEP mathematics assessment. Three levels were defined for each grade - Basic, Proficient, and Advanced. The levels were defined by a broadly representative panel of teachers, education specialists, and members of the general public.

For reporting purposes, the achievement levels for each grade are placed on the NAEP mathematics scale. Figure 1.3 presents the policy definitions of the achievement levels, while Chapter 3 contains specific descriptions for the levels at each grade.

It should be noted that setting achievement levels is a relatively new process for NAEP, and it is still in transition. Some evaluations have concluded that the percentage of students at certain levels may be underestimated. ${ }^{3}$ On the other hand, critiques of those evaluations have asserted that the weight of the empirical evidence does not support such conclusions. ${ }^{4}$ A further review is currently being conducted by the National Academy of Sciences.

The student achievement levels in this report have been developed carefully and responsibly, and the procedures used have been refined and revised as new technologies have become available. Upon review of the available information, the Commissioner of Education Statistics has judged that the achievement levels are in a developmental status. However, the Commissioner and the Governing Board also believe that the achievement levels are useful and valuable for reporting on the educational achievement of students in the United States. Chapter 3 presents results reported in terms of the mathematics achievement levels.


|  | Basic |
| :---: | :--- |
| Proficient | This level denotes partial mastery of prerequisite knowledge and skills that are <br> fundamental for proficient work at each grade. |
| This level represents solid academic performance for each grade assessed. <br> Students reaching this level have demonstrated competency over challenging <br> subject matter, including subject-matter knowledge, application of such knowledge <br> to real-world situations, and analytical skills appropriate to the subject matter. |  |
| Advanced | This level signifies superior performance. |

[^3]
## Item Maps

To better illustrate the NAEP mathematics scale, questions from the assessment are mapped onto the 0 -to- 500 scale at each grade level. These item maps are visual representations that compare questions with ability, and they indicate which questions a student can likely solve at a given performance level as measured on the NAEP scale. ${ }^{5}$ The mathematic achievement levels are also indicated on each item map.

Figures 1.4 through 1.6 are item maps for grades 4,8 , and 12, respectively. The 0 -to-500 mathematics scale includes all three grades; therefore, the majority of questions administered at grade 4 are targeted to the lower range of the scale, reflecting the typical performance of fourth graders. Similarly, most questions administered at grade 12 are targeted to the higher range of the scale. As a result, most fourth-grade questions map to the lower end of the scale, while most twelfth-grade questions map to the higher end. Questions administered at grade 8 are targeted more to the middle of the scale.

As an example of how to interpret the item maps, consider a multiple-choice question that requires students to identify cylindrical shapes and maps at a scale score of 208 for grade 4 (see Figure 1.4). Mapping a question at a score of 208 implies that students performing at or above this level on the NAEP mathematics scale have a 74 percent or greater chance of correctly answering this particular question. ${ }^{6}$ Students performing at a level lower than 208 would have less than a 74 percent chance of correctly answering the question. This mapping does not mean that students at or above the 208 level always answer the question correctly or that students below the 208 level always answer the question incorrectly. Students have a higher or lower probability of successfully answering the question depending on their overall ability as measured on the NAEP scale.

As another example, consider a constructed-response question that requires students to partition the area of a rectangle and maps at a score of 272 for grade 8 (see Figure 1.5). Scoring of this response allows for partial credit by using a four-point scoring guide. Mapping a question at a score of 272 implies that students performing at or above this level have a 65 percent or greater chance of receiving a score of 3 (Satisfactory) or 4 (Complete) on the question. Students performing at a level lower than 272 would have less than a 65 percent chance of receiving such a score.

[^4]NAEP Scale


Each grade 4 mathematics question was mapped onto the NAEP 0 to 500 mathematics scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probablity of successfully answering the question. (The probability was 74 percent for 4 -option multiple-choice question and 72 percent for 5 -option multiple choice question.) Only selected questions are presented. The fourth-grade mathematics achievement levels are referenced on the map.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),
1996 Mathematics Assessment.

## NAEP Scale



Each grade 8 mathematics question was mapped onto the NAEP 0 to 500 mathematics scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probablity of successfully answering the question. (The probability was 74 percent for 4 -option multiple-choice question and 72 percent for 5 -option multiple choice question.) Only selected questions are presented. The eighth-grade mathematics achievement levels are referenced on the map.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),
1996 Mathematics Assessment.

## NAEP Scale



Each grade 12 mathematics question was mapped onto the NAEP 0 to 500 mathematics scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probablity of successfully answering the question. (The probability was 74 percent for 4 -option multiple-choice question and 72 percent for 5 -option multiple choice question.) Only selected questions are presented. The twelfth-grade mathematics achievement levels are referenced on the map.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Sample Questions firom the NAEP 10.6 Assessment in Mathemntics

The NAEP 1996 assessment in mathematics is a rich collection of questions developed to survey the mathematical knowledge and skills of students in grades 4,8 , and 12. Each student received both multiple-choice and constructed-response questions. As shown in the item maps (see Figures 1.4 through 1.6), multiple-choice and constructed-response questions are used to assess all levels of mathematical knowledge and skills. The sample questions presented below represent the types of questions used (i.e., multiple-choice, short constructed-response, and extended constructed-response) but do not illustrate the breadth of the content assessed.

Figures 1.7 through 1.8 present samples of questions. The tables accompanying the questions show two types of percentages: (1) the overall percentage of students within a grade who successfully answered the question and (2) the percentages of students within each of the achievement level intervals - Basic, Proficient, and Advanced as well as below Basic - who successfully answered the question.

The second question in Figure 1.7 is a short constructed-response question administered at grade 4 . The question asked students to recognize the need for division and that the answer is the number of students remaining when 34 students are divided into groups of eight. One possible way of arriving at the correct answer would require students to divide 34 by 8 and note the remainder of 2 . Another path to the solution would be to note that 34 is 2 more than 8 times 4 . In either case, the number of remaining students, or substitutes, would be 2 .

The third question in Figure 1.8 is an extended constructed-response question administered at grade 8 . The question asked students to carefully note the context of the solution, daily ridership on Metro Rail, and then compare the relative value of two graphs. Students were asked to use both graphical representations of the data, one at a time, to argue for rapid increase in sales from March to October and to argue for little difference in the level of ticket sales between the same two months.

The third question in Figure 1.9 is a short constructed-response question administered at grade 12. The question asked students to compare the strength of two cherry-flavored drinks. As is the case with many of the questions included in the NAEP mathematics assessment, there are a number of ways for students to arrive at the correct answer. One approach might find students comparing $6 / 59$ or 10.16 percent cherry flavor for Luis with $5 / 47$ or 10.63 percent cherry flavor for Martin. The question measures students' ability to compare and order ratios and rates.

N stands for the number of stamps John had. He gave 12 stamps to his sister.
Which expression tells how many stamps John has now?
(A) $N+12$
(B) $N-12$
(C) $12-N$
(D) $12 \times N$

The correct answer is B.
This multiple-choice question measures algebra and functions.

| Grade 4 | PERCENTAGE "CORRECT" WITHIN ACHIEVEMENT LEVEL INTERVALS |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Overall <br> Percentage Correct | Below Basic <br> 213 and below* | Basic <br> 214 to $248 *$ | Proficient <br> 249 to $281 *$ | Advanced <br> 282 and above* |
| 67 | 44 | 73 | 90 | $* * *$ |

*NAEP mathematics composite scale range
***Sample size insufficient to permit reliable estimates

Ms. Hernandez formed teams of 8 students each from the 34 students in her class. She formed as many teams as possible, and the students left over were substitutes. How many students were substitutes?

Answer: $\qquad$

The correct answer is 2 .
This short constructed-response question measures number sense, properties, and operations. Students' responses were scored correct or incorrect.

| Grade 4 4 | PERCENTACE "CORRECT" WITHIN ACHIEVEMENT LEVEL INTERVALS |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Overall | Below Basic <br> 213 and below* | Basic <br> 214 to $248^{*}$ | Proficient <br> 249 to $281 *$ | Advanced <br> Percentage Correct |
| 39 | 5 | 42 | 86 | and above* |

[^5]Sam can purchase his lunch at school. Each day he wants to have juice that costs 50 ¢, a sandwich that costs 90 , and fruit that costs $35 ¢$. His mother has only $\$ 1.00$ bills. What is the least number of $\$ 1.00$ bills that his mother should give him so he will have enough money to buy lunch for five days?

This short constructed-response question measures number sense, properties, and operations. Students' responses were scored using a three-point scoring guide that allowed for partial credit. The following is a sample of a student response that received the highest score, Satisfactory. A Satisfactory response to this question gives the correct answer of nine dollar bills.


Grade 4
PERCENTAGE "SATISFACTORY" WITHIN ACHIEVEMENT LEVEL INTERVALS

| Overall <br> Percentage Satisfactory | Below Basic <br> 213 and below* | Basic <br> 214 to $248^{*}$ | Proficient <br> 249 to $281^{*}$ | Advanced <br> 282 and above* |
| :---: | :---: | :---: | :---: | :---: |
| 17 | 1 | 14 | 44 | $* * *$ |

*NAEP mathematics composite scale range
***Sample size insufficient to permit reliable estimates

## Figure 1.8 <br> NAEP 1996 Mathematics Sample Questions for Grade 8

A car odometer registered $41,256.9$ miles when a highway sign warned of a detour 1,200 feet ahead. What will the odometer read when the car reaches the detour?
(A) $42,456.9$
(B) $41,279.9$
(C) $41,261.3$
(D) $41,259.2$
(E) $41,257.1$

The correct answer is E.
This multiple-choice question measures the measurement strand.

| Grade 8 | PERCENTAGE "CORRECT" WHTHN ACHIEVEMENT LEVEL INTERVALS |  |
| :---: | :---: | :---: | :---: | :---: |

[^6]

In the figure above, what fraction of rectangle $A B C D$ is shaded?
(A) $\frac{1}{6}$
(B) $\frac{1}{5}$
(C) $\frac{1}{4}$
(D) $\frac{1}{3}$
(E) $\frac{1}{2}$

The correct answer is D.
This multiple-choice question measures number sense, properties, and operations.

| Grade 8 | PERCENTAGE "CORRECT" WITHIN ACHIEVEMENT LEVEL INTERVALS |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Overall <br> Percentage Correct | Below Basic 261 and below* | Basic 262 to 298* | Proficient 299 to 332 * | Advanced 333 and above* |
| 65 | 32 | 78 | 96 | $99!$ |

[^7]This question requires you to show your work and explain your reasoning. You may use drawings, words, and numbers in your explanation. Your answer should be clear enough so that another person could read it and understand your thinking. It is important that you show all of your work.

METRO RAIL COMPANY

| Month | Daily Ridership |
| :--- | :---: |
| October | 14,000 |
| November | 14,100 |
| December | 14,100 |
| January | 14,00 |
| February | 14,300 |
| March | 14,600 |

The data in the table above has been correctly represented by both graphs shown below.

Graph A


Graph B


Which graph would be best to help convince others that the Metro Rail Company made a lot more money from ticket sales in March than in October?

Explain your reason for making this selection.
Why might people who thought that there was little difference between October and March ticket sales consider the graph you chose to be misleading?

This extended constructed-response question measures data analysis, statistics, and probability. Students' responses were scored using a four-point scoring guide that allowed for partial credit. Scores of 3 (Satisfactory) and 4 (Complete) are illustrated below.

*NAEP mathematics composite scale range
***Sample size insufficient to permit reliable estimates

The following is a sample of a student response that received a Satisfactory score. A Satisfactory response to this question gives the correct response, Graph B, but provides an incomplete but partially correct explanation.
Graph B because itshous how the
a graph goes up so much.
because it shows a big jump
because all they and was make each square worth more ridechig

The following is a sample of a student response that received the highest score, Complete. A Complete response to this question gives the correct response, Graph B, and provides a complete explanation.

$$
\begin{aligned}
& \text { graphs } \\
& \text { Because it has a smaller scale for } \\
& \text { daily riderskit it lobes take a sweater } \\
& \text { increase } \\
& \text { Because it appears itsuncreased } \\
& \text { a lot when its only uncreadas } 600
\end{aligned}
$$

## Figure 1.9



If triangles $A D E$ and $A B C$ shown in the figure above are similiar, what is the value of $x$ ?
(A) 4
(B) 5
(C) 6
(D) 8
(E) $\quad 10$

The correct answer is A.
This multiple-choice question measures the geometry and spatial sense strand.

| Grade 12 | PERCENTACE "CORRECT" WITHIN ACHIEVEMENT LEVEL INTERVALS |  |
| :---: | :---: | :---: | :---: | :---: |

*NAEP mathematics composite scale range
***Sample size insufficient to permit reliable estimates

Which of the following could be the graph of a function?
(A)

(B)

(C)

(D)

(E)


The correct answer is E.
This multiple-choice question measures algebra and functions.

| Grade 12 | PERCENTAGE "CORRECT" WITHIN ACHIEVEMENT LEVEL INTERVALS |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Overall <br> Percentage Correct | Below Basic 287and below* | Basic $288 \text { to } 335^{*}$ | Proficient 336 to 366* | Advanced 367 and above* |
| 20 | 8 | 17 | 56 | *** |

[^8]Luis mixed 6 ounces of cherry syrup with 53 ounces of water to make a cherry-flavored drink. Martin mixed 5 ounces of the same cherry syrup with 42 ounces of water. Who made the drink with the stronger cherry flavor?

Give mathematical evidence to justify your answer.

This short constructed-response question measures number sense, properties, and operations. Students' responses were scored using a three-point scoring guide that allowed for partial credit. The following is a sample of a student response that received the highest score, Satisfactory. A Satisfactory response to this question identifies Martin and shows correct mathematical justification.


| Grade 12 | PERCENTACE "SATISFACTORY" WITHIN ACHEVEMENT LEVEL INTERVALS |  |
| :---: | :---: | :---: | :---: | :---: |

***Sample size insufficient to permit reliable estimates

## Dverview of the Remaining Report

Chapters 2 and 3 of this report present selected results in terms of the NAEP mathematics scale and achievement levels, respectively. Within each of these chapters, findings are presented for the nation, for regions, and for the major reporting subgroups described below. Appendix A presents more detailed descriptions of the reporting subgroups.

- Gender. Estimates are reported separately for males and females.
- Race/Ethnicity. Estimates are reported for students' race/ethnicity (self-identified), using the following mutually exclusive categories: White, Black, Hispanic, Asian/Pacific Islander, and American Indian (including Alaskan Native).
- Parents' Highest Level of Education. Estimates are reported based on students' reports of the highest level of education attained by at least one of their parents. These levels are: did not finish high school, graduated from high school, some education after high school, or graduated from college.
- Type of School. Estimates are reported for students attending public schools and nonpublic schools, including Catholic and other private schools.
- Title I Participation. Estimates are reported for students who are classified either as currently participating in Title I programs or services or as not participating in such programs or services.
- Free/Reduced-Price Lunch Program Eligibility. Estimates are reported for students classified as either currently eligible for the Department of Agriculture's Free/Reduced-Price Lunch Program or not eligible.

In addition to the national results, state-by-state results are included for the states and jurisdictions that participated in the mathematics assessment at grades 4 and 8.

This report examines and compares the mathematics performance of groups of students defined by demographic characteristics or by responses to background questions (e.g., males compared to females). It does not explore the relationships among combinations of these groups (e.g., White males compared to Black males).

The averages and percentages presented in this report are estimates because they are based on samples rather than on all members of each population. Consequently, the results are subject to a measure of uncertainty, reflected in the standard errors of the estimates. The comparisons presented in this report are based on statistical tests that consider the magnitude of the difference between the group averages or percentages and the standard errors of those statistics. Throughout this report, differences among reporting groups are defined as significant when they are significant from a statistical perspective. This statement means that observed differences in the sample are believed to reflect real differences in the population and are unlikely to result from chance factors. ${ }^{7}$ The term significant, therefore, is not intended to imply a judgment about the absolute magnitude of the educational relevance of the differences. It is intended to identify statistically dependable population differences to help focus subsequent dialogue among policy makers, educators, and the public.

[^9]Chapter 4 of this report discusses an investigation of steps to make NAEP a more inclusive measure of the achievement of all students. The NAEP 1996 assessments in mathematics and science examined revised inclusion criteria and accommodation for students with disabilities and limited English proficient students. Chapter 4 describes the changes to the inclusion criteria used in 1996 and the various accommodations available for the first time. This chapter focuses on describing the samples of students with disabilities and limited English proficient students rather than comparing these students with larger NAEP samples. The results reported in Chapters 2 and 3 are for the "reporting samples" from the 1996 assessment. The "reporting samples" used inclusion criteria equivalent to those used for the 1990 and 1992 assessments to allow for comparability of results across assessments.

This report also contains appendices that support or augment the results presented. Appendix A contains an overview of the NAEP mathematics framework and specifications, information on the national and state samples, and a more detailed description of the major reporting subgroups featured in Chapters 2 and 3. Appendix A also explains the need for reanalyzing the 1990 and 1992 data. (The 1990 [achievement level only] and 1992 [scale scores and achievement level] results presented in the report are based on a re-analysis of the data from these previous assessments.) The next two appendices present state-by-state results from NAEP (Appendix B) and non-NAEP (Appendix C) sources. Appendix D provides supporting material for Chapter 4 and Appendix E presents results for the grade 8 Asian /Pacific Islander sample. Finally, Appendix F presents the standard errors for the averages and percentages presented in the body of the report.

Detailed information about the measurement methodology and data analysis techniques will be available in the forthcoming NAEP 1996 Technical Report.

## Cautions in Interpretations

The reader is cautioned against interpreting the relationships among subgroup averages or percentages as causal relationships. Average performance differences between two groups of students may result, in part, from socioeconomic and other factors. For example, differences among racial/ethnic subgroups are almost certainly associated with a broad range of socioeconmic and educational factors not discussed in this report and possibly not addressed by the NAEP assessment program. Similarly, differences in performance between public and nonpublic school students may be better understood by accounting for other factors such as the composition of the student body, parents' education levels, and parental involvement. Finally, student participation roles and the motivation of students, particularly twelfth graders, to perform on an assessment like NAEP should be considered when interpreting the results. (A further discussion of twelfth graders' participation rates and motivation is presented in Appendix A.)

## Chapter 2

## Mathematics Scale Score Results: National and State Trends and Comparisons

## National and State Results

Overall, the mathematics results for the nation's fourth-, eighth-, and twelfth-grade students show continued improvement from 1990 to 1996 (see Figure 2.1). Performance on the 1992 mathematics assessment, when compared to 1990, showed a five-point increase at grades 8 and 12 and a seven-point increase at grade 4. The latest NAEP assessment in 1996 indicates continued improvement, when compared to 1992, showing a four-point gain in average mathematics scale scores at grades 4 and 8, and a five-point gain at grade 12. Combined, national performance in mathematics has risen 9 to 11 points since 1990.


* Indicates a significant difference from 1990.
† Indicates a significant difference from 1992.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

This encouraging national trend in the 1990s is generally reflected across the regions of the nation. (Although gains were also observed between 1990 and 1992, and between 1992 and 1996, not all were statistically significant.)

As shown in Table 2.1, performance differences among regions of the country still exist. In 1996, students attending schools in the Southeast region scored lower than those in the Northeast and Central regions at grade 4; the Central region at grade 8; and the Northeast, Central, and West regions at grade 12. In addition, at grade 4, students attending schools in West region scored lower than their counter parts in the Northeast and Central regions.

| Table 2.1 | Average Mothemotics Scale Scores by Region |  |  |  |  | THE NATION'S |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grade 4 | 1990 |  | 1992 |  | 1996 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| Nation | 100 | 213 | 100 | 220* | 100 | 224* $\dagger$ |
| Northeast | 22 | 215 | 21 | 224* | 22 | 228* |
| Southeast | 25 | 205 | 24 | 211 | 21 | 218* |
| Central | 25 | 216 | 27 | 224* | 25 | 231* |
| West | 27 | 216 | 28 | 219 | 32 | 220 |
| Grade 8 |  |  |  |  |  |  |
| Nation | 100 | 263 | 100 | 268* | 100 | 272* $\dagger$ |
| Northeast | 20 | 270 | 22 | 270 | 20 | 277 |
| Southeast | 25 | 255 | 25 | 261 | 23 | 266* |
| Central | 24 | 266 | 25 | 275* | 24 | 277* |
| West | 30 | 261 | 28 | 268 | 32 | 269 |
| Grade 12 |  |  |  |  |  |  |
| Nation | 100 | 294 | 100 | 299* | 100 | 304* $\dagger$ |
| Northeast | 24 | 300 | 24 | 303 | 22 | 307 |
| Southeast | 20 | 284 | 23* | 292* | 22 | 296* |
| Central | 27 | 297 | 25 | 304 | 24 | 310* |
| West | 29 | 294 | 27 | 299 | 33 | 303* |

* Indicates a significant difference from 1990.
† Indicates a significant difference from 1992.
Percentages may not total 100 due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, and 1996 Mathematics Assessments.

In addition to national results, fourth- and eighth-grade results were available for the states and jurisdictions that participated in the 1996 state-level mathematics assessment. State-by-state results were also available for many of these jurisdictions for 1990 (grade 8) and 1992 (grades 4 and 8).

In 1996, public school fourth graders in 16 of the 47 participating jurisdictions performed better than the national average of 222 , while 15 jurisdictions performed below the national average. (The remaining 16 jurisdictions performed at or around the national average.) Mirroring the trend for the nation, the average mathematics scores for 15 of the 39 jurisdictions that participated in the 1992 and 1996 mathematics assessments increased in the latest assessment. Only three jurisdictions had averages that declined during this time.

| Talde 2.2 | erage Mathema ade 4 Public Sch | Scores |  |
| :---: | :---: | :---: | :---: |
|  | 1996 Average Scale Score | Change from 1992Average Scale Score |  |
| Maine | 232 | 1 |  |
| Minnesota | 232 | 4\#+ |  |
| Connecticut | 232 | 5\# |  |
| Wisconsin | 231 | 3 |  |
| North Dakota | 231 | 2 |  |
| Indiana | 229 | 8\# |  |
| lowa $\ddagger$ | 229 | -1 |  |
| Massachusetts | 229 | 2 |  |
| Texas | 229 | 11 + |  |
| Nebraska | 228 | 2 |  |
| Montana $\ddagger$ | 228 | - |  |
| New Jersey $\ddagger$ | 227 | 0 |  |
| Utah | 227 | 2 |  |
| Michigan $\ddagger$ | 226 | 6+1 |  |
| Pennsylvania $\ddagger$ | 226 | 2 |  |
| Colorado | 226 | 5tt |  |
| Washington | 225 | - |  |
| Vermont $\ddagger$ | 225 |  |  |
| Missouri | 225 | 3 |  |
| North Carolina | 224 | 11 H |  |
| DDESS | 224 | - |  |
| Alaska $\ddagger$ | 224 | - |  |
| Oregon <br> West Virginia | $223$ | - + |  |
| DoDDS | 223 | 8 |  |
| Wyoming | 223 | -2 |  |
| Virginia | 223 | 2 |  |
| New York $\ddagger$ | 223 | 4\#t |  |
| Nation | 222 | $4+$ |  |
| Maryland | 221 | 3 |  |
| Rhode Island | 220 | 5\# |  |
| Kentucky | 220 | 5\# |  |
| Tennessee | 219 | 8\# |  |
| Nevada $\ddagger$ | 218 | 2 |  |
| Arizona $\ddagger$ | 218 | 2 |  |
| Arkansas Florida | 216 | $6 \pm$ |  |
| Georgia | 215 | 0 |  |
| Delaware | 215 | -3 \# |  |
| Hawaii | 215 | 1 |  |
| New Mexico | 214 | 1 |  |
| South Carolina $\ddagger$ | 213 | 1 |  |
| Alabama | 212 | 3 |  |
| California Louisiana | 209 | 1 |  |
| Louisiana | 209 | 5\# |  |
| Mississippi | 208 | 7\# |  |
| District of Columbia | 188 | -4 \# -5 t |  |

The change between 1992 and 1996 was calculated using unrounded average scale scores for the two assessments.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates in 1996 (see Appendix A).
\#t Indicates that the change since 1992 in average scale scores is significant at a 5 -percent level of significance using a multiple comparison procedure based on 39 jurisdictions (excluding the nation).

- Indicates that the jurisdiction did not participate in 1992.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992 and 1996
Mathematics Assessments.

The results at grade 8 are also encouraging. Of the 37 jurisdictions that participated in the 1992 and 1996 mathematics assessments, 13 showed improvements in the latest assessment (and of the 32 jurisdictions that participated in 1990 and 1996 all but 5 showed an improvement). No jurisdiction showed a significant decline in performance between 1992 and 1996. The 1996 results showed that eighth graders in 19 jurisdictions performed better than the national average of 271 , while 16 jurisdictions performed below the national average. (The remaining 9 jurisdictions performed at or about the national average.)

## Table 2.3 <br> Average Mathematics Scale Scores Grade 8 Public Schools



|  | 1996 Average Scale Score | Change from 1992 Average Scale Score | Change from 1990 Average Scale Score |
| :---: | :---: | :---: | :---: |
| North Dakota | 284 | 1 | 3* |
| Maine | 284 | 5 \# | - |
| Minnesota | 284 | 2 | 9 ** |
| lowa $\ddagger$ | 284 | 1 | 6** |
| Montana $\ddagger$ | 283 | - | 3 |
| Wisconsin $\ddagger$ | 283 | $5 \dagger$ | 8 ** |
| Nebraska | 283 | 5 \# | $7{ }^{* *}$ |
| Connecticut | 280 | 6 H | $10^{* *}$ |
| Vermont $\ddagger$ | 279 | - | - |
| Alaska $\ddagger$ | 278 | - | - |
| Massachusetts | 278 | $5 \dagger$ | - |
| Michigan $\ddagger$ | 277 | 10 \# | 12** |
| Utah | 277 | 2 | - |
| Oregon | 276 | - | $5^{* *}$ |
| Washington | 276 | - | - |
| Colorado | 276 | $3 \dagger$ | 8 ** |
| Indiana | 276 | 5 \# | 8 ** |
| DoDDS | 275 | - | - |
| Wyoming | 275 | 0 | $3^{* *}$ |
| Missouri | 273 | 2 | - |
| Nation | 271 | 5 † | 8* |
| New York $\ddagger$ | 270 | 4 | 9** |
| Texas | 270 | 6 \# | 12** |
| Virginia | 270 | 2 | $5^{* *}$ |
| Maryland $\ddagger$ | 270 | 5 | 9** |
| DDESS | 269 | - | - |
| Rhode Island | 269 | 3 \# | 9** |
| Arizona | 268 | 3 | 8 ** |
| North Carolina | 268 | 9 \# | $17^{* *}$ |
| Delaware | 267 | 4 + | 6** |
| Kentucky | 267 | 4 \# | $9^{* *}$ |
| West Virginia | 265 | 6 \# | 9 ** |
| Florida | 264 | 4 | 8** |
| Tennessee | 263 | 4 † | - |
| California | 263 | 2 | $6^{* *}$ |
| Georgia | 262 | 3 | 4 |
| Hawaii | 262 | 5 \# | $11^{* *}$ |
| New Mexico | 262 | 2 | $6^{* *}$ |
| Arkansas $\ddagger$ | 262 | 5 \# | $5^{* *}$ |
| South Carolina $\ddagger$ | 261 | 0 | - |
| Alabama | 257 | 4 | 4 |
| Lovisiana | 252 | 2 | $6^{* *}$ |
| Mississippi | 250 | 4 † | - |
| Guam | 239 | 4 | $7{ }^{* *}$ |
| District of Columbia | 233 | -2 | 1 |

The changes between scale scores were calculated using unrounded average scale scores for the two assessments.
$\ddagger$ Indicates jurisdiction did not satisfy one or more of the guidelines for school participation rates in 1996 (see Appendix A).

+ Indicates change in scale scores from 1992 is significant at a 5 -percent level of significance using a multiple comparison procedure based on 37 jurisdictions (excluding the nation).
$\dagger$ Indicates change in scale scores from 1992 is significant at a 5-percent level of significance if only one jurisdiction is being examined.
** Indicates change in scale scores from 1990 is significant at a 5 -percent level of significance using a multiple comparison procedure based on 32 jurisdictions (excluding the nation).
* Indicates change in scale scores from 1990 is significant at a 5 -percent level of significance if only one jurisdiction is being examined.
- Indicates jurisdiction did not participate in 1990 and/or 1992.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

## Performance of Selected Subgroups

## Gender

Over the years, the results of NAEP assessments in mathematics have shown few significant gender differences. The differences that were seen typically favored males at grade $12 .{ }^{8}$ These differences at grade 12 could possibly be explained by the differential course-taking and dropout rates by gender. ${ }^{9}$

The results of the NAEP mathematics assessments in 1990 and 1992 confirmed the finding of previous gender difference research. In 1990 and 1992, the only significant difference between male and female students occurred at grade 12 . At grades 4 and 8 , the performance of male and female students was not significantly different. However, the 1996 assessment showed a significant difference at the fourth grade, with male students scoring higher than female students. This fourth-grade difference was the only significant difference observed in 1996.

The trend of increasing scores observed for all students was also observed for male and female students at all grades. From 1990 to 1996, scores for male students increased by 8 to 12 points, while scores for female students increased by 9 to 12 points.

| Table 2.4 | Average Mothemotics Scole Scores by Cender |  |  |  |  | THE NATION's |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grade 4 | 1990 |  | 1992 |  | 1996 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| All Students | 100 | 213 | 100 | 220* | 100 | 224* $\dagger$ |
| Male | 52 | 214 | 50 | $221 *$ | 51 | 226* $\dagger$ |
| Female | 48 | 213 | 50 | 219* | 49 | 222*† |
| Grade 8 |  |  |  |  |  |  |
| All Students | 100 | 263 | 100 | 268* | 100 | 272* $\dagger$ |
| Male | 51 | 263 | 51 | 268* | 52 | 272* |
| Female | 49 | 262 | 49 | 269* | 48 | 272* $\dagger$ |
| Grade 12 |  |  |  |  |  |  |
| All Students | 100 | 294 | 100 | 299* | 100 | 304* $\dagger$ |
| Male | 48 | 297 | 49 | 301* | 48 | 305* $\dagger$ |
| Female | 52 | 291 | 51 | 298* | 52 | 303* $\dagger$ |

*Indicates a significant difference from 1990.
†Indicates a significant difference from 1992.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, and 1996
Mathematics Assessments.

[^10]
## Race/Ethnicity

Research during past decades and results from NAEP assessments in mathematics have shown significant performance differences among racial/ethnic subgroups. ${ }^{10}$ Some studies have suggested that the basis for the differences in levels of performance resides in the opportunities available to students, including opportunities to attend effective schools, ${ }^{11}$ opportunities afforded by social and economic factors of the home and school location, ${ }^{12}$ and opportunities brought by encouragement to continue studies in mathematics. ${ }^{13}$ The possibility of these factors should be recognized and considered when interpreting differences in subgroup performance.

The results for eighth-grade Asian/Pacific Islander students are not included in the main body of this report. A thorough investigation into the quality and credibility of these results, which included an independent review by the National Institute of Statistical Sciences, was initiated by NCES. Collateral results from the grade 8 state assessment program in mathematics suggested that the 1996 national results may substantially underestimate actual achievement of the Asian/Pacific Islander group. Because of its potential to misinform, NCES decided to omit the national grade 8 Asian/Pacific Islander results from the body of the report and to include them in an appendix. Appendix E includes 1996 results for this group along with a description of the findings that led to this decision.

As with previous NAEP assessments in a variety of academic subjects, differences in mathematics performance among racial/ethnic groups were evident at all three grades in 1996. As shown in Table 2.5, White and Asian/Pacific Islander fourth-, and twelfth-grade students and White eighth-grade students scored higher than their Black and Hispanic counterparts. At grade 4 American Indian students also outperformed Black and Hispanic students but scored lower than White students and Asian/Pacific Islanders. Finally, Hispanic eighth graders outperformed their Black counterparts.

Although scores have increased for many of the racial/ethnic groups since 1990, the differences in performance for White students and their Black and Hispanic counterparts at all three grades have remained stable. For example, in 1990, the average score differences between White students and Black and Hispanic twelfth grade students were 33 and 25 points, respectively. In 1996, these differences were 31 and 24 points, respectively.

[^11]| Table 2.5 | Average Mathematics Scale Scores by Race/Ethnicity |  |  |  |  | THE NATION'S |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grade 4 | 1990 |  | 1992 |  | 1996 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| All Students | 100 | 213 | 100 | 220* | 100 | 224* $\dagger$ |
| White | 70 | 220 | 70 | 228* | 68* $\dagger$ | 232* $\dagger$ |
| Black | 15 | 189 | 16* | 193 | 15* $\dagger$ | 200* |
| Hispanic | 10 | 198 | 9* | 202 | 13* $\dagger$ | 206* |
| Asian/Pacific Islander | 2 | 228 | 2 | 232 | 3* | 232 |
| American Indian | 2 | 208 | 1 | 211 | 2 | 216 |
| Grade 8 |  |  |  |  |  |  |
| All Students | 100 | 263 | 100 | 268* | 100 | 272* $\dagger$ |
| White | 71 | 270 | 70 | 278* | 69* | 282* |
| Black | 15 | 238 | 16 | 238 | 14* $\dagger$ | 243 |
| Hispanic | 10 | 244 | 10 | 247 | 12* $\dagger$ | 251 |
| Asian/Pacific Islander | $2!$ | 279 ! | 3 | 288 | -- | -- |
| American Indian | $2!$ | 246 ! | 1 | 255 | $1!$ | 264! |
| Grade 12 |  |  |  |  |  |  |
| All Students | 100 | 294 | 100 | 299* | 100 | $304 *$ + |
| White | 74 | 301 | 71* | 306* | 70* | $311^{*}+$ |
| Black | 14 | 268 | 15 | 276* | 14 | 280* |
| Hispanic | 8 | 276 | 9 | 284 | 11* | 287* |
| Asian/Pacific Islander | 3 | 311 | 4 | 316 | 4 | 319 |
| American Indian | $1!$ | *** | 1 | *** | 1! | 279 ! |

* Indicates a significant difference from 1990.
$\dagger$ Indicates a significant difference from 1992.
*** Sample size insufficient to permit a reliable estimate.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
-- Quality control activities and special analyses involving state assessement data raised concerns about the accuracy and precision of national grade 8 Asian/Pacific results. As a result, they are omitted from the body of this report. See Appendix A for a more detailed discussion. Percentages may not total 100 due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.


## Parents' Highest Level of Education

Students were asked to indicate the highest level of education completed by each parent. Four levels of education were identified: did not finish high school, graduated from high school, some education after high school, and graduated from college. A choice of "I don't know" was also available. For this analysis, the highest education level reported for either parent was used. It should be noted that 36 percent of fourth graders, 11 percent of eighth graders, and 3 percent of twelfth graders reported not knowing the education level of either of their parents.

Other research has suggested that connections between parents' education levels and students' achievements may result from the influence that parents have on students' attitudes toward mathematics. ${ }^{14}$ Because some research has questioned the accuracy of student-reported data among similar groups of students, caution should be used in interpreting the findings. ${ }^{15}$ Still, previous NAEP assessments in all subject areas, as well as other research, have found that the student-reported level of parental education exhibits a positive relationship with student performance on the assessments.

At all three grade levels in 1996, students who reported that neither parent had graduated from high school scored lower than those who reported that at least one parent had graduated from high school. This latter group, in turn, scored lower than those who reported that at least one parent had received some education after high school or graduated from college. At grade 12, students who reported that at least one parent had received some education after high school scored lower than those who reported that at least one parent had graduated from college.

From 1990 to 1996, average scale scores showed gains for all levels of student reported parental education, with one exception - grade 4 students who reported that neither parent had graduated from high school.

[^12]


* Indicates a significant difference from 1990.
† Indicates a significant difference from 1992.
Percentages may not total 100 due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, and 1996 Mathematics Assessments.


## Type of School

Previous NAEP mathematics assessments and other survey research on educational achievement have found significant differences in the performance of students attending public and nonpublic schools. ${ }^{16}$ However, the reader is cautioned against using NAEP results to make simplistic inferences about the relative effectiveness of public and nonpublic schools. Average performance differences between the two types of schools may be related in part to socioeconomic and sociological factors, such as the level of parental involvement in the child's education. To get a clearer picture of the differences between public and nonpublic schools, more in-depth investigations that are beyond the scope of the NAEP assessment program must be conducted.

In 1996, approximately 90 percent of the nation's fourth-, eighth-, and twelfth-grade students attended public schools, with the remaining students attending nonpublic schools (Catholic and other private schools). For each assessment, students attending nonpublic schools scored higher than those attending public schools.

In general, between 1990 and 1996, both public and nonpublic school students' average mathematics scores improved with the majority of the 8 - to 13 -point increases occurring between 1990 and 1992, with one exception - the largest increase for nonpublic school students at grade 4 occurred between 1992 and 1996. (The sample of twelfth graders in 1990 who attended nonpublic schools was not adequate to permit comparisons.)

[^13]| Table 2.7 | Average Mathematics Scale Scores by Type of School |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grade 4 | 1990 |  | 1992 |  | 1996 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| All Students | 100 | 213 | 100 | 220* | 100 | 224* $\dagger$ |
| Public Schools | 89 | 212 | 88 | 219* | 89 | 222** |
| Nonpublic Schools: | 11 | 224 | 12 | 228 | 11 | 237* $\dagger$ |
| Catholic | 7 | 219 | 8 | 228* | 7 | 232* |
| Other Private Schools | $4!$ | 233! | 4 | 230 | $4!$ | 247! |
| Grade 8 |  |  |  |  |  |  |
| All Students | 100 | 263 | 100 | 268* | 100 | 272* $\dagger$ |
| Public Schools | 92 | 262 | 89 | 267* | 89 | 271*+ |
| Nonpublic Schools: | 8 | 271 | 11 | 281* | 11 | 284* |
| Catholic | 5 | 271 | 6 | 278 | 6 | 283 |
| Other Private Schools | $3!$ | 272 ! | 5 | 284 | 4 | 286 |
| Grade 12 |  |  |  |  |  |  |
| All Students | 100 | 294 | 100 | 299* | 100 | 304* $\dagger$ |
| Public Schools | 91 | 294 | 87 | 297* | 88 | 303* $\dagger$ |
| Nonpublic Schools: | $9!$ | 300! | 13 | 314 | 12 | 314 |
| Catholic | $5!$ | 301! | 8 | 311 | 8 | 311 |
| Other Private Schools | $3!$ | 298! | $4!$ | 320! | 4 | 321 |

*Indicates a significant difference from 1990.
$\dagger$ Indicates a significant difference from 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
Percentages may not total 100 due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

## Participation in Title I Programs

The Improving America's Schools Act of 1994 (P.L. 103-382) reauthorized the Elementary and Secondary Education Act of 1965 (ESEA). Title I Part A of ESEA provides local education agencies with financial assistance to meet the educational needs of children who are failing or most at risk of failing. ${ }^{17}$ Title I programs are designed to help disadvantaged students meet challenging academic performance standards. Through Title I, schools are assisted in improving teaching and learning and in providing students with opportunities to acquire the knowledge and skills outlined in their state's content and performance standards. For high poverty Title I schools, all children in the school may benefit through participation in schoolwide programs. Title I funding supports state and local education reform efforts and promotes coordinating of resources to improve education for all students.

NAEP first collected student-level information on participation in Title I programs in 1994. Therefore, results comparing the performance of participating and nonparticipating students are not available for previous NAEP mathematics assessments. The NAEP program will continue to monitor the performance of Title I program participants in future assessments. The Title I information collected by NAEP refers to current participation in Title I services. Students who participated in such services in the past but do not currently receive services are not identified as Title I participants. Differences between students who receive Title I services and those who do not should not be viewed as an evaluation of Title I programs. Typically, Title I services are intended for students who score poorly on assessments. To properly evaluate Title I programs, the performance of students participating in such programs must be monitored over time and their progress must be assessed. ${ }^{18}$

The percentage of students participating in Title I programs and services was 22 percent at grade 4,12 percent at grade 8 , and 2 percent at grade 12 . At grades 4 and 8 , students who participate in Title I programs scored lower than their nonparticipating classmates. (The sample of twelfth graders was not adequate enough to permit a comparison between students who participated and those who did not.)

| Table 2.8 | Average Mathematics Scale Scores by Title I Participation |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | GRADE 4 |  | GRADE 8 |  | GRADE 12 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| All Students | 100 | 224 | 100 | 272 | 100 | 304 |
| Participated | 22 | 200 | 12 | 244 | $2!$ | 270 ! |
| Did Not Participate | 78 | 231 | 88 | 276 | 98 | 305 |

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^14]
## Eligibility for the Free/Reduced-Price Lunch Program

The free/reduced-price lunch component of the National School Lunch Program (NSLP), offered through the U.S. Department of Agriculture (USDA), is designed to ensure that children near or below the poverty line receive nourishing meals. ${ }^{19}$ Eligibility for the free/reduced-priced lunch program is included as an indicator of poverty. The program is available to public schools, nonprofit private schools, and residential child care institutions. Eligibility for free or reduced priced meals is determined through the USDA's Income Eligibility Guidelines.

NAEP first collected information on student-level eligibility for the federally funded NSLP in 1996. Although results cannot be presented for previous NAEP mathematics assessments, the NAEP program will continue to monitor the performance of these students in future assessments.

In 1996, the percentage of students eligible for the program was 31 percent at grade 4, 27 percent at grade 8 , and 13 percent at grade 12 . At all three grades, students who were eligible for NSLP scored lower than their classmates who were not eligible. (Information on eligibility was not available for 16 percent of fourth graders, 15 percent of eighth graders, and 27 percent of twelfth graders.)

| Table 2.9 | Average Mathematics Scale Scores by Free/Reduced-Price Lunch Program Eligibility |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | GRADE 4 |  | GRADE 8 |  | GRADE 12 |  |
|  | Percentige | Averoge Scale Score | Perentige | Averoge Scale Score | Percentige | Averge Scale Score |
| All Students | 100 | 224 | 100 | 272 | 100 | 304 |
| Eligible | 31 | 207 | 27 | 252 | 13 | 281 |
| Not Eligible | 53 | 231 | 55 | 280 | 60 | 307 |
| Information Not Available | 16 | 233 | 17 | 280 | 27 | 308 |

Percentanges may not total 100 due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^15]
## Chapter 3

## Achievement Level Results

## Achievement Level Descriptions

The percentages of students who attained each of the achievement levels in the NAEP 1996 mathematics assessment are presented in this chapter. Results are displayed for the nation, for regions, and for selected subgroups.

The National Education Statistics Act of 1994 requires the National Assessment Governing Board (NAGB) to develop "appropriate student performance levels" for reporting NAEP results. The law requires that these levels are "used on a developmental basis until the Commissioner of Education Statistics determines . . . that such levels are reasonable, valid, and informative to the public." It requires the Commissioner and the Governing Board to make clear the developmental status of such levels.

The student achievement levels in this report have been developed and adopted by NAGB, NAEP's independent policy-making body, with contributions from a wide variety of educators, business and government leaders, and interested citizens. These levels have been established to help Americans answer two questions that are important to parents and to all citizens in the communities and nation in which we live. These questions are: "What should students know and be able to do as they progress and graduate from school?" and "How good is good enough in terms of student achievement on NAEP?" Answering these questions obviously involves judgments. NAGB is not suggesting that there is a single answer to these questions. Rather, NAGB is trying to put forward reasonable judgments that can inform citizens across the United States - information they can use to answer these questions in their own schools and communities.

Developing carefully considered judgments about "what students should know and be able to do" and "how good is good enough" is both difficult and controversial. The Governing Board believes that these questions are so important that answers must be sought in an informed, responsible way. The process is subject to revision and refinement as appropriate.

The achievement levels in this report, approved by NAGB, are the result of a standardsetting process, designed and conducted by NAGB's contractor, American College Testing (ACT). This process was reviewed by scores of individuals, including policy makers, professional organizations, teachers, parents, and other members of the general public. To develop the levels, ACT convened a cross-section of educators and interested citizens from across the nation and asked them to recommend what students should know and be able to do
in mathematics. Prior to adopting these levels of student achievement, NAGB engaged a large number of persons to comment on the recommended levels and to review the results.

The result of the achievement level-setting process is a set of achievement level descriptions and a set of achievement level cutpoints on the 500 -point NAEP scale. The cutpoints are minimum scores that define Basic, Proficient, and Advanced performance at grades 4, 8, and 12. At present, evaluations conducted on the level-setting process and critiques of these evaluations have provided mixed reviews. Therefore, both the Governing Board and the Commissioner of Education Statistics regard the achievement levels as developmental; they should not be interpreted as statistically conclusive. Because these levels are still considered developmental, the reader of this report is advised to consider that status when interpreting the results. The reader should recognize that the results are based on the judgments of panels, approved by NAGB, of what Basic, Proficient, and Advanced students should know and be able to do in mathematics, as well as on their judgments regarding what percent of students at the borderline for each level should answer each question correctly. The latter information is used in translating the achievement level descriptions into cutpoints on the NAEP scale. NCES uses these levels in reporting NAEP results, but it does not currently adjudicate the reliability or validity of these achievement levels. Rather they are reported directly as adopted by NAGB.

The National Assessment Governing Board urges all who are concerned about "what students should know and be able to do" and "how good is good enough" to read and interpret these achievement levels recognizing that this is a developing, judgmental process and is subject to various interpretations. The decision to include the levels in NAEP reports is an attempt to make the assessment results more useful for parents, educators, and policy makers by providing performance standards against which to measure educational progress.

As explained in Chapter 1, these achievement levels - Basic, Proficient, and Advanced - are used to report NAEP results. The Basic level denotes partial mastery of the knowledge and skills that are fundamental for proficient work at a given grade. The Proficient level represents solid academic performance. Students reaching this level demonstrate competency with a range of challenging subject matter. The Advanced level signifies superior performance at a given grade.

Specific definitions of these levels of mathematics achievement as they apply at each of the three grades, are presented in Figures 3.1 through 3.3. For each grade, the definitions are cumulative from Basic through Advanced. In other words, students performing at the Proficient level also display the competencies associated with the Basic level, while students performing at the Advanced level demonstrate skills and knowledge associated with the preceding levels.

| Figure 3.1 | NAEP Mothematics AchieVement Levels - Grode 4 4 |
| :--- | :--- |

## Figure 3.2 NAEP Mathematics Achievement Levels - Grade 8

|  | Basic <br> (262) |
| :--- | :--- | | Eighth-grade students performing at the basic level should exhibit evidence of |
| :--- |
| conceptual and procedural understanding in the five NAEP content strands. This level |
| of performance signifies an understanding of arithmetic operations - including |
| estimation - on whole numbers, decimals, fractions, and percents. |

Eighth graders performing at the proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of basic level arithmetic operations - an understanding sufficient for problem solving in practical situations.
Quantity and spatial relationships in problems solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs; apply properties of informal geometry; and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability. the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content strands.
Eighth graders performing at the advanced level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth graders performing at the advanced level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

## Figure 3.3 NAEP Mathematics Achievement Levels - Grade 12

Basic Twelfth-grade students performing at the basic level should demonstrate procedural and conceptual knowledge in solving problems in the five NAEP content strands.

Twelfth grade students performing at the basic level should be able to use estimation to verify solutions and determine the reasonableness of results as applied to real-world problems. They are expected to use algebraic and geometric reasoning strategies to solve problems. Twelfth graders performing at the basic level should recognize relationships presented in verbal, algebraic, tabular, and graphical forms; and demonstrate knowledge of geometric relationships and corresponding measurement skills.

They should be able to apply statistical reasoning in the organization and display of data and in reading tables and graphs. They also should be able to generalize from patterns and examples in the areas of algebra, geometry, and statistics. At this level, they should use correct mathematical language and symbols to communicate mathematical relationships and reasoning processes; and use calculators appropriately to solve problems.

Proficient Twelfth-grade students performing at the proficient level should consistently integrate (336) mathematical concepts and procedures to the solutions of more complex problems in the five NAEP content strands.

Twelfth graders performing at the proficient level should demonstrate an understanding of algebraic, statistical, and geometric and spatial reasoning. They should be able to perform algebraic operations involving polynomials; justify geometric relationships; and judge and defend the reasonableness of answers as applied to real-world situations. These students should be able to analyze and interpret data in tabular and graphical form; understand and use elements of the function concept in symbolic, graphical, and tabular form; and make conjectures, defend ideas, and give supporting examples.

Twelfth-grade students performing at the advanced level should consistently demonstrate the integration of procedural and conceptual knowledge and the synthesis of ideas in the five NAEP content strands.

Twelfth-grade students performing at the advanced level should understand the function concept; and be able to compare and apply the numeric, algebraic, and graphical properties of functions. They should apply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous and discrete mathematics.

They should be able to formulate generalizations and create models through probing examples and counterexamples. They should be able to communicate their mathematical reasoning through the clear, concise, and correct use of mathematical symbolism and logical thinking.

## National and State Results

The percentage of students performing at or above the three achievement levels, as well as the percentage of those performing below the Basic level, are presented in Table 3.1. Consistent with the national scale score results discussed in Chapter 2, achievement level results reflect patterns of increasing achievement at all three grades. Compared to results for the 1990 mathematics assessment, the percentage of the nation's fourth-, eighth-, and twelfth-grade students performing at or above the Basic and Proficient levels has increased. In addition, the percentage of students at all three grades performing at or above the Basic level has increased in the four years between the 1992 and 1996 assessments. However, only 2 percent of fourth and twelfth graders and 4 percent of eighth graders performed at the Advanced level in 1996; only at grade 8 was there an appreciable change seen since the 1990 assessment.

In general, a greater percentage of students attending schools in the Northeast and Central regions were at or above the Basic and Proficient levels when compared with students in the Southeast and West regions. The two exceptions were (1) grade 8, at which the percentage of students in the Central region attaining the Proficient level was greater than that for the Southeast but not significantly different from those in the West regions, and (2) grade 12, at which the percentage of students in the West region attaining the Basic level was not significantly different from those in the Northeast or Central regions. Also at grade 12, the percentage of students in the West region attaining the Basic level was greater than that for the Southeast region.

The percentage of students at or above the Basic and Proficient level increased between 1990 and 1996 in many regions. (Significant changes in percentages are indicated in Table 3.1.) For example, at grade 4, increases in the percentages of students at or above the Basic level were seen in the Northeast, Southeast, and Central regions.


* Indicates a significant difference from 1990.
† Indicates a significant difference from 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

In addition to national results, fourth- and eighth-grade results are available for states and jurisdictions that took part in the 1996 state-level mathematics assessments. State-level results are also available for many of these jurisdictions for 1990 (grade 8) and 1992 (grades 4 and 8).

As shown in Table 3.2, of the 39 jurisdictions that participated in the 1992 and 1996 mathematics assessments, the percentage of fourth graders performing at or above the Basic level increased for 9 jurisdictions between 1992 and 1996. (For the remaining 27 jurisdictions, no changes in the percentage of students at this level were observed.) For 7 of the 39 jurisdictions, the percentage of students performing at or above the Proficient level showed improvements in the latest assessment. No jurisdiction showed a decline in the percentage of fourth graders attaining the Proficient level during this time. Across the jurisdictions, 3 percent or less of the students were at the Advanced level.

Table 3.2

$\ddagger$ Indicates jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
H Indicates change in percentages from 1992 is significant at a 5 -percent level of significance using a multiple comparison procedure based on 39 jurisdictions (excluding the nation).
$\dagger$ Indicates change in percentages from 1992 is significant at a 5 -percent level of significance if only one jurisdiction is being examined.

- Indicates jurisdiction did not participate in 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),
1992 and 1996 Mathematics Assessments.

Between 1990 and 1996, the percentage of eighth graders performing at or above the Basic level increased for 27 jurisdictions while no decreases in the percentage of students at this level were observed (see Table 3.3). For 26 of the 32 jurisdictions that participated in the 1990 and 1996 mathematics assessments, the percentage of students performing at or above the Proficient level increased with no jurisdiction showing a decline in the percentage of eighth graders attaining this level. Across the jurisdictions, between 0 and 7 percent of the students were at the Advanced level in 1996.

$\ddagger$ Indicates jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
** Indicates change in percentages from 1990 is significant at a 5 -percent level of significance using a multiple comparison procedure based on 32 jurisdictions (excluding the nation).

* Indicates change in percentages from 1990 is significant at a 5 -percent level of significance if only one jurisdiction is being examined.
\# Indicates change in percentages from 1992 is significant at a 5 -percent level of significance using a multiple comparison procedure based on 37 jurisdictions (excluding the nation).
$\dagger$ Indicates change in percentages from 1992 is significant at a 5 -percent level of significance if only one jurisdiction is being examined.
*** Sample size insufficient to permit a reliable estimate.
- Indicates jurisdiction did not participate in 1990 and/or 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996
Mathematics Assessments.


## Performance of Selected Subgroups

In this section, variations in performance among the major reporting subgroups are discussed. Again, the reader is cautioned against making simple or causal inferences related to the performances of various subgroups of students, the effectiveness of public and nonpublic schools, or the impact of Title I and Free/Reduced Price lunch programs.

## Gender

In 1996, gender differences were seen at grades 4 and 12 in the percentages of students attaining the Proficient and Advanced levels (see Table 3.4). At both grades, the percentages of male students at the Advanced level and at or above the Proficient level were greater than the percentages of female students. As for the Basic level at grades 4 and 12 and for all levels at grade 8, no significant differences between males and females were observed.

At all three grades in 1996, the percentages of male and female students at or above the Proficient and Basic level increased since 1990 with one exception - male twelfth graders at or above Proficient. Between 1992 and 1996, the percentage of female students at or above the Basic level increased for grades 8 and 12. The only significant difference in the percentage at or above the Proficient level was for fourth grade males.


* Indicates a significant difference from 1990.
$\dagger$ Indicates a significant difference from 1992.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.


## Race/Ethnicity

In 1996, differences among the racial/ethnic groups were apparent at grades 4, 8, and 12 (see Table 3.5 ${ }^{20}$. At the Advanced level, the percentages of White students were greater than those of Black students at grades 4 and 12, and Hispanic students at grade 8.

In 1996, at grade 4, the percentages of White and Asian/Pacific Islander students at or above the Basic and Proficient levels were greater than those for their Black, Hispanic, or American Indian peers. Also, the percentage of American Indian fourth graders at or above the Basic level were greater than the percentage of Black fourth graders. The percentages of White fourth graders at or above the Basic and Proficient levels were greater in 1996 than in 1990. For Black students, the percentage at or above Basic was greater in 1996 than the percentage in 1990.

In 1996, at grade 8, the percentages of White students at or above the Basic and Proficient levels were greater than the percentages for Black and Hispanic students. Also, the percentage of Hispanic eighth graders attaining the Basic level was greater than that for their Black classmates. Between 1990 and 1996, the percentages of students at or above the Basic and Proficient levels increased for White students.

In 1996, at grade 12, the percentages of White and Asian/Pacific Islander students at or above the Basic and Proficient levels were greater than those for Black and Hispanic students. In 1996, the percentages of White students at or above the Basic and Proficient levels were greater than in 1990. The percentage of White twelfth graders at or above Basic in 1996 was also greater than that in 1992.

[^16]

## Parents' Highest Level of Education

Table 3.6 presents achievement level results based on students' reports of their parents' highest levels of education. The background questionnaire that accompanied the assessment asked students to indicate the education level of each of their parents. The highest level of education was used for reporting. For example, for students who indicated that their fathers had graduated from high school and their mothers had received some education after high school, the level of education used would be some education after high school. Appendix A presents more detailed discussion of the parental education variable.


* Indicates a significant difference from 1990.
$\dagger$ Indicates a significant difference from 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996
Mathematics Assessments.

A substantial number of fourth-grade students, 36 percent, did not know the education level of either parent. The problem was less severe at grades 8 and 12; the percentages of students who did not know their parents' education level were 11 and 3 percent, respectively.

As shown in Table 3.6, parental education and student achievement are positively related. This mirrors the average scale score results discussed in the previous chapter. At grade 12 , students who reported that at least one parent graduated from college were more likely to reach the Advanced level than were those who reported lower levels of parental education. At grade 8 , students who reported that at least one parent graduated from college were more likely to reach the Advanced level than were those who reported that their parents had graduated from high school.

At the Basic and Proficient levels, the patterns also support the positive relationship between parental education and mathematics achievement. In general, the percentage of students at each grade who attained the Basic and Proficient levels increased as parental education increased.

In general, the percentages of students at or above the Basic level increased between 1990 and 1996, regardless of parental education. ${ }^{21}$ Gains in the percentages of students at or above Basic were also seen between 1992 and 1996 for eighth and twelfth graders who reported that at least one parent had received some education after high school. As for the percentages of students at or above the Proficient level, gains between 1990 and 1996 were observed for fourth graders who reported that at least one parent had graduated from high school or from college and for eighth graders who reported that at least one parent had received some education after high school or had graduated from college.

[^17]
## Type of School

Table 3.7 shows the differences between public and nonpublic schools at all three grades in terms of the percentages of students at or above each of the achievement levels. In 1996, the percentages of fourth-, eighth-, and twelfth-grade students attending nonpublic schools who were at or above the Basic and Proficient levels were greater than the percentages of students attending public schools.

Since 1990, the percentage of students attending public schools and who were at or above the Basic and Proficient levels increased at grades 4 and 8 . Similar increases were seen at the Basic level at grade 12. The same was true for fourth graders attending nonpublic schools. Between 1990 and 1996, the percentage of eighth graders attending nonpublic schools who attained the Proficient and Advanced levels also increased. No significant differences were observed for twelfth graders attending nonpublic schools.

Between 1992 and 1996, the percentages of students at grades 8 and 12 who attended public school and who were at or above the Basic level increased. For fourth graders attending nonpublic schools, the percentage of students at or above the Basic and Proficient level also increased. During this time period, no significant differences were observed for eighth and twelfth graders attending nonpublic schools.


* Indicates a significant difference from 1990.
$\dagger$ Indicates a significant difference from 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately
determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.


## Participation in Title I Programs

Table 3.8 presents the achievement levels for each grade in terms of students' participation in Title I programs. In 1996, at grades 4 and 8 , the percentages of students who were not currently receiving Title I services and who were at or above the Basic and Proficient levels were greater than those of students who were receiving Title I services. (Grade 12 differences are not discussed here because the nature of the grade 12 sample prohibits an accurate estimation of the variability of the percentage of Title I recipients.)

| Table 3.8 | Percentage Attaining Mathematics Achievement Levels by Title I Participation |  |
| :---: | :---: | :---: |



[^18]
## Eligibility for the Free/Reduced-Price Lunch Program

Table 3.9 presents the achievement levels for each grade by students' eligibility for the free/ reduced-price lunch component of the National School Lunch Program (NSLP). At all three grades, the percentages of students who were not eligible for this program and who were at or above the Basic and Proficient levels were greater than those of students who were eligible. This was also true for fourth and eighth graders at the Advanced level.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

## Chapter 4

## Exploring a More Inclusive NAEP

The 1996 national and state NAEP mathematics assessments were conducted in a manner that ensured the reporting of valid trend results. Samples of students were assessed using materials and administration procedures consistent with those used for the 1990 and 1992 assessments. The results reported in Chapters 1, 2, and 3 of this report are based on these samples. In addition to these core assessment activities, the 1996 assessment included supplemental samples of schools and students. The supplemental samples were designed to allow the program to study the feasibility and impact of increasing the numbers of Limited English Proficient (LEP) students and students with disabilities that are included in NAEP and assessed in an appropriate manner. Specifically, revised inclusion rules were implemented in one sample and assessment accommodations and adaptations were permitted in another. This chapter describes these additional samples and procedural revisions and presents some initial results on research issues pertinent to the development of a more inclusive NAEP.

Because it serves as the Nation's Report Card, the intent of NAEP has been to report results that reflect the achievement of all students at a given grade or age. Practical realities and fiscal constraints, however, have always excluded at least some small percentage of students from the determination of NAEP results. For example, in its most recent assessments the small percentage of students who receive home schooling, who attend ungraded schools, who attend special schools for the deaf and blind, or who are incarcerated were not included in NAEP samples because of the logistical challenges and costs associated with identifying and assessing such students.

When reporting on the educational achievement of students in a particular grade, NAEP attempts to include all students who are enrolled in that grade at the sampled schools. NAEP samples include students with disabilities (including, students who have Individualized Education Programs (IEPs) or who are receiving special services as a result of section 504 of the Rehabilitation Act) and limited English proficient (LEP) students in approximately the same percentages in which they are found in the general school population. Although NAEP has traditionally included a substantial percentage of these students in its assessment results, the program has always recognized that a subset of a given school's students may not be able to participate in the assessment.

In the past, schools have been allowed to exclude students from NAEP for a number of reasons. Some students, such as those with significant cognitive disabilities, might not, as part of their normal educational program, have participated in any large-scale standardized assessments if their teachers judged them to be incapable of such participation. Other students
might have been incapable of taking assessments such as NAEP in English. And some students might not have participated because NAEP was unable to provide the accommodations or adaptations that would have made their inclusion possible.

To facilitate the consistent implementation of the program's policies, NAEP has provided specific criteria that staff from the sampled schools (typically the team responsible for the student's IEP or the school staff person most knowledgeable about each student) can use to determine those students who should be included in the assessment. By using these criteria, considerable numbers of students with disabilities or LEP students have been assessed. For example, NAEP 1994 results indicate that nearly 13 percent of the nation's fourth graders, 10 percent of the eighth graders, and 8 percent of twelfth graders are classified as students with disabilities or LEP students. More than half of the students with disabilities and LEP students sampled for NAEP ( 59 percent at fourth grade, 56 percent at eighth grade, and 58 percent at twelfth grade) were assessed as part of the NAEP 1994 assessment. However, the remaining 41 to 44 percent were not assessed. ${ }^{22}$

In recent years, a number of policy, legislative civil rights, and technical considerations have caused NAEP to look more closely at its administration and assessment procedures and to consider changes that can increase participation among students with disabilities or LEP students. ${ }^{23}$ Based on previous studies ${ }^{24,25}$ as well as recommendations from various offices in the U.S. Department of Education, program procedures have been modified to increase participation among students with disabilities and LEP students. Modifications were made in two areas. ${ }^{26}$ First, inclusion criteria were revised with the intention of making them clearer, more inclusive, and more likely to be applied consistently across jurisdictions participating in the state assessment program. Second, a variety of assessment accommodations and adaptations was offered to students with disabilities whose IEPs specified such accommodations for testing or LEP students who were, in the opinion of their instructors, unable to take the assessment in English.

However, several important technical issues needed to be solved before the procedural modifications could be implemented as official NAEP policy. One issue is the effect of procedural modifications on NAEP's capacity to provide accurate comparisons over time. One of the NAEP's goals is to report on trends in academic achievement. Accurately reporting changes requires keeping assessment procedures and instrumentation comparable during the period over which measurement is sought. Modifying inclusion criteria and providing accommodations can significantly expand the number of students with disabilities and LEP

[^19]students included in NAEP assessments. Although this expansion is desirable, it can cloud the interpretation of changes in achievement over time, since assessments conducted using revised procedures might include results for students that would not have been included in previous assessments.

Another issue is the validity of results from nonstandard administrations (i.e., administrations in which accommodations were allowed) and their comparability to results obtained under standard conditions. Specifically, data obtained under nonstandard conditions may not be able to be summarized and reported in terms of the same NAEP scale used for results obtained under standard conditions. That is, do scale score results obtained under nonstandard conditions convey the same information about educational achievement as corresponding results obtained under standard conditions?

The 1996 national and state mathematics assessments included supplemental samples of schools and students to allow research into inclusion, accommodation, and score validity issues, and to provide a bridge to future mathematics assessments in which revised inclusion criteria and the provision of accommodations are standard program practice.

Preliminary answers to several important research questions have been obtained.

- The introduction of the revised inclusion criteria, without the provision of accommodations, had little effect on the percentage of the total population that was assessed in NAEP at either the national or state level.
- Likewise, the introduction of the revised inclusion criteria, without the provision of accommodations, had, at most, a limited effect on the percentage of students with disabilities or LEP students who were assessed in NAEP at either the national or state level.
- The provision of accommodations and adaptations clearly increased participation rates for students with disabilities and LEP students at grades 4 and 8 . When accommodation or adaptations were available, more than 70 percent of both of these groups were assessed at each of these two grades. These numbers are substantially higher than the program has achieved in past assessments that did not offer accommodations and adaptations. On the other hand, providing accommodations at grade 12 had little effect.
- A portion of the population of students with disabilities was assessed with accommodations or adaptations when these were available but was assessed under standard conditions when special administration procedures were not available. A similar pattern of results was not evident among LEP students. The potential impact on trend measurement of this "switching" phenomenon is a topic for expanded analysis and discussion in future NAEP reports.

Though providing useful information, the analyses discussed in this chapter are only the first step in what is an ongoing research and development effort. Whether or not changes in inclusion and administration procedures affected overall scale score results is a topic for expanded analysis and discussion in future NAEP reports. A comprehensive research report on this and other inclusion issues will be published later in 1997.

## The Three NAEP 1996 Mathenaties Samples

The design of the NAEP 1996 mathematics assessment required three distinct national samples of schools and two distinct samples of schools within each jurisdiction that participated in the state assessment program. In the first of these school samples (denoted S1), the assessment was conducted using the same inclusion criteria used during the 1990 and 1992 NAEP assessments in mathematics. In the second school samples (denoted S2), revised inclusion criteria were used. No assessment accommodations or adaptations were offered to students in S1 or S2 schools. Samples of each type were identified at all three grades in the national assessment and at grades 4 and 8 for jurisdictions participating in the state assessment.

In the third sample (denoted S3), the assessment was conducted using inclusion criteria that were effectively identical to those used in S 2 schools. The S 3 sample was distinguished, however, by the availability of a variety of assessment accommodations and adaptations. To ensure sufficient amounts of data for planned analyses, students with disabilities and LEP students were oversampled in national S2 and S3 schools and all students in S3 that received an accommodation at a given grade were administered the same NAEP assessment booklet. Because of concerns about feasibility and an interest in managing the burden on participating jurisdictions, separate S 3 samples were not obtained for the state assessment.

Data from S1 and a portion of S2 (students without IEPs or equivalent plans) were combined and analyzed as the reporting sample appropriate for national and state comparisons to previous NAEP mathematics assessments. The results in Chapters 1 through 3 of this report are based on this data set. By comparing results obtained from S1 to those from S2, the NAEP program will be able to assess the effects of changing inclusion criteria on inclusion rates and assessment results. Similarly, by comparing results obtained from S2 to those from S3, the program will be able to assess the effects of providing accommodations and adaptations. Finally, by comparing results from S1 and S3, the program will be able to assess the effects of jointly changing the inclusion criteria and providing accommodations and adaptations.

## National and State Percentages of Students with Disabilities and LEP Students

As part of its standard data collection procedures, NAEP records whether or not each student in the sample has a disability or is LEP. Prior to the assessment, the NAEP school coordinator, a staff member designated by the school as the NAEP liason, is presented with a list of sampled students and, in consultation with appropriate school staff, is asked to identify students with disabilities or students classified by the school as LEP. Table 4.1 presents the percentages of the national NAEP population at each grade identified as students with disabilities, LEP students, or both.

Eleven percent of the nation's fourth grade students, 9 percent of the nation's eighth grade students, and 5 percent of the nation's twelfth grade students are identified as students with disabilities (i.e., combining "SD Only" and "Both SD and LEP"). Five percent of the nation's fourth graders, and 2 percent of the nation's eighth and twelfth graders are identified as LEP students (i.e., combining "LEP Only" and "Both SD and LEP").

Analogous results for grade 4 and grade 8 public schools are provided in Appendix D (Tables D.1 and D.2) for the nation and for each of the jurisdictions participating in the state assessment. ${ }^{27}$ The results indicate substantial variation across states and jurisdictions in the percentages of students with disabilities and LEP students. (See Appendix D for further discussion.)

| Table 4.1 | Percentage of National Population Identified as SD, LEP, or Both Grades 4, 8, and 12, Public and Nompublic Schools |  |
| :---: | :---: | :---: |


|  | Total | SD Only | Both SD and LEP | LEP Only |
| :--- | :---: | :---: | :---: | :---: |
| Grade 4 | 15 | 10 | 1 | 4 |
| Grade 8 | 11 | 9 | 0 | 2 |
| Frade 12 | 8 | 5 | 0 | 2 |

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^20]
## Effect of Inclusion Criteria and Provision of Accommodation on the Participation Rates

Revised inclusion criteria for NAEP were implemented on an experimental basis in the S 2 and S3 samples for the 1996 assessment. The revision had four goals:

1. increase inclusion rates for students with disabilities
2. bring NAEP inclusion rules for LEP students more in line with those used in state testing programs
3. allow for more consistent inclusion decisions across states and jurisdictions
4. ensure that inclusion decisions were related to the subject-matter instruction given to the student rather than less relevant considerations

Original inclusion criteria (used in S1) provided a basis for determining whether students could be excluded from the assessment. Based on the S1 criteria (i.e., the criteria used in NAEP's mathematics assessments in 1990 and 1992), students with disabilities could be excluded only if they were mainstreamed in academic subjects less than 50 percent of the time AND/OR judged to be incapable of participating meaningfully in the assessment. LEP students could be excluded if they were native speakers of a language other than English AND enrolled in an English speaking school for less than two years AND judged to be incapable of taking part in the assessment.

The guidelines used in S 2 were revised to emphasize criteria for the inclusion rather than exclusion of students with disabilities and LEP students. Although the original criteria did instruct school staff, when in doubt, to include students, the revised criteria were designed to communicate more clearly a presumption of inclusion except under special circumstances. Students with IEPs were to be included in the NAEP assessment except in the following cases:

1. The school's IEP team determined that the student could not participate, OR,
2. The student's cognitive functioning was so severely impaired that she or he could not participate, OR,
3. The student's IEP required that the student had to be tested with an accommodation or adaptation and that the student could not demonstrate his or her knowledge without that accommodation.

Under the revised criteria, all LEP students receiving academic instruction in English for three years or more were to be included in the assessment. Those LEP students receiving instruction in English for less than three years were to be included unless school staff judged them as being incapable of participating in the assessment in English.

In the S3 sample, the revised criteria were used and various accommodations and adaptations were made available. NAEP attempted to assess students with disabilities under conditions identical to those under which they normally participate in large-scale assessments. To the extent possible, NAEP offered S3 students the assessment accommodations that were specified in their IEP or equivalent document. For example, if a student's IEP specified that he
or she could only be assessed with extended assessment time, NAEP provided this accommodation. Thus, students whose IEPs required accommodations or adaptations were included in NAEP if the program was able to offer their accommodation.

An array of assessment accommodations were permitted. In general, most accommodations that schools routinely provided for their own testing were allowed in S3. These permitted accommodations included:

- One-on-one testing
- Small group testing
- Extended time
- Oral reading of directions
- Signing of directions
- Use of magnifying equipment
- Use of an amanuensis

NAEP also developed a Braille-version of the mathematics instrument at grade 8 and a largeprint version at grades 4 and 8 . These modified-format booklets were made available to students who normally would have been assessed using Braille or large-print materials.

It should be noted that students assessed under one of the special conditions typically received some combination of accommodations and adaptations. For example, students assessed in small groups (as opposed to standard NAEP sessions of roughly 30 students) usually received extended time and had directions and/or assessment questions read aloud as needed. In one-on-one administrations students often received assistance in recording answers, had directions and questions read aloud, and were afforded extra time.

NAEP goals and plans regarding LEP students were somewhat different. As with students with disabilities, the new inclusion criteria emphasized inclusion rather than exclusion and LEP students were eligible for any of the accommodations previously listed. However, field test experience had suggested that many LEP students did not have IEPs that specified assessment accommodations. Because the majority of these students are native Spanish speakers, a translation of the instrument seemed to offer an opportunity to include many students who had been excluded in the past. Therefore, in addition to the accommodations listed above, LEP students at grades 4 and 8 were offered a bilingual version of the assessment which displayed Spanish and English versions of questions on facing pages. In S3, this version was administered to LEP students whose teachers believed that the student could only participate in NAEP if given this version or that the student could best show his or her mathematical abilities working with this instrument. Students who took this booklet were typically assessed in a small-group setting and given extra time.

Table 4.2 presents the national percentages excluded from the NAEP 1996 mathematics assessment for the S1, S2 and S3 samples. Exclusion percentages for the S1 and S2 samples for public school students at the national level, and state-by-state are presented in Appendix D (Table D.2). Overall, comparisons of exclusion percentages in S1 and S2 indicate that the revised inclusion criteria, without the provision of accommodations, had little effect on the percentage of the population assessed in NAEP at either the national or state level.
Percentage of National Population Excluded Table 4.2 From the Assessment, Grades 4, 8, and 12, Public and Nonpublic Schools


| S1: Using <br> Original <br> Inclusion <br> Criteria | S2: Using <br> Revised <br> Inclusion <br> Criteria | S3: Using Revised <br> Criteria and <br> Providing <br> Accomodations/ <br> Adaptions |  |
| :--- | :---: | :---: | :---: |
| Fracte 4 | 6 | $8^{*}$ | $4 \dagger$ |
| Grade 8 | 4 | 4 | 3 |
| Grocule 12 | 3 | 3 | 3 |

* Indicates a significant difference between S1 and S2 results.
$\dagger$ Indicates a significant difference between S2 and S3 results.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Although in one instance a difference was found in the national data at grade 4, a corroborating pattern of findings was not evident in the state assessment results.

As shown in Table 4.2, the national exclusion rates at grade 4 were 2 percentage points higher in S2 than in S1. However, the grade 4 state assessment results do not corroborate this finding. Observed state-level exclusion percentages were not consistently lower in one or the other sample types and differences between the samples in exclusion percentages were not statistically significant for any of the jurisdictions. For grades 8 and 12, the national exclusion percentages are nearly identical for the S 1 and S 2 samples and do not differ significantly. At grade 8 , the state public school results are in agreement in showing little evidence of an effect.

As noted earlier, comparisons of S3 national results with those obtained in S1 and S2 help to assess the effects of providing accommodations. As shown in Table 4.2, in grade 4 using the revised inclusion criteria in conjunction with the provision of accommodations resulted in lower exclusion rates than those obtained using only the revised criteria. In S2, where accommodations were not available, 8 percent of the population was excluded. In S 3 , where the same inclusion criteria were used but accommodations were provided, a smaller percentage of the population (4 percent) was excluded. However, jointly using the revised inclusion criteria and providing accommodations resulted in an exclusion rate that did not differ significantly from those obtained using the original criteria in the absence of accommodations (i.e., sample S1). At grades 8 and 12, providing accommodations and adaptations, in combination with or in addition to the introduction of revised inclusion criteria, had little effect on exclusion percentages. Differences between S1, S 2 and S 3 exclusion rates were quite small and only the 1 percentage point difference between S 2 and S 3 exclusion rates at grade 8 was statistically significant.

At the national level, and in many of the jurisdictions that participated in the state assessment, students with disabilities and LEP students constituted a relatively modest percentage of the total school population. Because the effects of the inclusion criteria and the provision of accommodations and adaptations were confined to these groups, examining exclusion rates among the total population may not provide a sufficiently sensitive measure of their effects. Examining inclusion rates among students with disabilities and LEP students provides a more in-depth analyses and affords a potentially different perspective on the procedural changes.

Table 4.3 contains national percentages of students with disabilities assessed under standard conditions, and with the provision of adaptations or accommodations, as well as the total percentages of students with disabilities that were assessed. Appendix D (Tables D. 3 and D.4) contains the analogous results for the state assessment. At grade 4, the observed percentage of students with disabilities who were assessed in $S 2$ was 11 points lower than the corresponding percentage in Sl. Although this observed difference is consistent with the results on exclusion rates, it is not statistically significant. State results for grade 4 show no consistent pattern of increased inclusion and none of the differences between S1 and S2 inclusion percentages are statistically significant. National results for grades 8 and 12 indicate smaller observed differences which also do not differ significantly. State results for grade 8 are again consistent with national results. Table 4.3 also presents national results on inclusion percentages for LEP students. There were no significant differences between S1 and S2 LEP inclusion percentages at the national level for any of the three grades, again suggesting that revisions to the inclusion criteria had little, if any, impact on the percentage of LEP students that were assessed. At the state level, only very limited evidence to the contrary can be found. ${ }^{28}$

[^21]

## Grade 4



* Indicates a significant difference between S1 and S2 results.
$\dagger$ Indicates a significant difference between S2 and S3 results. \# Indicates a significant difference between S 1 and S 3 results.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Although changes to the inclusion criteria did not have any significant effect, the provision of accommodations and adaptations did increase grade 4 and grade 8 participation rates for students with disabilities and for LEP students. More than 70 percent of students with disabilities and LEP students were assessed in the S3 samples. At these two grade levels, the S3 rates of inclusion for students with disabilities and LEP students were significantly higher than those observed in S2. For students with disabilities at these two grades, S3 inclusion rates were also higher than those obtained in Sl. For LEP students, observed inclusion percentage were 15 to 18 percent higher in S3 than in S1 but these differences are not statistically significant, due at least partly to the relatively small numbers of these students in each of the samples. In contrast, grade 12 results do not provide a clear picture on the effects of providing accommodations. The pattern of observed inclusion rates are consistent with those evident at grades 4 and 8 in that higher percentages were obtained in S3 than in the other two samples. However, the differences across sample types at grade 12 were, for the most part, smaller than those evident at the other two grades and were not statistically significant.

As discussed above, although expanded inclusion for students with disabilities and LEP students is desirable, it presents challenges regarding the measurement of trends. Changes in overall rates of exclusion present one such challenge. The overall exclusion rate data presented in Table 4.2 suggest that such changes are small and perhaps can be ignored when measuring trends. This issue will be analyzed and discussed in greater detail in forthcoming NAEP reports. However, additional challenges to trend measurement are associated with the availability of accommodations and adaptations. In any population of students with disabilities or LEP students, some students may be capable of taking the assessment under standard conditions, but they may do somewhat better or be more comfortable with an accommodation or adaptation. Results obtained with accommodations may be more valid, particularly from the perspective of the individual student. However, assessing such students without the benefit of accommodations or adaptation in one assessment and providing such accommodations in a later assessment year can complicate the interpretation of trend results.

Results in Table 4.3 suggest that there is a portion of the students with disabilities population that is assessed with accommodations or adaptations when possible but are assessed under standard conditions when special administration procedures are not available. At all three grades, the percentage of students with disabilities who were assessed without accommodations or adaptations was lower in S3 than in S2. For example, the percentages were 12 percent lower in S 3 than in S 2 at grades 4 and 8 and 16 percent lower at grade 12. A comparison of the results for S 3 and S 1 reveal a similar pattern of observed differences. However, only the grade 4 result is statistically significant.

This same phenomenon was not evident among LEP students. There is no consistent pattern of results indicating that fewer LEP students are assessed under standard conditions when accommodations or adaptations are present. Furthermore, at all three grades, the percentages of students in S 3 who were assessed without accommodations and adaptations did not differ from those in S 1 and S2. The potential effect on trend measurement of this "switching" phenomenon in the students with disabilities population, and its absence among the LEP population, are additional topics for expanded analysis and discussion in future NAEP reports.

## Concluding Comments

Increasing the numbers of students with disabilities and LEP students who meaningfully participate in the NAEP assessment remains an important program goal. To the extent possible, NAEP results should represent the performance of all students. Greater inclusiveness in a nationally visible program like NAEP emphasizes that all students, including those with special needs, are entitled to a quality education and that we, as a nation, care about the educational achievement of all our students. The NAEP program benefits from greater inclusiveness in other ways. Other things being equal, greater inclusiveness improves NAEP's validity because achievement comparisons across assessment years, or across jurisdictions participating in the state assessment, can be made with greater confidence. However, increasing the participation of students with disabilities and LEP students must be accomplished in a way that does not jeopardize the program's ability to meet another important goal-the measurement of educational progress over time. The results described in this chapter were made possible by embedding within the NAEP 1996 assessment an experimental design that allowed the program to accomplish three objectives: (1) maintain comparability of results with previous mathematics assessments, (2) study the impacts of proposed procedural changes on important program results, such as inclusion rates and estimates of achievement, and (3) provide a bridge to future assessments in which the proposed procedural changes have become standard NAEP policy.

Although they provide useful information, the analyses discussed in this chapter are only the first step in an ongoing research and development effort. Additional questions remain about the validity of results when accommodations or adaptations are used and about their comparability to results obtained under standard conditions. The impact of providing accommodations or adaptations on NAEP estimates of scale score and achievement level distributions, for the total population and for some of NAEP's traditional reporting subgroups (e.g., Black and Hispanic students), is another issue that requires further study. In-depth analyses of the data gathered with NAEP's SD/LEP Questionnaires can provide more detailed information about the nature and extent of student disabilities, the exposure of these students to appropriate grade-level curriculum, the assessment practices that schools use with these students, and the nature of the students excluded from NAEP assessments. Analyses pertinent to these and other research issues will be included in future NAEP reports.

## Appendix A

## Duervien of Procedures Used fior the <br> NAEP 1996 Mathenutics Assessment

## Introduction

Conducting a large-scale assessment such as the National Assessment of Educational Progress (NAEP) entails the successful coordination of numerous projects, committees, procedures, and tasks. This appendix provides an overview of the NAEP 1996 mathematics assessment's primary components - framework, development, administration, scoring, and analysis. A more extensive review of the procedures and methods used in the mathematics assessment will be included in two subsequent technical reports - NAEP 1996 Technical Report and Technical Report of the NAEP 1996 State Assessment Program in Mathematics.

## The NAEP 1996 Mathematics Assessment

The 1996 assessment was the first update of the NAEP mathematics assessment framework ${ }^{1}$ since the release of the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics. ${ }^{2}$ This update reflected refinements in the specifications governing the development of the 1996 assessment while ensuring comparability of results across the 1990, 1992, and 1996 assessments. The refinements that distinguish the framework of the assessment conducted in 1996 from the framework of the assessments conducted in 1990 and 1992 include the following:

[^22]- moving away from the rigid content-strand-by-cognitive-process matrix that governed the development of earlier assessments. Classifying specific questions into cells of a matrix required those questions to measure a unique content strand at a unique cognitive level. This stipulation often decontextualized the questions and limited the possibility of assessing students' abilities to reason in rich problem-solving situations and to make connections among content strands within mathematics.
- allowing individual questions on the assessment to be classified in one or more content strands when appropriate. Knowledge or skills from more than one content strand is often needed to answer a question. The option to classify questions in multiple ways provides a greater opportunity to measure student ability in content settings that closely approximate real-world reasoning and problem-solving situations. (However, to develop content strands scales, the primary content classification was used for questions with multiple classifications.)
- including the mathematics ability categories (conceptual understanding, procedural knowledge, and problem solving) as well as the process goals from the NCTM Standards (i.e., communication and connections) to achieve a balance of questions that measured a range of cognitive outcomes.
- continuing the move towards including more constructed-response questions.
- creating "families" of questions that probe a student's understanding of mathematics vertically within a content strand or horizontally across content strands.
- revising the number sense, properties, and operations and geometry and spatial sense content strands to reflect the NCTM Standards emphasis on developing and assessing students' abilities to make sense of both number and operation and spatial settings.
These refinements to the NAEP mathematics framework were made so that the 1996 assessment would: (1) more adequately reflect recent curricular emphases and objectives and yet (2) maintain a connection with the 1990 and 1992 assessments to measure trends in student performance. Prior to the 1996 assessment, investigations were conducted to ensure that results from the assessment could be reported on the existing NAEP mathematics scale. The conclusion drawn from these investigations was that results from the 1990, 1992, and 1996 assessments could be reported on a common scale and trends in mathematics performance since 1990 examined. Figure A. 1 describes the five mathematics content strands that constitute the NAEP assessment.


## Figure A. 1

 Descriptions of the Five NAEP Mathematics Content StrandsData Analysis, Statistics, and Probability
This content strand emphasizes the appropriate methods for gathering data, the visual exploration of data, various ways of representing data, and the development and evaluation of arguments based on data analysis. At grade 4, students are asked to apply their understanding of numbers and quantities by solving problems that involve data. Fourth graders are asked to interact with a variety of graphs, to make predictions from data and explain their reasoning, to deal informally with measures of central tendency, and to use the basic concepts of chance in meaningful contexts. At grade 8, students are asked to analyze statistical claims and to design experiments, and they are asked to use simulations to model real-world situations. This strand focuses on eighth graders' basic understanding of sampling, their ability to make predictions based on experiments or data, and their ability to use some formal terminology related to probability, data analysis, and statistics. At grade 12, the strand focuses on the ability to apply the concepts of probability and to use formulas and more formal terminology to describe a variety of situations. For twelfth graders, the strand also emphasizes a basic understanding of how to use mathematical equations and graphs to interpret data.

## Algebra and Functions

This content strand extends from work with simple patterns at grade 4 to basic algebra concepts at grade 8 to sophisticated analysis at grade 12. It involves not only algebra but also precalculus and some topics from discrete mathematics. Students were expected to use algebraic notation and thinking in meaningful contexts to solve mathematical and real-world problems, specifically addressing an increasing understanding of the use of functions (including algebraic and geometric) as a representational tool. The grade 4 assessment involved informal demonstration of students' abilities to generalize from patterns, including the justification of their generalizations. Students were expected to translate between mathematical representations, to use simple equations, and to do basic graphing. At grade 8, the assessment included more algebraic notation, stressing the meaning of variables and an informal understanding of the use of symbolic representations in problem-solving contexts. Students were asked to use variables to represent a rule underlying a pattern. Eighth graders were asked to demonstrate a beginning understanding of equations and functions and the ability to solve simple equations and inequalities. By grade 12, students were asked about basic algebraic notation and terminology as they relate to representations of mathematical and real-world situations. Twelfth graders were asked to use functions as a way of representing and describing relationships.

## Figure A. 1 <br> (continued) Descriptions of the Five NAEP Mathematics Content Strands

## Number Sense, Properties, and Operations

This content strand focuses on students' understanding of numbers (whole numbers, fractions, decimals, integers, real numbers, and complex numbers), operations, and estimation and their application to real-world situations. At grade 4, this strand emphasizes the development of number sense through connecting various models to their numerical representations and an understanding of the meaning of addition, subtraction, multiplication, and division. At grade 8, number sense is extended to include positive and negative numbers, and the strand addresses properties and operations involving whole numbers, fractions, decimals, integers, and rational numbers. At grade 12, this strand includes real and complex numbers and allows students to demonstrate competency up to the precalculus or calculus level.

## Measurement

This content strand focuses on an understanding of the process of measurement and the use of numbers and measures to describe and compare mathematical and realworld objects. Students are asked to identify attributes, select appropriate units and tools, apply measurement concepts, and communicate measurement-related ideas. At grade 4, the strand focuses on time, money, temperature, length, perimeter, area, capacity, weight/mass, and angle measure. At grades 8 and 12, the strand includes these measurement concepts, but, the focus shifts to more complex measurement problems that involve volume or surface area or that require students to combine shapes and to translate and apply measures. Eighth- and twelfth-grade students also solve problems involving proportional thinking (such as scale drawing or map reading) and do applications that involve the use of complex measurement formulas.

## Geometry and Spatial Sense

This content strand is designed to extend beyond low-level identification of geometric shapes to include transformations and combinations of those shapes. Informal constructions and demonstrations (including drawing representations) along with their justifications, take precedence over more traditional types of compass-andstraightedge constructions and proofs. At grade 4, students are asked to model properties of shapes under simple combinations and transformations, and they are asked to use mathematical communication skills to draw figures from verbal descriptions. At grade 8, students are asked to expand their understanding to include properties of angles and polygons. They are also asked to apply reasoning skills to make and validate conjectures about transformations and combinations of shapes. At grade 12, students are asked to demonstrate an understanding of transformational geometry and to apply concepts of proportional thinking to various geometric situations.

## The Assessment Design

Students participating in the assessment received a booklet containing a set of general background questions, a set of subject-specific background questions, and a combination of cognitive questions grouped in sets called blocks. At each grade level, the blocks of questions consisted of multiple-choice and constructed-response questions. Two types of constructedresponse questions were included - short and extended constructed response. Short constructed-response questions required students to provide answers to computation problems or to describe solutions in one or two sentences. Extended constructed-response questions required students to provide longer answers (e.g., a description of possibilities, a more involved computational analysis, or a description of a pattern and its implications). Students were expected to adequately answer the short constructed-response questions in 2 to 3 minutes and the extended constructed-response questions in approximately 5 minutes. Short constructedresponse questions which first appeared in the assessment in 1996 were graded to allow for partial credit (i.e., giving student credit for answers that are partially correct) according to a unique scoring rubric developed for each constructed-response question. Short constructedresponse questions included in the 1990 and 1992 mathematics assessments were dichotomously scored (i.e., correct or incorrect). The extended constructed-response questions included in the 1992 and 1996 assessments were scored allowing for partial credit.

The blocks of questions contained several other features. Five to seven of the blocks at each grade level allowed for the use of calculators. At grade 4, students were provided fourfunction calculators, and at grades 8 and 12 , students were provided scientific calculators. Prior to the assessment, all students were trained to use these calculators. For several blocks, students were given manipulatives (including geometric shapes, three dimensional models, and spinners). For two of the blocks at each grade level, students were given rulers at grade 4 and rulers and protractors at grades 8 and 12 so they could answer questions dealing with measurements and draw specified geometric shapes.

As part of the national assessment, other blocks of questions were developed for each of the grade levels. Each grade level had two estimation blocks that employed a paced-audiotape format to measure students' estimation skills. Each grade level also had two 30-minute theme blocks that contained a mixture of multiple-choice and constructed-response questions. All the questions in these theme blocks related to some aspect of a rich problem setting that served as a unifying theme for the entire block. Neither the estimation nor the theme block components were included in the state assessment. Results for the estimation and theme blocks are not included in this report but will be featured in future reports on the NAEP 1996 mathematics assessment.

Of the 17 blocks in the national sample at grade 4 and 19 blocks in the national sample at grades 8 and 12, 3 were carried forward from the 1990 assessment and 5 were carried forward from the 1992 assessment to allow for the measurement of trends across time. The remaining blocks of questions at each grade level contained new questions that were developed for the 1996 assessment as specified by the updated framework.

The data in Table A. 1 reflect the number of questions by type and grade level for the 1990, 1992, and 1996 assessments. As mentioned earlier, the 1996 assessment continued the shift toward more constructed-response questions, including extended constructed-response questions that required students to provide an answer and a corresponding explanation.

Each booklet also included three sets of student background questions. The first set consisting of general background questions included questions about race or ethnicity, mother's and father's level of education, reading materials in the home, homework, attendance, and academic expectations. The second set consisting of mathematics background questions included questions about instructional activities, courses taken, use of specialized resources such as calculators in mathematics classes, and views on the utility and value of the subject. (Students were given 5 minutes to complete each set of questions, with the exception of the fourth graders, who were given more time because the general background questions were read aloud to them.) The third set of questions followed the cognitive question blocks and contained five questions about students' motivation to do well on the assessment, their perception of the difficulty of the assessment, and their familiarity with the types of cognitive questions included.

| Table A. 1 | Distribution of Questions by Question Type |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | GRADE 4 |  |  | GRADE 8 |  |  | GRADE 12 |  |  |
|  | 1990 | 1992 | 1996 | 1990 | 1992 | 1996 | 1990 | 1992 | 1996 |
| Multiple-Choice | 102 | 99 | 81 | 149 | 118 | 102 | 156 | 115 | 99 |
| Short Constructed-Response * | 41 | 59 | 64 | 42 | 65 | 69 | 47 | 64 | 74 |
| Extended Constructed-Response ** | - | 5 | 13 | - | 6 | 12 | - | 6 | 11 |
| Total | 143 | 163 | 158 | 191 | 189 | 183 | 203 | 185 | 184 |

* Short constructed-response questions included in the 1990 and 1992 assessments were scored dichotomously. New short constructed-response questions included in the 1996 assessment were scored to allow for partial credit. ** No extended constructed-response questions were included in the 1990 assessment.

The booklets were carefully balanced to accommodate time requirements for the question types in each block, using information gathered from field testing. Information on the design of the assessment is presented in the forthcoming NAEP 1996 Technical Report.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment-a mathematics teacher questionnaire, a school characteristics and policy questionnaire, and an SD/LEP student questionnaire.

The teacher questionnaire was administered to the mathematics teacher of the fourthand eighth-grade students participating in the assessment. The questionnaire consisted of three sections and took approximately 20 minutes to complete. The first section focused on the teacher's general background and experience; the second section focused on the teacher's background related to mathematics; and the third section focused on classroom information about mathematics instruction.

The school characteristics and policy questionnaire was given to the principal or other administrator in each participating school and took about 20 minutes to complete. The questions asked about the principal's background and experience, school policies, programs, facilities, and the demographic composition and background of the students and teachers.

The SD/LEP student questionnaire was completed by a school staff member knowledgeable about those students who were selected to participate in the assessment and who were identified as (1) having an Individualized Education Plan (IEP) or equivalent plan (for reasons other than being gifted or talented) or (2) having limited English proficiency (LEP). A SD/LEP student questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each questionnaire took approximately 3 minutes to complete and asked about the student and the special programs in which he or she participated.

## National and State Samples

The national and regional results presented in this report are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. The samples were selected using a complex multistage sampling design that involved sampling students from selected schools within selected geographic areas across the country. The sample design had the following stages:

1. selection of geographic areas (a county, group of counties, or metropolitan statistical area)
2. selection of schools (public and nonpublic) within the selected areas
3. selection of students within selected schools

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences between the student samples and the respective populations from which they were drawn. Sampling weights account for disproportionate representation due to the oversampling of students who attend schools with high concentrations of Black and/or Hispanic students and who attend nonpublic schools. Sampling weights also account for lower sampling rates for very small schools.

Table A. 2 provides a summary of the weighted and unweighted student sample sizes for the national mathematics assessment. The numbers reported include public and nonpublic school students.


|  | Number of Schools | Unweighted Student <br> Sample Size | Weighted Student <br> Sample Size |
| :---: | :---: | :---: | :---: |
| Grade 4 | 281 | 6,627 | $3,714,998$ |
| Grade 8 | 261 | 7,146 | $3,570,116$ |
| Grade 12 | 264 | 6,904 | $2,830,443$ |

[^23]The results of the 1996 state assessment program in mathematics provided in this report are based on state-level samples of fourth- and eighth-grade students. The samples of both public and nonpublic school fourth- and eighth-grade students were selected based on a twostage sample design that entailed selecting schools within participating jurisdictions and selecting students within schools. The first-stage samples of schools were selected with probability proportional to the fourth- or eighth-grade enrollment in those schools. Special procedures were used for jurisdictions that have many small schools and for jurisdictions that have a small number of schools.

As with the national samples, the jurisdiction samples were weighted to allow for valid inferences about the populations of interest. Tables A.3a through A.3d contain the unweighted number of participating schools and students as well as weighted school and student participation rates. Two weighted school participation rates are provided for each jurisdiction. The first rate is the weighted percentage of schools participating in the assessment before substitution. This rate is based only on the number of schools that were initially selected for the assessment. The numerator of this rate is the sum of the number of students represented by each initially selected school that participated in the assessment. The denominator is the sum of the number of students represented by each of the initially selected schools that had eligible students enrolled. This rate included both participating and nonparticipating schools.

The second school participation rate is the weighted participation rate after substitution. The numerator of this rate is the sum of the number of students represented by each of the participating schools, whether originally selected or substituted. The denominator is the same as that for the weighted participation rate for the initial sample. This statement means that for a given jurisdiction, the weighted participation rate after substitution is at least as great as the weighted participation rate before substitution.

Also presented in Table A.3a through A3d are the weighted percentages of students who participated after make-up sessions were completed. This rate reflects the percentage of the eligible student population from participating schools within the jurisdiction, and this percentage represents the students who participated in the assessment in either an initial session or a make-up session. The numerator of this rate is the sum, across all assessed students, of the number of students that each selected student who was eligible to participate represents, including students who did not participate.

| Table A.3a | NAEP 1996 School and Student | the wariows |
| :---: | :---: | :---: |
|  | Participation Rates by States | MR0 |
|  | Grade 4 Public Schools |  |


|  | Weighted School Participation |  | Total <br> Number of Schools Participating | Weighted Student Participation Rate | Total <br> Number of Students Assessed |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Before Substitutes | After Substitutes |  |  |  |
| Nation | 83 | 83 | 209 | 95 | 5,215 |
| Alabama | 79 | 93 | 99 | 96 | 2,541 |
| Alaska $\ddagger$ | 91 | 91 | 113 | 91 | 2,304 |
| Arizona | 87 | 87 | 91 | 95 | 2,113 |
| Arkansas $\ddagger$ | 76 | 78 | 81 | 96 | 2,047 |
| California | 80 | 94 | 99 | 94 | 2,063 |
| Colorado | 99 | 99 | 107 | 95 | 2,609 |
| Connecticut | 100 | 100 | 105 | 96 | 2,565 |
| Delaware | 100 | 100 | 51 | 94 | 1,984 |
| District of Columbia | 100 | 100 | 108 | 95 | 2,574 |
| Florida | 100 | 100 | 106 | 94 | 2,549 |
| Georgia | 98 | 98 | 103 | 95 | 2,542 |
| Hawaii | 100 | 100 | 106 | 95 | 2,578 |
| Indiana | 87 | 91 | 96 | 96 | 2,470 |
| lowa $\ddagger$ | 79 | 87 | 95 | 97 | 2,359 |
| Kentucky | 88 | 96 | 102 | 95 | 2,579 |
| Lovisiana | 100 | 100 | 108 | 95 | 2,671 |
| Maine | 87 | 87 | 97 | 94 | 2,115 |
| Maryland | 93 | 93 | 99 | 96 | 2,465 |
| Massachusetts | 97 | 97 | 103 | 95 | 2,497 |
| Michigan $\ddagger$ | 76 | 88 | 94 | 94 | 2,382 |
| Minnesota | 91 | 93 | 99 | 94 | 2,425 |
| Mississippi | 92 | 97 | 103 | 96 | 2,716 |
| Missouri | 96 | 99 | 107 | 95 | 2,643 |
| Montana $\ddagger$ | 70 | 81 | 99 | 96 | 2,251 |
| Nebraska | 100 | 100 | 132 | 95 | 2,678 |
| Nevada $\ddagger$ | 84 | 86 | 95 | 92 | 2,193 |
| New Jersey $\ddagger$ | 73 | 73 | 78 | 95 | 1,961 |
| New Mexico | 100 | 100 | 107 | 94 | 2,389 |
| New York $\ddagger$ | 73 | 86 | 90 | 94 | 2,248 |
| North Carolina | 97 | 97 | 106 | 96 | 2,658 |
| North Dakota | 75 | 96 | 120 | 96 | 2,666 |
| Oregon | 86 | 90 | 95 | 95 | 2,233 |
| Pennsylvania $\ddagger$ | 73 | 86 | 90 | 95 | 2,347 |
| Rhode Island | 89 | 99 | 104 | 95 | 2,461 |
| South Carolina $\ddagger$ | 87 | 88 | 92 | 95 | 2,364 |
| Tennessee | 94 | 94 | 98 | 96 | 2,473 |
| Texas | 95 | 97 | 104 | 96 | 2,413 |
| Utah | 100 | 100 | 106 | 95 | 2,625 |
| Vermont $\ddagger$ | 78 | 81 | 100 | 96 | 2,136 |
| Virginia | 100 | 100 | 104 | 95 | 2,586 |
| Washington | 99 | 99 | 105 | 94 | 2,640 |
| West Virginia | 100 | 100 | 109 | 95 | 2,530 |
| Wisconsin | 92 | 94 | 99 | 95 | 2,437 |
| Wyoming | 100 | 100 | 115 | 96 | 2,758 |
| DDESS | 100 | 100 | 38 | 95 | 1,313 |
| DoDDS | 100 | 100 | 93 | 94 | 2,604 |

National results are based on the national assessment samples, not on aggregated state assessment program samples.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A)
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## NAEP 1996 School and Student <br> Table A.3b

|  | Weighted School Parricipation |  | Total <br> Number of Schools Participating | Weighted Student Participation Rate | Total Number of Students Assessed |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Before <br> Substitutes | After Substitutes |  |  |  |
| Nation | 79 | 79 | 77 | 97 | 1,412 |
| Alabama $\ddagger$ | 72 | 72 | 10 | 97 | 239 |
| Alaska | - | - | - | - | - |
| Arizona $\ddagger$ | 78 | 87 | 10 | 99 | 185 |
| Arkansas | 86 | 86 | 8 | 97 | 174 |
| California $\ddagger$ | 73 | 73 | 11 | 98 | 256 |
| Colorado $\ddagger$ | 76 | 76 | 10 | 96 | 174 |
| Connecticut $\ddagger$ | 75 | 75 | 13 | 96 | 245 |
| Delaware $\ddagger$ | 41 | 43 | 13 | 95 | 337 |
| District of Columbia $\ddagger$ | 63 | 66 | 18 | 96 | 395 |
| Florida $\ddagger$ | 66 | 73 | 12 | 96 | 232 |
| Georgia | 99 | 99 | 13 | 94 | 251 |
| Hawaii | - | - | - | - | - |
| Indiana $\ddagger$ | 79 | 86 | 15 | 96 | 297 |
| lowa $\ddagger$ | 82 | 82 | 15 | 96 | 284 |
| Kentucky | 87 | 87 | 13 | 97 | 300 |
| Louisiana | 86 | 86 | 19 | 97 | 444 |
| Maine $\ddagger$ | 71 | 74 | 8 | 97 | 101 |
| Maryland $\ddagger$ | 57 | 57 | 11 | 98 | 269 |
| Massachusetts $\ddagger$ | 84 | 84 | 15 | 96 | 305 |
| Michigan | 86 | 94 | 18 | 97 | 342 |
| Minnesota $\ddagger$ | 78 | 78 | 15 | 96 | 277 |
| Mississippi $\ddagger$ | 79 | 79 | 11 | 96 | 268 |
| Missouri | 99 | 100 | 23 | 95 | 449 |
| Montana | 94 | 94 | 10 | 95 | 173 |
| Nebraska | 91 | 91 | 22 | 99 | 433 |
| Nevada | 91 | 100 | 9 | 96 | 173 |
| New Jersey $\ddagger$ | 64 | 75 | 16 | 94 | 334 |
| New Mexico | 90 | 90 | 13 | 94 | 212 |
| New York $\ddagger$ | 83 | 91 | 23 | 96 | 495 |
| North Carolina | - | - | - | - | - |
| North Dakota $\ddagger$ | 68 | 68 | 12 | 95 | 152 |
| Oregon $\ddagger$ | 34 | 34 | 4 | 96 | 69 |
| Pennsylvania $\ddagger$ | 66 | 66 | 19 | 96 | 401 |
| Rhode Island | - | - | - | - | - |
| South Carolina | - | - | - | - | - |
| Texas $\ddagger$ | 64 | 64 | 4 | 96 | 101 |
| Utah $\ddagger$ | 81 | 81 | 7 | 95 | 146 |
| Vermont $\ddagger$ | 74 | 74 | 9 | 97 | 145 |
| Virginia | - | - | - | - | - |
| Washington | - | - | - | - | - |
| West Virginia | - | - | - | - | - |
| Wisconsin $\ddagger$ | 68 | 73 | 25 | 97 | 480 |
| Wyoming | 82 | 95 | 7 | 96 | 84 |
| DDESS | - | - | - | - | - |
| DoDDS | - | - | - | - | - |
| Guam $\ddagger$ | 78 | 78 | 9 | 94 | 317 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for nonpublic school participation rates (see Appendix A).
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.


|  | Weighted School Parricipation |  | Total <br> Number of Schools Participating | Weighted Student Participation Rate | Total Number of Students Assessed |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Before Substitutes | After Substitutes |  |  |  |
| Nation | 81 | 82 | 192 | 92 | 5,590 |
| Alabama | 84 | 90 | 97 | 93 | 2,261 |
| Alaska $\ddagger$ | 92 | 92 | 53 | 80 | 1,462 |
| Arizona | 87 | 87 | 93 | 91 | 2,136 |
| Arkansas $\ddagger$ | 70 | 71 | 77 | 92 | 1,845 |
| California | 83 | 94 | 101 | 90 | 2,290 |
| Colorado | 100 | 100 | 108 | 91 | 2,530 |
| Connecticut | 100 | 100 | 102 | 91 | 2,485 |
| Delaware | 100 | 100 | 30 | 90 | 1,798 |
| District of Columbia | 100 | 100 | 32 | 85 | 1,693 |
| Florida | 100 | 100 | 104 | 91 | 2,401 |
| Georgia | 99 | 99 | 100 | 90 | 2,364 |
| Hawaii | 100 | 100 | 51 | 91 | 2,189 |
| Indiana | 88 | 91 | 96 | 93 | 2,347 |
| lowa $\ddagger$ | 74 | 84 | 93 | 93 | 2,169 |
| Kentucky | 88 | 92 | 101 | 94 | 2,461 |
| Lovisiana | 100 | 100 | 112 | 89 | 2,599 |
| Maine | 90 | 90 | 93 | 92 | 2,258 |
| Maryland $\ddagger$ | 86 | 86 | 89 | 91 | 2,137 |
| Massachusetts | 92 | 92 | 98 | 92 | 2,280 |
| Michigan $\ddagger$ | 70 | 86 | 90 | 90 | 2,155 |
| Minnesota | 86 | 88 | 96 | 92 | 2,425 |
| Mississippi | 89 | 95 | 103 | 93 | 2,487 |
| Missouri | 93 | 96 | 105 | 91 | 2,386 |
| Montana $\ddagger$ | 72 | 75 | 75 | 92 | 1,912 |
| Nebraska | 99 | 100 | 116 | 91 | 2,610 |
| Nevada $\ddagger$ | 38 | 38 | 28 | 90 | 983 |
| New Hampshire $\ddagger$ | 66 | 69 | 62 | 89 | 1,723 |
| New Jersey $\ddagger$ | 64 | 65 | 69 | 93 | 1,655 |
| New Mexico | 100 | 100 | 90 | 90 | 2,371 |
| New York $\ddagger$ | 71 | 80 | 84 | 91 | 1,962 |
| North Carolina | 100 | 100 | 107 | 91 | 2,638 |
| North Dakota | 83 | 95 | 108 | 94 | 2,602 |
| Oregon | 86 | 92 | 98 | 90 | 2,323 |
| Rhode Island | 90 | 90 | 42 | 89 | 2,055 |
| South Carolina $\ddagger$ | 86 | 87 | 91 | 89 | 2,143 |
| Tennessee | 92 | 92 | 98 | 91 | 2,300 |
| Texas | 90 | 95 | 100 | 92 | 2,245 |
| Utah | 100 | 100 | 95 | 91 | 2,697 |
| Vermont $\ddagger$ | 74 | 74 | 75 | 93 | 2,001 |
| Virginia | 100 | 100 | 106 | 91 | 2,545 |
| Washington | 94 | 95 | 103 | 90 | 2,434 |
| West Virginia | 100 | 100 | 106 | 92 | 2,578 |
| Wisconsin $\ddagger$ | 78 | 78 | 90 | 92 | 2,165 |
| Wyoming | 100 | 100 | 70 | 93 | 2,696 |
| DDESS | 100 | 100 | 12 | 95 | 620 |
| DoDDS | 100 | 100 | 57 | 94 | 2,160 |
| Guam | 100 | 100 | 6 | 86 | 928 |

[^24]

|  | Weighted School Parricipation |  | Total <br> Number of Schools Participating | Weighted Student Participation Rate | Total <br> Number of Students Assessed |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Before <br> Substitutes | After Substitutes |  |  |  |
| Nation | 81 | 81 | 78 | 97 | 1,556 |
| Alabama $\ddagger$ | 64 | 64 | 9 | 92 | 119 |
| Alaska | - | - | - | - | - |
| Arizona | 5 | - | - | - | - |
| Arkansas $\ddagger$ | 51 | 60 | 5 | 98 | 62 |
| California $\ddagger$ | 75 | 75 | 13 | 97 | 232 |
| Colorado | - | - | - | - | - |
| Connecticut $\ddagger$ | 63 | 65 | 19 | 94 | 265 |
| Delaware $\ddagger$ | 38 | 40 | 12 | 96 | 281 |
| District of Columbia $\ddagger$ | 47 | 47 | 16 | 95 | 222 |
| Florida | - | - | - | - | - |
| Georgia | 88 | 88 | 10 | 97 | 267 |
| Hawaii | - | - | - | - | - |
| Indiana | - | - | - | - | - |
| lowa | 88 | 88 | 15 | 96 | 282 |
| Kentucky $\ddagger$ | 67 | 67 | 11 | 98 | 218 |
| Louisiana $\ddagger$ | 73 | 73 | 22 | 96 | 426 |
| Maine | - | 64 | - | - | - |
| Maryland $\ddagger$ | 60 | 64 | 18 | 97 | 301 |
| Massachusetts $\ddagger$ | 70 | 74 | 18 | 95 | 301 |
| Michigan $\ddagger$ | 80 | 88 | 18 | 96 | 293 |
| Minnesota $\ddagger$ | 75 | 75 | 15 | 96 | 250 |
| Misssissippi | , | - | - | 96 | 5 |
| Missouri | 94 | 100 | 22 | 96 | 353 |
| Montana $\ddagger$ | 78 | 78 | 9 | 95 | 121 |
| Nebraska $\ddagger$ | 83 | 85 | 20 | 95 | 358 |
| Nevada $\ddagger$ | 78 | 78 | 6 | 95 | 101 |
| New Hampshire $\ddagger$ | 85 | 85 | 12 | 96 | 212 |
| New Jersey $\ddagger$ | 68 | 71 | 22 | 94 | 320 |
| New Mexico $\ddagger$ | 87 | 87 | 12 | 89 | 228 |
| New York $\ddagger$ | 88 | 90 | 30 | 95 | 539 |
| North Carolina | - | - | - | - | - |
| North Dakota | 86 | 86 | 12 | 96 | 194 |
| Oregon $\ddagger$ | 22 | 22 | 3 | 93 | 43 |
| Rhode Island $\ddagger$ | 81 | 81 | 26 | 96 | 423 |
| South Carolina $\ddagger$ | 76 | 76 | 10 | 96 | 164 |
| Tennessee Texas | - 93 | - 93 | 9 | - 92 | - 166 |
| Utah $\ddagger$ | 43 | 43 | 2 | 93 | 40 |
| Vermont $\ddagger$ | 73 | 73 | 9 | 95 | 114 |
| Virginia | - | - | - | - | - |
| Washington | 86 | 86 | 9 | 97 | 182 |
| West Virginia | - | - |  | - | - |
| Wisconsin $\ddagger$ | 68 | 73 | 28 | 94 | 362 |
| Wyoming $\ddagger$ | 74 | 74 | 5 | 97 | 51 |
| DDESS | - | - | - | - | - |
| DoDDS | - | - | - | - | - |
| Guam $\ddagger$ | 76 | 76 | 8 | 95 | 202 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for nonpublic school participation rates (see Appendix A).
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

In carrying out the 1996 state assessment program, the National Center for Education Statistics (NCES) established participation rate standards that jurisdictions were required to meet in order for their results to be reported (see notations in Tables A.3a throught A.3d). NCES also established additional standards that required the annotation of published results for jurisdictions whose sample participation rates were low enough to raise concerns about their representativeness.

No jurisdictions at grade 4 and three states at grade 8 (Nevada, New Hampshire, and New Jersey) failed to meet the initial public school participation rate of 70 percent. For these states, results for eighth-grade public school students are not reported in this or any report of NAEP 1996 mathematics findings. Several other jurisdictions whose results were published received a notation to indicate possible nonresponse bias.

The following 10 jurisdictions failed to meet the initial nonpublic school participation rate of 70 percent at grade 4: Delaware, the District of Columbia, Florida, Maryland, New Jersey, North Dakota, Oregon, Pennsylvania, Texas, and Wisconsin. Twelve jurisdictions failed to meet the initial nonpublic school participation rate of 70 percent at grade 8: Alabama, Arkansas, Connecticut, Delaware, the District of Columbia, Kentucky, Maryland, New Jersey, Oregon, Utah, Wisconsin, and Wyoming. For these jurisdictions, results for fourth- or eighthgrade nonpublic school students are not reported in this or any report of NAEP 1996 mathematics findings. As with public schools, several other jurisdictions whose nonpublic school results were published received a notation to indicate nonresponse bias.

NCES standards require weighted school participation rates before substitution of at least 85 percent to guard against potential bias due to school nonresponse. The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that declined to participate in the assessment. However, considerable technical consideration has been given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate the possibility of bias because of the nonparticipation of initially selected schools. Thus, for the weighted school participation rates that included substitute schools, the guideline was set at 90 percent. This is expressed in the following guideline:

## A jurisdiction will receive a notation if its weighted participation rate for the initial sample of schools was below 85 percent $\underline{\text { AND }}$ the weighted school participation rate after substitution was below 90 percent.

Nine states did not meet this guideline for public schools at grade 4: Arkansas, Iowa, Michigan, Montana, Nevada, New Jersey, New York, Pennsylvania, and Vermont. Fourteen jurisdictions did not meet this guideline for nonpublic schools at grade 4: Alabama, Arizona, California, Colorado, Connecticut, Guam, Indiana, Iowa, Maine, Massachusetts, Minnesota, Mississippi, Utah, and Vermont. Seven jurisdictions did not meet this guideline for public schools at grade 8: Arkansas, Iowa, Michigan, Montana, New York, Vermont, and Wisconsin. Twelve jurisdictions did not meet this guideline for nonpublic schools at grade 8: California, Guam, Louisiana, Massachusetts, Michigan, Minnesota, Montana, Nebraska, Nevada, Rhode Island, South Carolina, and Vermont.

To help ensure adequate sample representation for each jurisdiction participating in the 1996 state assessment program, NAEP provided substitutes for nonparticipating public and nonpublic schools. (When possible, a substitute school was provided for each initially selected school that declined participation.) For jurisdictions that used substitute schools, the assessment results were based on the student data from all schools participating from both the original sample and the list of substitutes (unless an initial school and its substitute eventually participated, in which case only the data from the initial school were used). For jurisdictions that did not use substitute schools, the participation rates were based on participating schools from the original sample.

The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, inadequate representation of an important segment of a jurisdiction's population is of concern, regardless of the overall participation rate. At grade 4, Alaska and South Carolina (for public schools) and New York (for nonpublic schools) failed to meet the following NCES guideline concerning strata-specific participation rates. At grade 8 , Alaska, Maryland, and South Carolina (for public schools) and New Hampshire, New Mexico, and New York (for nonpublic schools) failed to meet this NCES guideline.

> A jurisdiction that is not already receiving a notation for problematic overall school or student participation rates will receive a notation if the sampled students within participating schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than five percent of the jurisdiction's weighted assessable student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the students, whether or not the student was classified as a student with a disability (SD) or of limited English proficiency (LEP), and the type of assessment session (monitored or unmonitored). In addition, for public schools, classes of schools were determined by school level of urbanization, minority enrollment, and median household income of the area in which the school is located. For nonpublic schools, classes of schools were determined by type and location of schools.

This guideline addresses the concern that if nonparticipating schools were concentrated within a particular class of schools, the potential for substantial bias remained, even though the overall level of school participation appeared to be satisfactory. Nonresponse adjustment cells for schools were formed within each jurisdiction, and the schools within each cell were similar in terms of minority enrollment, degree of urbanization, and/or median household income for public schools, and school type and location for nonpublic schools, as appropriate for each jurisdiction. If more than 5 percent (weighted) of the sample schools (after substitution) were nonparticipants from a single adjustment cell, then the potential for nonresponse bias was too great.

In one state (Alaska), the public school student participation rate for grade 8 fell below the NCES-prescribed criteria of 85 percent. No other notations related to student participation rates appear in NAEP 1996 mathematics reports.

## Students with Disabilities (SD) and Limited English Proficient (LEP) Students

It is NAEP's intent to assess all selected students. Therefore, every effort is made to ensure that all selected students who are capable of participating in the assessment are assessed. Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. These criteria are described in Chapter 4 of this report. The results discussed in Chapters 1 through 3 are based on the national and state "reporting samples." The reporting samples used inclusion criteria equivalent to those used for the 1990 and 1992 assessments to allow for comparability of results across assessments. Sample information for the SD and LEP populations for the reporting samples are presented in Tables A.4a. through A.4d.

## Scoring

Materials from the 1996 assessment were shipped to National Computer Systems, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by the Educational Testing Service (ETS). Each constructed-response question had a unique scoring rubric that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four- or five-level rubrics, while the short constructed-response questions first appearing in the 1996 assessment were rated according to three-level rubrics that permitted partial credit. Other short constructed-response questions that appeared in previous assessments were scored as either correct or incorrect.

For the national and state mathematics assessments more than 4.8 million constructed responses were scored. This number includes rescoring to monitor inter-rater reliability and trend reliability. In other words, scoring reliability was calculated within year (1996) and across years (1990, 1992, and 1996). The overall within-year percentages of agreement for the 1996 national reliability samples were 96 percent at grade 4,96 percent at grade 8 , and 96 percent at grade 12. The percentages of agreement across the assessment years for the national interyear reliability sample were 96 percent (1990 to 1996) and 94 percent (1992 to 1996) at grade 4,95 percent (1990 to 1996) and 94 percent (1992 to 1996) at grade 8, and 95 percent (1990 to 1996) and 93 percent (1992 to 1996) at grade 12.

## Table A.4a

NAEP 1996 SD and LEP Participation Rates by States Grade 4 Public Schools


|  | Total Percentage of Students - SD and LEP |  | Percentage of Students SD |  | Percentage of Students LEP |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Identified | Excluded | Identified | Excluded | Identified | Excluded |
| Nation | 15 | 6 | 11 | 5 | 4 | 2 |
| Alabama | 11 | 6 | 11 | 6 | 0 | 0 |
| Alaska | 21 | 4 | 13 | 3 | 9 | 1 |
| Arizona | 22 | 13 | 10 | 7 | 13 | 7 |
| Arkansas | 9 | 7 | 9 | 6 | 0 | 0 |
| California | 33 | 16 | 9 | 6 | 26 | 13 |
| Colorado | 16 | 9 | 13 | 7 | 4 | 2 |
| Connecticut | 15 | 8 | 12 | 6 | 3 | 2 |
| Delaware | 14 | 7 | 11 | 5 | 2 | 2 |
| District of Columbia | 14 | 11 | 9 | 7 | 6 | 5 |
| Florida | 19 | 10 | 14 | 7 | 6 | 3 |
| Georgia | 13 | 7 | 12 | 6 | 2 | 1 |
| Hawaii | 14 | 6 | 9 | 4 | 5 | 1 |
| Indiana | 11 | 5 | 11 | 5 | 1 | 0 |
| lowa | 12 | 5 | 10 | 4 | 2 | 1 |
| Kentucky | 10 | 6 | 10 | 6 | 0 | 0 |
| Lovisiana | 13 | 7 | 13 | 7 | 1 | 0 |
| Maine | 16 | 8 | 15 | 8 | 0 | 0 |
| Maryland | 14 | 8 | 13 | 7 | 1 | 1 |
| Massachusets | 16 | 8 | 14 | 7 | 2 | 2 |
| Michigan | 12 | 7 | 10 | 6 | 2 | 1 |
| Minnesota | 13 | 6 | 10 | 4 | 3 | 1 |
| Mississippi | 7 | 5 | 7 | 5 | 0 | 0 |
| Missouri | 15 | 5 | 14 | 5 | 1 | 0 |
| Montana | 10 | 5 | 10 | 5 | 0 | 0 |
| Nebraska | 16 | 5 | 14 | 5 | 2 | 1 |
| Nevada | 15 | 8 | 10 | 6 | 6 | 3 |
| New Jersey | 10 | 6 | 9 | 5 | 2 | 1 |
| New Mexico | 22 | 12 | 14 | 8 | 10 | 5 |
| New York | 17 | 9 | 10 | 6 | 7 | 4 |
| North Carolina | 15 | 7 | 13 | 7 | 2 | 1 |
| North Dakota | 11 | 4 | 10 | 3 | 0 | 0 |
| Oregon | 20 | 9 | 13 | 6 | 7 | 3 |
| Pennsylvania | 10 | 5 | 9 | 4 | 1 | 1 |
| Rhode Island | 18 | 6 | 13 | 5 | 5 | 2 |
| South Carolina | 13 | 6 | 12 | 6 | 0 | 0 |
| Tennessee | 14 | 7 | 12 | 6 | 1 | 1 |
| Texas | 25 | 11 | 13 | 8 | 14 | 5 |
| Utah | 13 | 6 | 11 | 5 | 2 | 1 |
| Vermont | 14 | 6 | 14 | 6 | 1 | 0 |
| Virginia | 14 | 7 | 12 | 6 | 2 | 1 |
| Washington | 14 | 6 | 11 | 5 | 3 | 1 |
| West Virginia | 13 | 8 | 13 | 8 | 0 | 0 |
| Wisconsin | 12 | 8 | 11 | 8 | 2 | 1 |
| Wyoming | 12 | 4 | 12 | 4 | 1 | 0 |
| DDESS | 9 | 4 | 8 | 3 | 1 | 1 |
| DoDDS | 10 | 5 | 9 | 4 | 2 | 1 |
| Guam | 16 | 13 | 7 | 6 | 9 | 7 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
SD = Students with Disabilities (the term previously used was IEP).
LEP = Limited English Proficient student.
To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment. A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## NAEP 1996 SD and LEP Participation Rates <br> Table A.4b by States Grade 4 Nonpublic Schools

|  | Total Percentage of Students - SD and LEP |  | Percentage of Students SD |  | Percentage of Students LEP |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Identified | Excluded | Identified | Excluded | Identified | Excluded |
| Nation | 2 | 1 | 2 | 1 | 0 | 0 |
| Alabama | 0 | 0 | 0 | 0 | 0 | 0 |
| Alaska | - | - | - | - | - | - |
| Arizona | 43 | 3 | 3 | 3 | 43 | 3 |
| Arkansas | 4 | 2 | 2 | 1 | 1 | 1 |
| California | 0 | 0 | 0 | 0 | 0 | 0 |
| Colorado | 5 | 5 | 5 | 5 | 0 | 0 |
| Connecticut | 7 | 1 | 5 | 0 | 3 | 1 |
| Delaware | 1 | 1 | 1 | 1 | 0 | 0 |
| District of Columbia | 1 | 0 | 0 | 0 | 1 | 0 |
| Florida | 1 | 1 | 1 | 1 | 0 | 0 |
| Georgia | 5 | 2 | 5 | 2 | 0 | 0 |
| Hawaii | - | - | - | - | - | - |
| Indiana | 3 | 1 | 2 | 0 | 1 | 1 |
| lowa | 2 | 1 | 2 | 1 | 0 | 0 |
| Kentucky | 2 | 1 | 2 | 1 | 0 | 0 |
| Louisiana | 7 | 2 | 7 | 1 | 0 | 0 |
| Maine | 3 | 0 | 3 | 0 | 0 | 0 |
| Maryland | 0 | 0 | 0 | 0 | 0 | 0 |
| Massachusetts | 6 | 1 | 6 | 1 | 0 | 0 |
| Michigan | 3 | 0 | 2 | 0 | 0 | 0 |
| Minnesota | 5 | 1 | 5 | 1 | 0 | 0 |
| Mississippi | 8 | 0 | 8 | 0 | 0 | 0 |
| Missouri | 3 | 0 | 3 | 0 | 0 | 0 |
| Montana | 0 | 0 | 0 | 0 | 0 | 0 |
| Nebraska | 2 | 0 | 1 | 0 | 1 | 0 |
| Nevada | 0 | 0 | 0 | 0 | 0 | 0 |
| New Jersey | 3 | 0 | 3 | 0 | 0 | 0 |
| New Mexico | 15 | 7 | 7 | 7 | 12 | 4 |
| New York | 14 | 1 | 1 | 1 | 14 | 1 |
| North Carolina | - | - | - | - | - | - |
| North Dakota | 4 | 2 | 4 | 2 | 0 | 0 |
| Oregon | 0 | 0 | 0 | 0 | 0 | 0 |
| Pennsylvania | 2 | 1 | 2 | 1 | 0 | 0 |
| Rhode Island | - | - | - | - | - | - |
| South Carolina | - | - | - | - | - | - |
| Tennessee | , | 3 |  | 3 | 0 | 0 |
| Texas | 3 | 3 | 3 | 3 | 0 | 0 |
| Utah | 0 | 0 | 0 | 0 | 0 | 0 |
| Vermont | 0 | 0 | 0 | 0 | 0 | 0 |
| Virgina | - | - | - | - | - | - |
| Washington | - | - | - | - | - | - |
| West Virginia | - | 0 | 0 | 0 | - | 0 |
| Wisconsin | 0 | 0 | 0 | 0 | 0 | 0 |
| Wyoming | 4 | 4 | 4 | 4 | 0 | 0 |
| DDESS | - | - | - | - | - | - |
| DoDDS | - | - | - | - | - | - |
| Guam | 3 | 2 | 0 | 0 | 3 | 2 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
SD = Students with Disabilities (the term previously used was IEP).
LEP = Limited English Proficient student.
To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment. A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Table A.4c

# NAEP 1996 SD and LEP Participation Rates by States Grade 8 Public Schools 

|  | Total Percentage of Students - SD and LEP |  | Percentage of Students SD |  | Percentage of Students LEP |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Identified | Excluded | Identified | Excluded | Identified | Excluded |
| Nation | 11 | 5 | 9 | 4 | 3 | 1 |
| Alabama | 15 | 8 | 15 | 8 | 0 | 0 |
| Alaska | 15 | 5 | 10 | 5 | 5 | 1 |
| Arizona | 17 | 9 | 9 | 5 | 9 | 4 |
| Arkansas | 12 | 7 | 11 | 7 | 1 | 1 |
| California | 20 | 10 | 8 | 4 | 13 | 6 |
| Colorado | 11 | 4 | 10 | 4 | 2 | 1 |
| Connecticut | 16 | 9 | 14 | 7 | 3 | 2 |
| Delaware | 12 | 8 | 11 | 8 | 1 | 0 |
| District of Columbia | 12 | 9 | 9 | 7 | 3 | 2 |
| Florida | 15 | 9 | 12 | 7 | 3 | 2 |
| Georgia | 9 | 6 | 8 | 5 | 1 | 1 |
| Hawaii | 13 | 6 | 10 | 5 | 4 | 2 |
| Indiana | 12 | 5 | 11 | 5 | 1 | 0 |
| lowa | 11 | 4 | 10 | 4 | 0 | 0 |
| Kentucky | 10 | 5 | 10 | 5 | 0 | 0 |
| Louisiana | 9 | 5 | 8 | 5 | 1 | 0 |
| Maine | 10 | 4 | 9 | 4 | 0 | 0 |
| Maryland | 12 | 6 | 10 | 5 | 1 | 1 |
| Massachusetts | 15 | 7 | 14 | 6 | 1 | 1 |
| Michigan | 9 | 5 | 8 | 4 | 1 | 0 |
| Minnesota | 11 | 3 | 10 | 3 | 1 | 0 |
| Mississippi | 11 | 7 | 11 | 7 | 0 | 0 |
| Missouri | 12 | 8 | 12 | 7 | 1 | 1 |
| Montana | 10 | 3 | 9 | 3 | 0 | 0 |
| Nebraska | 11 | 4 | 10 | 4 | 1 | 1 |
| Nevada | 21 | 10 | 13 | 7 | 9 | 4 |
| New Hampshire | 15 | 4 | 15 | 4 | 0 | 0 |
| New Jersey | 12 | 6 | 9 | 5 | 3 | 2 |
| New Mexico | 18 | 8 | 12 | 4 | 7 | 4 |
| New York | 13 | 7 | 11 | 6 | 3 | 2 |
| North Carolina | 9 | 4 | 8 | 4 | 1 | 1 |
| North Dakota | 11 | 4 | 11 | 4 | 1 | 0 |
| Oregon | 12 | 4 | 11 | 4 | 1 | 1 |
| Rhode Island | 16 | 7 | 13 | 5 | 3 | 2 |
| South Carolina | 10 | 6 | 9 | 5 | 0 | 0 |
| Tennessee | 11 | 4 | 11 | 4 | 0 | 0 |
| Texas | 16 | 8 | 11 | 6 | 6 | 3 |
| Utah | 12 | 6 | 10 | 5 | 2 | 2 |
| Vermont | 13 | 5 | 12 | 4 | 1 | 0 |
| Virginia | 13 | 7 | 11 | 6 | 1 | 1 |
| Washington | 12 | 5 | 10 | 5 | 2 | 1 |
| West Virginia | 12 | 8 | 12 | 8 | 0 | 0 |
| Wisconsin | 12 | 7 | 11 | 7 | 1 | 1 |
| Wyoming | 8 | 1 | 8 | 1 | 0 | 0 |
| DDESS | 12 | 4 | 11 | 3 | 1 | 1 |
| DoDDS | 6 | 2 | 6 | 2 | 1 | 1 |
| Guam | 9 | 4 | 6 | 2 | 3 | 2 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
SD = Students with Disabilities (the term previously used was IEP).
LEP $=$ Limited English Proficient student.
To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment. A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics
Assessment.

## Table A.4d <br> NAEP 1996 SD and LEP Participation Rates by States Grade 8 Nonpublic Schools

|  | Total Percentage of Students - SD and LEP |  | Percentage of Students SD |  | Percentage of Students LEP |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Identified | Excluded | Identified | Excluded | Identified | Excluded |
| Nation | 3 | 0 | 2 | 0 | 1 | 0 |
| Alabama | 1 | 0 | 1 | 0 | 0 | 0 |
| Alaska | - | - | - | - | - | - |
| Arizona | - | - | - | - | - | - |
| Arkansas | 0 | 0 | 0 | 0 | 0 | 0 |
| California | 1 | 0 | 1 | 0 | 0 | 0 |
| Colorado | - | - | - | - | - | - |
| Connecticut | 5 | 0 | 4 | 0 | 1 | 0 |
| Delaware | 2 | 0 | 2 | 0 | 0 | 0 |
| District of Columbia | 1 | 0 | 0 | 0 | 1 | 0 |
| Florida |  | - | - | - | - | - |
| Indiana | - | - | - | - | - | - |
| Alaska | - | - | - | - | - | - |
| Georgia | 3 | 1 | 3 | 1 | 0 | 0 |
| lowa | 2 | 0 | 2 | 0 | 0 | 0 |
| Kentucky | 2 | 0 | 2 | 0 | 0 | 0 |
| Louisiana | 4 | 2 | 4 | 2 | 0 | 0 |
| Maine | - | - | - | - | - | - |
| Maryland | 1 | 1 | 1 | 1 | 0 | 0 |
| Massachusetts | 2 | 1 | 2 | 1 | 0 | 0 |
| Michigan | 2 | 0 | 2 | 0 | 0 | 0 |
| Minnesota | 2 | 0 | 1 | 0 | 1 | 0 |
| Mississippi | - | - | , | - | - | 0 |
| Missouri | 3 | 1 | 3 | 1 | 0 | 0 |
| Montana | 7 | 1 | 6 | 1 | 1 | 0 |
| Nebraska | 1 | 0 | 1 | 0 | 0 | 0 |
| Nevada | 0 | 0 | 0 | 0 | 0 | 0 |
| New Hampshire | 4 | 1 | 3 | 0 | 1 | 1 |
| New Jersey | 7 | 3 | 5 | 1 | 2 | 2 |
| New Mexico | 0 | 0 | 0 | 0 | 0 | 0 |
| New York | 0 | 0 | 0 | 0 | 0 | 0 |
| North Carolina | - | - | - | - | - | - |
| North Dakota | 2 | 0 | 2 | 0 | 0 | 0 |
| Oregon | 0 | 0 | 0 | 0 | 0 | 0 |
| Rhode Island | 1 | 0 | 1 | 0 | 0 | 0 |
| South Carolina | 0 | 0 | 0 | 0 | 0 | 0 |
| Tennessee | - | - | - | - | - | - |
| Texas | 1 | 1 | 1 | 1 | 0 | 0 |
| Utah | 4 | 4 | 4 | 4 | 0 | 0 |
| Vermont | 3 | 0 | 3 | 0 | 0 | 0 |
| Virginia | - | - | - | - | - | - |
| Washington | 0 | 0 | 0 | 0 | 0 | 0 |
| West Virginia | - | - | - | - | - | - |
| Wisconsin | 1 | 0 | 1 | 0 | 0 | 0 |
| Wyoming | 6 | 0 | 6 | 0 | 0 | 0 |
| DDESS | - | - | - | - | - | - |
| DoDDS |  |  |  |  | - |  |
| Guam | 1 | 0 | 0 | 0 | 1 | 0 |

National results are based on the national assessments samples, not on aggregated state assessment program samples.
SD = Students with Disabilities (the term previously used was IEP).
LEP = Limited English Proficient student.
To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment. A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Data Analysis and IRT Scaling

Subsequent to the professional scoring, all information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information had been compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse. Through stratification, the weighting assured that the representation of certain subpopulations corresponded to figures from the U.S. Census and the Current Population Survey. ${ }^{3}$

Analysis were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses subsequent to the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the block were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

It is standard ETS practice to treat all nonrespondents to the last question in a block as if they had not reached the question. For multiple-choice and short constructed response questions, this practice produces a reasonable pattern of results in that the proportion reaching the last question is not dramatically smaller than the proportion reaching the next-to-last question. However, for blocks that ended with extended constructed-response questions, the standard ETS practice would result in extremely large drops in the proportion of students attempting the final question. Therefore, for blocks ending with an extended constructedresponse question, students who answered the next-to-last question but did not respond to the extended constructed-response question were classified as having intentionally omitted the last question.

Item response theory (IRT) was used to estimate average mathematics scale scores for the nation, for various subgroups of interest within the nation, and for the states and territories. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups such as those defined by grades and characteristics, including gender and race/ethnicity.

Because of the BIB-spiraling design used by NAEP, students do not receive enough questions about a specific topic to provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would lead to misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale score level. Consequently, NAEP constructs sets of plausible values designed to represent the distribution of performance in the population.

[^25]A plausible value for an individual is not a scale score for that individual but may be regarded as a representative value from the distribution of potential scale scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the NAEP mathematics scale are based on the plausible values. They estimate values that would have been obtained had individual scale scores been observed—that is, had each student responded to a sufficient number of cognitive questions so that scores could be precisely estimated. ${ }^{4}$

For the 1990, 1992, and 1996 mathematics assessments, a scale ranging from 0 to 500 was created to report performance for each content strand. The scales summarize student performance across all three question types in the assessment (multiple-choice, short constructed-response, and extended constructed-response).

Each content strand scale is based on the distribution of student performance across all three grades in the national assessment (grades 4,8 , and 12). The scales have an average of 250 and a standard deviation of 50 . In addition, a composite scale was created as an overall measure of students' mathematical performance. This composite scale is a weighted average of the separate scales for the content strands. The weight for each content strand corresponds to the relative importance of each strand in the NAEP 1996 mathematics framework.

In producing the mathematics scales, three distinct IRT models were used. Multiplechoice questions were scaled using the three-parameter logistic (3PL) model; short constructedresponse questions rated as correct or incorrect were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level rubric, as well as extended constructed-response questions rated on a four- or five-level rubric, were scaled using a generalized partial-credit (GPC) model. ${ }^{5}$ Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex constructed-response questions.

The mathematics scale is composed of three types of questions: multiple-choice, constructed-response (scored dichotomously as correct or incorrect) and constructed-response (scored according to a partial-credit model). One natural question about the scale concerns the amount of information contributed by each type of question. Unfortunately, this question has no simple answer for the NAEP mathematics assessment, due to the complex procedures used to form the composite mathematics scale.

The information provided by a given question is determined by the IRT model used to scale the question and is a function of its item parameters. ${ }^{6}$ Thus, the answer to the query "How much information do the different types of questions provide?" will differ for each level of

[^26]mathematics performance. When considering the composite mathematics scale, the answer is even more complicated. The mathematics data are scaled separately by the content strands. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when they are derived from the same calibration. Because the composite scale is based on five separate calibrations, there is no direct way to compare the information provided by the questions on the composite scale.

## NAEP Reporting Froups

In this report, results are provided for groups of students defined by shared characteristicsregion of the country, gender, race or ethnicity, parental education, type of school, participation in Title I programs, and eligibility for the Free/Reduced-Price School Lunch program. Based on criteria described later in this appendix, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. For public school students, the minimum requirement is at least 62 students in a particular subgroup from at least 5 primary sampling units (PSUs). ${ }^{7}$ For nonpublic school students, the minimum requirement is 62 students from at least 6 different schools for the state assessment program or from at least 5 PSUs for the national assessment. However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations referred to in this report are presented below.

## Region

Results are reported for four regions of the nation: Northeast, Southeast, Central, and West. Figure A. 2 shows how states are subdivided into these regions. All 50 states and the District of Columbia are listed. Territories and the two Department of Defense Educational Activities jurisdictions are not assigned to any region.

Regional results are based on national assessment samples, not on aggregated state assessment program samples. Thus, the regional results are based on a sample that is different and separate from that used to report the state results.

[^27]

| Northeast | Southeast | Central | West |
| :--- | :--- | :--- | :--- |
| Connecticut | Alabama | Illinois | Alaska |
| Delaware | Arkansas | Indiana | Arizona |
| District of Columbia | Florida | lowa | California |
| Maine | Georgia | Kansas | Colorado |
| Maryland | Kentucky | Michigan | Hawaii |
| Massachusetts | Louisiana | Minnesota | Idaho |
| New Hampshire | Mississippi | Missouri | Montana |
| New Jersey | North Carolina | Nebraska | Nevada |
| New York | South Carolina | North Dakota | New Mexico |
| Pennsylvania | Tennessee | Ohio | Oklahoma |
| Rhode Island | Virginia* | South Dakota | Oregon |
| Vermont | West Virginia | Wisconsin | Texas |
| Virginia* |  |  | Utah |
|  |  |  | Washington |
|  |  |  | Wyoming |

* Note: The part of Virginia that is included in the Washington, DC metropolitan area is included in the Northeast region; the remainder of the state is included in the Southeast region.


## Gender

Results are reported separately for males and females.

## Race/Ethnicity

The race/ethnicity variable is derived from two questions asked of students and school records, and it is used for race/ethnicity subgroup comparisons. Two questions from the set of general student background questions were used to determine race/ethnicity:

If you are Hispanic, what is your Hispanic background?
Ol am not Hispanic
O Mexican, Mexican American, or Chicano

- Puerto Rican

O Cuban
Other Spanish or Hispanic background

Students who responded to this question by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the question, or provided information that was illegible or could not be classified, responses to the following question were examined to determine their race/ethnicity.

```
Which best describes you?
    OWhite (not Hispanic)
    O Black (not Hispanic)
    O Hispanic ("Hispanic" means someone who is Mexican, Mexican
        American, Chicano, Puerto Rican, Cuban, or other Spanish or
        Hispanic background)
    O Asian or Pacific Islander ("Asian or Pacific Islander" means someone
        who is from a Chinese, Japanese, Korean, Filipino, Vietnamese, or
        other Asian or Pacific Islander background.)
    O American Indian or Alaskan Native ("American Indian or Alaskan
        Native" means someone who is from one of the American Indian
        tribes or one of the original people of Alaska.)
    O Other (specify)
```

Students' race/ethnicity was then assigned on the basis of their responses. For students who filled in the seventh oval ("Other") and provided illegible information or information that could not be classified or who did not respond at all, race/ethnicity was assigned as determined by school records. ${ }^{8}$

Race/ethnicity could not be determined for students who did not respond to either of the demographic questions and whose schools did not provide information about race/ethnicity.

Details of how race/ethnicity classifications were derived is presented so that readers can determine how useful the results are for their particular purposes. Also, some students indicated that they were from a Hispanic background (e.g., Puerto Rican or Cuban) and that a racial/ethnic category other than Hispanic best described them. These students were classified as Hispanic based on the rules described above. Furthermore, information from the schools did not always correspond to how students described themselves. Therefore, the racial/ethnic results presented in this report attempt to provide a clear picture based on several sources of information.

As noted in Chapters 2 and 3, scale score and achievement level results for eighth grade Asian/Pacific Islander students are not included in the main body of the NAEP 1996 Mathematics Report Card. The decision not to publish these results is discussed in Appendix E.

[^28]
## Parents' Highest Level of Education

The variable representing the level of parental education is derived from responses to two questions from the set of general student background questions. Students were asked to indicate the extent of their mother's education.

How far in school did your mother go?
O She did not finish high school.
O She graduated from high school.
O She had some education after high school.
O She graduated from college.
Ol don't know.
Students were asked a similar question about their father's education level.

```
How far in school did your father go?
    O He did not finish high school.
    OHe graduated from high school.
    O He had some education after high school.
    He graduated from college.
    OI don't know.
```

The information was combined into one parental education reporting variable determined through the following process. If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. If a student did not know the level of education for both parents or did not know the level for one parent and did not respond for the other, the parental education level was classified as "I don't know." If the student did not respond for either parent, the student was recorded as having provided no response. (Nationally, 36 percent of fourth graders, 11 percent of eighth graders, and 3 percent of twelfth graders reported that they did not know the education level of either of their parents.)

## Type of School

Results are reported by the type of school that the student attends-public or nonpublic. Nonpublic schools include Catholic and other private schools. Although Bureau of Indian Affairs (BIA) schools and Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS) are not included in either the public or nonpublic categories, they are included in the overall national results. (A separate sample for DDESS was included as a jurisdiction in the state assessment.)

## Title I Participation

Based on available school records, students were classified either as currently participating in a Title I program or receiving Title I services or as not receiving such services. The classification applies only to the school year when the assessment was administered (i.e., the 1995-96 school year) and is not based on participation in previous years. If the school does not offer any Title I programs or services, all students in that school would be classified as not participating.

## Eligibility for the Free/Reduced-Price School Lanch Program

Based on available school records, students were classified as either currently eligible for the free/reduced-price lunch component of the Department of Agriculture's National School Lunch Program or not eligible. The classification applies only to the school year when the assessment was administered (i.e., the 1995-96 school year) and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available."

## Cantions in Interpretations

As described earlier, the NAEP mathematics scale makes it possible to examine relationships between students' performance and various background factors measured by NAEP. However, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not capture the influence of unmeasured variables. The results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

## Guidelines for Analysis and Reporting

This report describes mathematics performance for fourth, eighth, and twelfth graders and compares the results for various groups of students within these populations (e.g., those who have certain demographic characteristics or who responded to a specific background question in a particular way.) It also examines the results for individual demographic groups and individual background questions. However, it does not include an analysis of the relationships among combinations of these subpopulations or background questions.

## Estimating Variability

Because the statistics presented in this report are estimates of group and subgroup performance based on samples of students rather than the estimates that could be calculated if every student in the nation answered every question, the degree of uncertainty associated with the estimates should be taken into account. Two components of uncertainty are accounted for in the variability of statistics based on student ability: (1) the uncertainty due to sampling only a relatively small number of students and (2) the uncertainty due to sampling only a relatively small number of cognitive questions. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any content strand, the scale score for any single student would be imprecise. In this case, plausible values technology can be used to describe the performance of groups and subgroups of students, but the underlying imprecision involved in this step adds another component of variability to statistics based on NAEP scale scores. ${ }^{9}$ Appendix F provides the standard errors for the results presented in this report.

Typically, when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the standard errors may be quite large. Throughout this report, estimates of standard errors subject to a large degree of uncertainty are followed by the "!" symbol. In such cases, the standard errors-and any confidence intervals or significance tests involving these standard errors-should be interpreted cautiously. Additional details concerning procedures for identifying such standard errors are discussed in the forthcoming NAEP 1996 Technical Report.

The reader is reminded that, like findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources-inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling error is difficult to estimate, and because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

[^29]
## Drawing Inferences from the Results

Because the percentages of students in these subpopulations and their average scale scores are based on samples rather than on the entire population of fourth, eighth, or twelfth graders in the nation or a jurisdiction, the numbers reported are estimates. As such, they are subject to a measure of uncertainty, reflected in the standard error of the estimate. When the percentages or average scale scores of certain groups are compared, the standard error should be taken into account, and observed similarities or differences should not be relied on solely. Therefore, the comparisons discussed in this report are based on statistical tests that consider the standard errors of those statistics and the magnitude of the difference among the averages or percentages.

The results from the sample taking into account the uncertainty associated with all samples are used to make inferences about the population. Using confidence intervals based on the standard errors provides a way to make inferences about the population averages and percentages in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample average scale score $+/-2$ standard errors approximates a 95 -percent confidence interval for the corresponding population quantity. This statement means that one can conclude with approximately a 5 percent level of significance that the average performance of the entire population of interest (e.g., all fourth-grade students in public schools in a jurisdiction) is within +/- 2 standard errors of the sample average.

As an example, suppose that the average mathematics scale score of the students in a particular group was 256 with a standard error of 1.2. A 95 -percent confidence interval for the population quantity would be as follows:

$$
\begin{aligned}
& \text { Average } \pm 2 \text { standard errors } \\
& 256 \pm 2 \times 1.2 \\
& 256 \pm 2.4 \\
& 253.6,258.4
\end{aligned}
$$

Thus, one can conclude with a 5 percent level of significance that the average scale score for the entire population of students in that group is between 253.6 and 258.4.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. For extreme percentages, confidence intervals constructed in the above manner may not be appropriate, and accurate confidence intervals can be constructed only by using procedures that are quite complicated.

Extreme percentages, defined by both the magnitude of the percentage and the size of the sample from which it was derived, should be interpreted with caution. (The forthcoming NAEP 1996 Technical Report contains a more complete discussion of extreme percentages.)

## Analyzing Group Difierences <br> in Averages and Percentages

The statistical tests determine whether the evidence, based on the data from the groups in the sample, is strong enough to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher than or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. If the evidence is not sufficiently strong (i.e., the difference is not statistically significant), the averages or percentages are described as being not significantly different, regardless of whether the sample averages or percentages appear to be approximately the same or widely discrepant.

The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain a estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the standard error of the difference between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

$$
\text { Standard Error of the Difference }=\mathrm{SE}_{\mathrm{A}-\mathrm{B}}=\sqrt{\mathrm{SE}_{\mathrm{A}}^{2}+\mathrm{SE}_{\mathrm{B}}{ }^{2}}
$$

Similar to how the standard error for an individual group average or percentage is used, the standard error of the difference can be used to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups $+/-$ two standard errors of the difference represents an approximate 95 -percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant (different) at the 0.05 level. In this report, differences among groups that involve poorly defined variability estimates (i.e., denoted with a !) or extreme percentages are not discussed.

As an example, to determine whether the average mathematics scale score of Group A is higher that that of Group B, suppose that the sample estimates of the average scale scores and standard errors were as follows:

| Group | Average Scale <br> Score | Standard Error |
| :---: | :---: | :---: |
| A | 218 | 0.9 |
| B | 216 | 1.1 |

The difference between the estimates of the average scale scores of Groups A and B is two points (218-216). The standard error of this difference is

$$
\sqrt{0.9^{2}+1.1^{2}=1.4}
$$

Thus, an approximate 95 -percent confidence interval for this difference is

$$
\text { Difference }+/-2 \text { standard errors of the difference }
$$

$$
\begin{gathered}
2 \pm 2 \times 1.4 \\
2 \pm 2.8 \\
-0.8,4.8
\end{gathered}
$$

The value zero is within the confidence interval, therefore, there is insufficient evidence to claim that Group A outperformed Group B.

The procedures described in this section and the certainty ascribed to intervals (e.g., a 95-percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, in Chapters 2 and 3 of this report, many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), adjustments (called multiple comparison procedures) must be made to the methods described in the previous section. One such procedure, the Bonferroni method, was used in the analyses described in this report to confidence intervals for the differences among groups when sets of comparisons were considered. ${ }^{10}$ Thus, the confidence intervals for the sets of comparisons in the text are more conservative than those described on the previous pages.

Most of the multiple comparisons in this report pertain to relatively small sets or families of comparisons. For example, for discussions concerning comparisons of parents' level of education, six comparisons were conducted—all pairs of the four parental education levels. In these situations, Bonferroni procedures were appropriate. However, for the cross-state comparisons with a large family of comparisons, the False Discovery Rate (FDR) procedure ${ }^{11}$ was used to control the certainty level.

Unlike the Bonferroni procedure which controls the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, Bonferroni procedures are considered conservative for large families of comparisons. ${ }^{12}$ Therefore, the FDR procedure is more suitable for cross-state comparisons. A detailed description of the Bonferroni and FDR procedures appears in NAEP 1996 Technical Report and NAEP 1996 Technical Report for the State Assessment Program in Mathematics.

[^30]
## Revisions to the NAEP 1990 and 1992 Mathematies Findings

After the NAEP 1994 assessment has been conducted, two technical problems were discovered in the procedures used to develop the NAEP mathematics scale and achievement levels determined for the 1990 and 1992 mathematics assessments. These errors affected the mathematics scale scores reported in 1992 and the achievement level results reported in 1990 and 1992. The National Center for Education Statistics(NCES) and the National Assessment Governing Board (NAGB) have evaluated the impact of these errors and have reanalyzed and reported the revised results from both mathematics assessments. The technical errors have been corrected and the revised national and state scale score results for 1992 and achievement level results for 1990 and 1992 are presented in the NAEP 1996 mathematics reports.

Although the two technical problems that were discovered are discussed in greater detail in the NAEP 1996 Technical Report and NAEP 1996 Technical Report of the State Assessment in Mathematics, a brief summary is presented below.

The first technical problem resulted from an error in the computer program used to compute NAEP scale score results. The error occurred in the convention used to handle omitted responses in the item response theory (IRT) scaling of the partial-credit constructedresponse questions, and it was limited only to those questions. In analyses of the NAEP 1992 mathematics assessment, this error caused all blank responses to partial-credit constructedresponse questions (both omitted and not-reached responses) to be treated as missing-an acceptable treatment, but not the conventional choice for NAEP. (Because the NAEP 1990 mathematics assessment did not include these types of questions, the error did not occur.) The national and state assessments results were recalculated using the intended convention for the treatment of omitted responses.

In general, the effect of this technical problem on the previously reported NAEP 1992 mathematics findings was minimal, and it had little impact on policy-related interpretations. The recalculated 1992 mathematics scale score results, at the national and state levels, are quite similar to those published in the 1992 mathematics reports.

The second technical problem involved the development of the NAEP mathematics achievement level cut scores, and it concerned the mapping of the NAGB-approved achievement levels onto the NAEP mathematics scale. This error affected the achievement level results reported for the 1990 and 1992 mathematics assessments. In deriving the final levels recommended to NAGB, panelists' ratings for the multiple-choice and constructedresponse questions were combined to obtain an overall rating for the questions. When combined, the ratings were weighted based on the amount of information provided by each type of question. In other words, some of the questions "counted more" toward the overall cut scores than others. However, because the weighting was carried out incorrectly, the constructedresponse questions received more weight than intended. Therefore, the cut scores established by mapping the achievement levels onto the NAEP mathematics scale were incorrect, and the percentages of students at or above these levels were incorrectly estimated.

The program that mapped the achievement levels to the NAEP scale was corrected to appropriately weight the constructed-response questions, and revised mathematics achievement
level cut scores were developed based on the corrected scaling procedures. As a result, the cut scores for the three achievement levels at each grade were raised, and the percentages of students at or above the achievement levels were recalculated based on the corrected cut scores. Revised 1990 and 1992 percentages, for the national and state assessments, are presented in this report.

## Grade 12 Participation Rates and Motivation

NAEP has been described as a "low-stakes" assessment. That is, students receive no individual scores and their NAEP performance has no effect on their grades, promotions, or graduation. There has been continued concern that this lack of consequences affects participation rates of students and schools, as well as the motivation of students to perform well on NAEP. Of particular concern has been the performance of twelfth graders, who typically have lower student participation rates than fourth and eighth graders, and who are more likely to omit responses compared to the younger cohorts.

## Participation Rates

In NAEP, there has been a consistent pattern of lower participation rates for older students. In the 1994 NAEP assessments, for example, the student participation rates were 93 percent and 91 percent at grades 4 and 8 , respectively. At the twelfth grade, however, the participation rate was 81 percent. School participation rates (the percentage of sampled schools that participated in the assessment) have also typically decreased with grade level. Again citing the 1994 assessments, the school participation rate was 86 percent for the fourth grade; 86 percent for the eighth grade; and 79 percent for the twelfth grade.

The effect of participation rates on student performance, however, is unclear. Students may choose not to participate in NAEP for many reasons, such as desire to attend regular classes so as not to miss important instruction, or fear of not doing well on NAEP. Similarly, there are a variety of reasons for which various schools do not participate. The sampling weights and nonresponse adjustments, described earlier in this appendix, provide an approximate statistical adjustment for nonparticipation. However, the effect of some school and student nonparticipation may have some undetermined effect on results.

## Motivation

To the extent that students in the NAEP sample are not trying their hardest, NAEP results may underestimate student performance. The concern increases as students get older, and is particularly pronounced for twelfth graders. The students themselves furnish some evidence about their motivation. As part of the background questions, students were asked how important it was to do well on the NAEP mathematics assessment. They were asked to indicate whether it was very important, important, somewhat important, or not very important to them (see Table A.5). The percentage of students indicating they thought it was either important or very important to do well was 86 percent for fourth graders, 58 percent for eighth graders, and 31 percent for twelfth graders. Motivation to do well decreased at each higher grade assessed.

| Table A. 5 | Students' Report on How Important It Was for Them to Perform Well on the NAEP Mathematics Assessment |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | GRADE 4 |  | GRADE 8 |  | GRADE 12 |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| Not Very Important | 5 (0.3) | 219 (2.7) | 14 (0.7) | 278 (1.8) | 33 (0.9) | 306 (1.5) |
| Somewhat Important | 9 (0.5) | 228 (2.2) | 28 (0.8) | 275 (1.3) | 35 (0.7) | 305 (1.1) |
| Important | 24 (0.6) | 228 (1.4) | 34 (0.7) | 274 (1.2) | 23 (0.6) | 304 (1.2) |
| Very Important | 62 (1.0) | 223 (0.8) | 24 (0.7) | 263 (1.4) | 8 (0.6) | 293 (1.9) |

The standard errors of the estimated percentages and average scale scores appear in parentheses.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Several factors may contribute to this pattern. The NAEP was administered in the late winter, when high school seniors often have other things on their minds. More recently, the addition to NAEP of more constructed-response questions, which in many instances take longer for the student to answer, may also have had some effect on twelfth graders completing the assessment. As with participation rates, however, the combined effect of these and other factors is unknown.

It is also interesting to note that students who indicated it was very important for them to do well on NAEP did not have the highest average scores. In fact, at grades 8 and 12, students who reported it was not very important to do well also had higher average scores than those who reported it was very important to do well. These data further cloud the relationship between motivation and performance on NAEP.

## Need for Future Researeh

More research is needed to delineate the factors that contribute to nonparticipation and lack of motivation. To that end, NCES plans to commission a study of high school transcripts to learn more about the academic performance of twelfth grade students who do not participate in the assessment. In addition, NCES is currently investigating how various types of incentives can be effectively used to increase participation in NAEP.

## Appendix B

## 1996 State-Level Results for Selected Subyroups

This appendix includes state-by-state results from the NAEP 1996 state assessment program in mathematics for selected subgroups discussed in Chapters 2 and 3. Average scale scores and achievement level results are presented for gender, race/ethnicity, parental education, type of school (public and nonpublic), Title I participation, and eligibility for the free/reduced-price lunch program. In all the tables in this appendix, DDESS refers to Department of Defense Domestic Department Elementary and Secondary Schools and DoDDS refers to overseas Department of Defense Dependents Schools.

## Average Mathematics Scale Scores and Achievement Levels by Gender Grade 4, Public Schools Only



National results are based on the national assessment samples, not on aggregated state assessment program samples.
*** Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^31]

National results are based on the national assessment samples, not on aggregated state assessment program samples.
***Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.


[^32]
## Table B. 4 <br> Mathemotics Achievement Levels (continued) by Race/Ethnicity Grade 4, Public Schools Only



National results are based on the national assessment samples, not on aggregated state assessment program samples.
*** Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

# Table B. 5 

 Average Mathematics Scale Scoresby Race/Ethnicity Grade 8, Public Schools Only


National results are based on the national assessment samples, not on aggregated state assessment program samples.
***Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

-     - Quality control activities and special analyses involving state assessement data raised concerns about the accuracy and precision of national grade 8 Asian/Pacific results. As a result, they are omitted from the body of this report. See Appendix A for a more detailed discussion.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

Table B. 6 Mathematics Achievement Levels by Race/Ełhnicity Grade 8, Public Schools Only


[^33]| Table B. 6 (continued) | Muthematics Achievement Levels by Race/Ethnicity Grade 8, Public Schools Only |  |
| :---: | :---: | :---: |



[^34]| Table B. 7 | Average Mathematics Scale Scores by Parents' Highest Level of Education Grade 4, Public Schools Only |  |
| :---: | :---: | :---: |



[^35]
# Table B. 8 



National results are based on the national assessment samples, not on aggregated state assessment program samples.
*** Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.


[^36]

[^37]

National results are based on the national assessment samples, not on aggregated state assessment program samples.
***Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

\section*{Table B. 10 Mathematics Achievement Levels by (continued) Parents' Highest Level of Educution Grade 8, Public Schools Only <br> REPOR NATION' CARD | Napp |
| :--- | :--- |
| $\overline{y n}$ |}



[^38]

[^39]
## Table B. 12

## Average Mathematics Scale Scores and Achievement Levels by Type of School Grade 8



National results are based on the national assessment samples, not on aggregated state assessment program samples.
*** Sample size insufficient to permit reliable estimates.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.


[^40]Table B. 14
Average Mathematics Scile Scores and Achievement Levels by Titile I Participation Grade 8, Public Schools Only


[^41]
## Table B. 15



[^42]

[^43]

[^44]

[^45]
## Appendix C

## State-Level Contextual Vaniables

To help better place results from the NAEP 1996 state assessment program into context, this appendix presents selected state-level data from sources other than NAEP. The information presented are taken from the Digest of Education Statistics 1996.

## Table C.Ia <br> School System Characteristics from Non-NAEP Sources



|  | Estimated Total and School-Age Resident Population: 1995 (Estimates as of July 1) ${ }^{1}$ |  | Entollment in Public Elementary and Secondary Schools: Fall 1994² |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Total, All Ages (in thousands) | 5. to 17 -year-olds (in thousands) | Total | Kindergarten through Grade 8 | Grades 9 to 12 |
| United States | 262,755 | 49,149 | 44,108,775 | 31,894,333 | 12,214,442 |
| Alabama | 4,253 | 779 | 736,472 | 535,187 | 201,285 |
| Alaska | 604 | 136 | 127,057 | 93,719 | 33,338 |
| Arizona | 4,218 | 837 | 737,424 | 542,904 | 194,520 |
| Arkansas | 2,484 | 477 | 447,565 | 319,282 | 128,283 |
| California | 31,589 | 5,984 | 5,407,043 | 3,955,434 | 1,451,609 |
| Colorado | 3,747 | 712 | 640,521 | 469,755 | 170,766 |
| Connecticut | 3,275 | 570 | 506,824 | 375,638 | 131,186 |
| Delaware | 717 | 127 | 106,813 | 76,819 | 29,994 |
| District of Columbia | 554 | 75 | 80,450 | 62,126 | 18,324 |
| Florida | 14,166 | 2,403 | 2,108,968 | 1,567,328 | 541,640 |
| Georgia | 7,201 | 1,372 | 1,270,948 | 934,650 | 336,298 |
| Hawaii | 1,187 | 213 | 183,795 | 133,675 | 50,120 |
| Idaho | 1,163 | 258 | 240,448 | 168,887 | 71,561 |
| Illinois | 11,830 | 2,205 | 1,916,172 | 1,368,041 | 548,131 |
| Indiana | 5,803 | 1,079 | 968,933 | 678,943 | 289,990 |
| lowa | 2,842 | 541 | 499,550 | 344,754 | 154,796 |
| Kansas | 2,565 | 510 | 460,838 | 329,211 | 131,627 |
| Kentucky | 3,860 | 712 | 657,642 | 467,005 | 190,637 |
| Louisiana | 4,342 | 903 | 797,933 | 583,892 | 214,041 |
| Maine | 1,241 | 230 | 212,601 | 155,903 | 56,698 |
| Maryland | 5,042 | 904 | 790,938 | 580,903 | 210,035 |
| Massachusetts | 6,074 | 1,019 | 893,727 | 658,507 | 235,220 |
| Michigan | 9,549 | 1,837 | 1,614,784 | 1,170,251 | 444,533 |
| Minnesota | 4,610 | 925 | 821,693 | 581,426 | 240,267 |
| Mississippi | 2,697 | 553 | 505,962 | 366,846 | 139,116 |
| Missouri | 5,324 | 1,012 | 878,541 | 628,286 | 250,255 |
| Montana | 870 | 179 | 164,341 | 116,748 | 47,593 |
| Nebraska | 1,637 | 329 | 287,100 | 203,055 | 84,045 |
| Nevada | 1,530 | 279 | 250,747 | 185,336 | 65,411 |
| New Hampshire | 1,148 | 219 | 189,319 | 138,851 | 50,468 |
| New Jersey | 7,945 | 1,386 | 1,174,206 | 862,331 | 311,875 |
| New Mexico | 1,685 | 362 | 327,248 | 229,168 | 98,080 |
| New York | 18,136 | 3,177 | 2,766,208 | 1,949,245 | 816,963 |
| North Carolina | 7,195 | 1,285 | 1,156,767 | 847,463 | 309,304 |
| North Dakota | 641 | 129 | 119,288 | 83,419 | 35,869 |
| Ohio | 11,151 | 2,087 | 1,814,290 | 1,295,289 | 519,001 |
| Oklahoma | 3,278 | 648 | 609,718 | 442,607 | 167,111 |
| Oregon | 3,141 | 587 | 521,945 | 371,967 | 149,978 |
| Pennsylvania | 12,072 | 2,125 | 1,765,891 | 1,244,103 | 521,788 |
| Rhode Island | 990 | 170 | 147,487 | 107,913 | 39,574 |
| South Carolina | 3,673 | 682 | 648,673 | 468,798 | 179,875 |
| South Dakota | 729 | 154 | 143,482 | 101,805 | 41,677 |
| Tennessee | 5,256 | 945 | 881,355 | 640,534 | 240,821 |
| Texas | 18,724 | 3,819 | 3,677,171 | 2,720,623 | 956,548 |
| Utah | 1,951 | 491 | 474,675 | 328,482 | 146,193 |
| Vermont | 585 | 110 | 104,533 | 75,590 | 28,943 |
| Virginia | 6,618 | 1,149 | 1,060,809 | 774,319 | 286,490 |
| Washington | 5,431 | 1,033 | 938,314 | 673,107 | 265,207 |
| West Virginia | 1,828 | 325 | 310,511 | 212,808 | 97,703 |
| Wisconsin | 5,123 | 1,009 | 860,686 | 601,215 | 259,471 |
| Wyoming | 480 | 104 | 100,369 | 70,185 | 30,184 |

[^46]

|  | Poverty status of 5- to 17-year-olds: 19943 |  | Number of Children (Birth to age 21) Served Under State-Operated Individuals With Disabilities Education Act and Chapter 1 of the Education Consolidation and Improvement Act Programs ${ }^{4}$ |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Number in Poverty (in thousands) | Percent in Poverty | 1993-94 School Year | Percent Change: 1990-91 to 1993-94 |
| United States | 9,974 | 20.1 | 5,318,021 | 11.7 |
| Alabama | 157 | 19.5 | 99,760 | 5.1 |
| Alaska | 15 | 11.7 | 18,006 | 22.1 |
| Arizona | 189 | 23.4 | 69,530 | 21.5 |
| Arkansas | 87 | 20.4 | 53,187 | 11.2 |
| California | 1,550 | 25.3 | 533,807 | 13.7 |
| Colorado | 69 | 9.9 | 66,595 | 16.6 |
| Connecticut | 100 | 18.6 | 71,863 | 11.3 |
| Delaware | 10 | 9.8 | 15,196 | 6.3 |
| District of Columbia | 29 | 30 | 6,994 | 11.2 |
| Florida | 563 | 22.1 | 289,539 | 22.7 |
| Georgia | 267 | 18.5 | 123,143 | 20.7 |
| Hawaii | 21 | 12 | 15,248 | 15.8 |
| Idaho | 39 | 15.5 | 23,536 | 6.9 |
| Illinois | 405 | 18 | 257,986 | 7.9 |
| Indiana | 164 | 13.7 | 127,961 | 11.6 |
| lowa | 74 | 13.5 | 63,373 | 4.4 |
| Kansas | 97 | 19.5 | 50,438 | 11.6 |
| Kentucky | 200 | 26.6 | 80,539 | 1.4 |
| Louisiana | 337 | 36.8 | 86,931 | 18 |
| Maine | 20 | 9.6 | 29,350 | 4.9 |
| Maryland | 143 | 17.2 | 97,998 | 6.6 |
| Massachusetts | 121 | 12.2 | 160,275 | 3.7 |
| Michigan | 326 | 17.9 | 181,251 | 8.6 |
| Minnesota | 115 | 13.7 | 90,918 | 12.4 |
| Mississippi | 138 | 28.2 | 64,153 | 5.3 |
| Missouri | 204 | 23.6 | 114,008 | 11.8 |
| Montana | 20 | 12.3 | 18,401 | 7 |
| Nebraska | 43 | 12.5 | 37,112 | 13.3 |
| Nevada | 45 | 16.2 | 25,242 | 36.9 |
| New Hampshire | 23 | 12.2 | 23,354 | 18.8 |
| New Jersey | 211 | 14.6 | 190,003 | 4.8 |
| New Mexico | 111 | 29.2 | 43,474 | 20.6 |
| New York | 769 | 23.5 | 365,697 | 18.9 |
| North Carolina | 206 | 18.4 | 136,513 | 10.9 |
| North Dakota | 15 | 11.6 | 12,440 | -0.5 |
| Ohio | 448 | 19.5 | 219,875 | 7 |
| Oklahoma | 140 | 21.5 | 73,130 | 11.4 |
| Oregon | 81 | 13.7 | 63,212 | 14.6 |
| Pennsylvania | 400 | 19 | 210,826 | -3.9 |
| Rhode Island | 24 | 13.3 | 23,582 | 11.9 |
| South Carolina | 121 | 18.7 | 81,930 | 5.4 |
| South Dakota | 32 | 18.2 | 15,907 | 6.1 |
| Tennessee | 206 | 20.1 | 119,146 | 13.6 |
| Texas | 1,084 | 26.8 | 411,917 | 17.5 |
| Utah | 46 | 9.9 | 51,950 | 8.8 |
| Vermont | 7 | 7 | 10,452 | -14.8 |
| Virginia | 157 | 12.6 | 131,599 | 15.5 |
| Washington | 146 | 14.6 | 101,254 | 18.6 |
| West Virginia | 66 | 22 | 44,528 | 3.2 |
| Wisconsin | 120 | 12.1 | 102,412 | 17.8 |
| Wyoming | 12 | 10.7 | 12,480 | 11.4 |

[^47]| Table C.Tc (continued) | School System Characteristics from Non-NAEP Sources |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Elementary and Secondary Education Expenditures | Pupi-Teacher Ratios in Public Elementary and Secondary | Estimated Average Public and (cu | Salaries of Teachers Schools <br> s) |
|  | Per Capita: 1991-92 ${ }^{5}$ | Schools: Fall $1994^{\circ}$ | NEA: 1995-967 | AFT: 1994-95 ${ }^{8}$ |
| United States | 896.57 | 17.3 | 37,846 | 36,744 |
| Alabama | 585.31 | 17.2 | 31,307 | 30,545 |
| Alaska | 1,713.81 | 17.6 | 49,620 | 47,864 |
| Arizona | 835.69 | 19.3 | 32,484 | 32,223 |
| Arkansas | 705.09 | 17.1 | 29,322 | 28,950 |
| California | 868.44 | 24 | 42,516 | 40,667 |
| Colorado | 900.58 | 18.4 | 35,364 | 34,571 |
| Connecticut | 1,124.30 | 14.4 | 50,400 | 50,598 |
| Delaware | 905.69 | 16.6 | 40,533 | 39,076 |
| District of Columbia | 1,066.24 | 13.2 | 43,700 | 43,142 |
| Florida | 819.3 | 19.1 | 33,320 | 32,590 |
| Georgia | 805.85 | 16.3 | 34,307 | 32,198 |
| Hawaii | 702.2 | 17.9 | 35,807 | 37,443 |
| Idaho | 775.69 | 19.1 | 30,891 | 29,784 |
| Illinois | 801.64 | 17.3 | 41,008 | 39,445 |
| Indiana | 857.87 | 17.5 | 37,805 | 36,799 |
| lowa | 917.11 | 15.7 | 32,376 | 31,511 |
| Kansas | 856.45 | 15.1 | 35,518 | 32,085 |
| Kentucky | 654.64 | 17 | 33,018 | 32,272 |
| Louisiana | 814.21 | 16.6 | 26,800 | 26,811 |
| Maine | 962.73 | 13.8 | 32,869 | 31,972 |
| Maryland | 877.49 | 17 | 41,215 | 40,661 |
| Massachusetts | 811.98 | 14.8 | 43,756 | 40,976 |
| Michigan | 1,012.79 | 20.1 | 49,168 | 46,575 |
| Minnesota | 1,060.85 | 17.5 | 36,937 | 35,948 |
| Mississippi | 639.56 | 17.5 | 27,689 | 26,818 |
| Missouri | 781.87 | 15.5 | 33,341 | 31,209 |
| Montana | 934.99 | 16.3 | 29,364 | 28,785 |
| Nebraska | 924.51 | 14.5 | 31,496 | 30,922 |
| Nevada | 897.18 | 18.7 | 36,167 | 38,010 |
| New Hampshire | 889.57 | 15.6 | 35,792 | 34,721 |
| New Jersey | 1,263.17 | 13.8 | 47,910 | 47,038 |
| New Mexico | 827.45 | 17.2 | 29,349 | 28,394 |
| New York | 1,224.39 | 15.2 | 48,115 | 47,612 |
| North Carolina | 788.77 | 16.2 | 30,564 | 30,793 |
| North Dakota | 832.42 | 15.3 | 26,969 | 26,317 |
| Ohio | 813.62 | 16.6 | 37,835 | 36,971 |
| Oklahoma | 778.17 | 15.5 | 28,909 | 28,745 |
| Oregon | 956.96 | 19.9 | 39,650 | 38,871 |
| Pennsylvania | 910.93 | 17.1 | 46,916 | 44,510 |
| Rhode Island | 864.33 | 14.7 | 42,160 | 40,729 |
| South Carolina | 800.23 | 16.4 | 31,568 | 30,366 |
| South Dakota | 819.08 | 14.4 | 26,346 | 26,037 |
| Tennessee | 586.25 | 18.6 | 33,451 | 31,270 |
| Texas | 885.47 | 15.7 | 32,000 | 31,223 |
| Utah | 830.92 | 24.3 | 30,452 | 28,919 |
| Vermont | 1,120.15 | 13.8 | 36,295 | 35,207 |
| Virginia | 854.34 | 14.6 | 34,687 | 33,907 |
| Washington | 1,045.76 | 20.2 | 38,025 | 36,160 |
| West Virginia | 865.8 | 14.8 | 32,155 | 31,944 |
| Wisconsin | 1,015.96 | 15.9 | 38,571 | 37,617 |
| Wyoming | 1,328.26 | 15 | 31,571 | 31,285 |

[^48]${ }^{6}$ U.S. Department of Education, National Center of Education Statistics, Common Core of Data surveys.
7 National Education Association (NEA), Estimates of School Statistics, and unpublished data. (Latest edition 1995-96, Copyright © 1996 by the National Education Association. All rights reserved.)
${ }^{8}$ American Federation of Teachers (AFT), Survey and Analysis of Salary Trends, various years.
Information reprinted from the Digest of Education Statistics 1996 (NCES 96-133).

## Appendix D

## State-Level SD/LEP Infiormation

This appendix contains national and state-level public school results on identification and inclusion rates for students with disabilities and LEP students. Results are presented for grades 4 and 8, the grades at which the 1996 state NAEP mathematics assessment was conducted.

Table D. 1 presents the percentages of the NAEP population at each grade that were identified as students with disabilities, LEP students, or both. In the nation's public schools, 12 percent of the fourth graders were identified as students with disabilities (including those who were also identified as LEP students). The percentage of fourth-grade students with disabilities ranged from 7 percent (in Guam) to 15 percent (in Massachusetts) with the majority of jurisdictions ( 35 of 47 that participated) identifying between 10 and 14 percent of their students as students with disabilities.

Nationally, nearly 5 percent of the nation's public school students are identified as LEP (including those who were also identified as students with disabilities). However, substantial variability across jurisdictions is evident in the percentages of fourth-grade students so identified. In five jurisdictions (Arizona, California, New Mexico, Texas and Guam), the percentage of LEP students in the fourth grade is at least 10 percent with California identifying nearly one in four of its fourth graders as LEP. In contrast, 28 of the 47 jurisdictions identified 2 percent or less of their fourth-graders as limited English proficient.

Consistent with past NAEP data, the national percentages of public school students with disabilities and LEP students are lower at grade 8 than at grade 4. Nationally, about 10 percent of public school eighth-graders are identified as students with disabilities (including those who were also identified as LEP students). The percentages range from 4 percent (in Guam) to 14 percent (in Massachusetts) with 34 of the 44 participating jurisdictions identifying between 9 and 12 percent of eighth-graders as students with disabilities. Approximately 4 percent of the nation's eighth-graders are identified as LEP (including those who were also identified as students with disabilities). Only one jurisdiction (California) identifies more than 10 percent of its population as being limited English proficient while in 33 of the 44 participating jurisdictions 2 percent or less of the eighth grade public school population is so identified.

Table D. 2 presents the percentages of the NAEP population at each grade that were excluded from the assessment in the S1 and S2 samples.

At grade 4, the national public school results again indicate that slightly more students were excluded using the revised criteria. However, the grade 4 state assessment results do not corroborate this finding. Observed state-level exclusion percentages were not consistently lower in one or the other sample types and differences between the samples in exclusion percentages were not statistically significant for any of the jurisdictions. At grade 8, national and state public school results are in agreement, with one exception, in showing little evidence of any effect. In
one jurisdiction, Delaware, a smaller percentage of the population was excluded in S 2 using the revised criteria than in Sl using the original criteria. It should be noted, however, that because of its size, fewer schools are represented in each of the Delaware's samples than in most of the other jurisdictions. Furthermore, results from the remaining jurisdictions do not suggest a clear pattern of greater inclusion for either of the sets of criteria.

At the national level, and in many of the jurisdictions that participated in the state assessment, students with disabilities and LEP students constitute a relatively modest percentage of the total school population. Consequently, examining exclusion rates (as was done in Chapter 4) may not, in some cases, provide a sufficiently sensitive measure of the effects of the inclusion criteria changes. Further analyses of national inclusion rates among students with disabilities and LEP students were included in Chapter 4. However, due to space limitations, similar analyses at the state level were not included in the main body of the report. These analyses are included in this appendix. ${ }^{1}$

Table D. 3 presents the percentages of assessed students with disabilities for the nation's public schools and for each of the jurisdictions participating in the state assessment. Considerable variability across jurisdictions is evident in the percentages of students with disabilities who are assessed in NAEP. The District of Columbia assessed less than 20 percent of its grade 4 students with disabilities, and less than 30 percent of its grade 8 students with disabilities, regardless of which inclusion criteria was used. In contrast, several jurisdictions (Nebraska, North Dakota, and Wyoming) assessed more than 60 percent of their students with disabilities at both grades, again regardless of which inclusion criteria were employed.

Comparisons of the S1 (sample using existing inclusion criteria) and S2 (sample using the revised inclusion criteria) inclusion percentages for students with disabilities across jurisdictions provide little evidence of a systematic or unidirectional effect due to changes in inclusion criteria. At grade 4, observed inclusion percentages using the revised criteria were higher for 17 of 46 jurisdictions. ${ }^{2}$ Averaged over jurisdictions, the difference in S1 and S2 inclusion percentages was about 2 percentage points ( 49 percent in S 1 versus 47 percent in S 2 ). At grade 8, observed inclusion percentages using the revised criteria were higher for 23 of 44 jurisdictions. In only one of these jurisdictions (Delaware) was the difference between S1 and S2 inclusion rates significantly different. Averaged over jurisdictions, the S1 and S2 inclusion percentages were virtually identical ( 53 percent).

Table D. 4 presents LEP student inclusion percentages for the nation's public schools and for the small number of jurisdictions participating in the NAEP state assessment in which samples of LEP students were sufficiently large to permit meaningful analysis. There were no significant differences between S1 and S2 LEP student inclusion percentages at the national level, again suggesting that revisions to the inclusion criteria had little, if any, impact on the percentage of LEP students that were assessed. Some limited evidence to contrary can be found in the state assessment results. For one jurisdiction (Texas at grade 8), a larger percentage of LEP students was included in S 2 than in S 1 . However, in the case of Texas, grade 4 results are not consistent with those obtained at grade 8 . On balance, the bulk of the evidence suggests that the percentage of LEP students assessed were not impacted by the changes made to the NAEP inclusion criteria.

[^49]| Table D. | Percentages of Students Identified as SD, LEP or Both by State, Grades 4 and 8, Public Schools Only |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Grade 4 |  |  |  | Grade 8 |  |  |  |
|  | Total | $\begin{aligned} & \text { SD } \\ & \text { Only } \end{aligned}$ | $\begin{gathered} \text { SD and } \\ \text { LEP } \\ \text { Both } \end{gathered}$ | $\begin{aligned} & \text { LEP } \\ & \text { Only } \end{aligned}$ | Total | $\begin{aligned} & \text { SD } \\ & \text { Only } \end{aligned}$ | $\begin{gathered} \text { SD and } \\ \text { LEP } \\ \text { Both } \end{gathered}$ | $\begin{aligned} & \text { LEP } \\ & \text { Only } \end{aligned}$ |
| Nation | 16 (0.5) | 11 (0.5) | $1(0.1)$ | $4(0.3)$ | 12 (0.6) | 9 (0.5) | 1 (0.1) | 3 (0.3) |
| Alabama | 12 (0.7) | 11 (0.7) | 0 (0.0) | 0 (0.1) | 13 (1.0) | 13 (1.0) | $0(0.0)$ | 0 (0.2) |
| Alaska | 20 (1.3) | 12 (0.8) | 1 (0.2) | 7 (1.0) | 15 (1.3) | 11 (0.9) | 0 (0.2) | 4 (0.8) |
| Arizona | 21 (1.6) | 9 (0.8) | 1 (0.2) | 11 (1.6) | 17 (1.1) | 8 (0.6) | 1 (0.3) | 8 (0.9) |
| Arkansas | 10 (0.8) | 9 (0.8) | 0 (0.0) | 1 (0.2) | 11 (0.8) | 11 (0.8) | 0 (0.1) | 1 (0.2) |
| California | 33 (1.9) | 6 (0.7) | 2 (0.3) | 24 (1.9) | 20 (1.2) | 7 (0.6) | 1 (0.2) | 12 (1.2) |
| Colorado | 15 (0.9) | 11 (0.7) | 0 (0.1) | 3 (0.6) | 12 (0.8) | 10 (0.6) | 0 (0.2) | 2 (0.4) |
| Connecticut | 16 (1.1) | 13 (0.8) | 1 (0.2) | 2 (0.7) | 15 (0.7) | 13 (0.7) | 0 (0.1) | 2 (0.3) |
| Delaware | 14 (0.8) | 12 (0.8) | 0 (0.1) | 1 (0.1) | 13 (0.6) | 12 (0.6) | 0 (0.1) | 1 (0.1) |
| District of Columbia | 14 (0.6) | 8 (0.4) | 0 (0.1) | 5 (0.4) | 13 (0.8) | 9 (0.7) | 0 (0.1) | 4 (0.4) |
| Florida | 19(1.3) | 13 (0.9) | 1 (0.3) | 5 (0.9) | 16 (1.0) | 12 (0.7) | 0 (0.1) | 4 (0.6) |
| Georgia | 13 (1.0) | 11 (0.9) | 0 (0.1) | 2 (0.6) | 10 (0.8) | 9 (0.6) | 0 (0.1) | 1 (0.4) |
| Hawaii | 14 (0.7) | 9 (0.6) | 0 (0.1) | 5 (0.5) | 12 (0.7) | 8 (0.7) | 0 (0.2) | 3 (0.4) |
| Indiana | 11 (0.9) | 10 (0.8) | 0 (0.2) | 0 (0.3) | 12 (0.6) | 12 (0.7) | 0 (0.1) | 1 (0.2) |
| lowa | 13 (0.9) | 11 (0.7) | 0 (0.1) | 2 (0.7) | 13 (0.6) | 12 (0.6) | 0 (0.1) | 0 (0.2) |
| Kentucky | 10 (0.8) | 10 (0.8) | 0 (0.0) | 0 (0.1) | $9(0.6)$ | 9 (0.6) | 0 (0.0) | 0 (0.1) |
| Louisiana | 14 (0.9) | 13 (0.8) | 0 (0.1) | 1 (0.3) | 10 (0.8) | 10 (0.7) | $0(0.0)$ | 1 (0.3) |
| Maine | 15 (0.8) | 14 (0.8) | 0 (0.2) | 0 (0.2) | 12 (0.7) | 11 (0.6) | 0 (0.1) | 1 (0.2) |
| Maryland | 14 (0.9) | 13 (0.9) | 0 (0.0) | 1 (0.3) | 12 (0.7) | 11 (0.6) | 0 (0.1) | 1 (0.2) |
| Massachusetts | 18 (1.0) | 14 (0.8) | 1 (0.2) | 3 (0.7) | 17 (1.1) | 14 (0.9) | 0 (0.1) | 2 (0.5) |
| Michigan | 11 (0.8) | 10 (0.8) | 0 (0.2) | 1 (0.4) | 9 (0.7) | 8 (0.7) | 0 (0.1) | 1 (0.3) |
| Minnesota | 14 (0.7) | 11 (0.7) | 0 (0.1) | 3 (0.4) | 11 (0.7) | 10 (0.6) | 0 (0.1) | 1 (0.4) |
| Mississippi | 8 (0.7) | 8 (0.7) | 0 (0.0) | 0 (0.1) | 11 (0.8) | 11 (0.8) | $0(0.0)$ | 0 (0.1) |
| Missouri | 14 (0.8) | 13 (0.8) | 0 (0.1) | 0 (0.2) | 12 (0.7) | 11 (0.6) | 0 (0.1) | 0 (0.4) |
| Montana | 10 (0.7) | 9 (0.7) | 0 (0.1) | 0 (0.1) | $9(0.8)$ | 9 (0.8) | 0 (0.1) | 0 (0.2) |
| Nebraska | 15 (1.1) | 14 (0.9) | 0 (0.1) | 2 (0.6) | 12 (1.0) | 11 (0.8) | 0 (0.1) | 1 (0.3) |
| Nevada | 16(1.1) | $9(0.6)$ | 1 (0.2) | 7 (0.9) | , | $\ddagger$ | $\ddagger$ | , |
| New Jersey | 11 (1.0) | 9 (0.9) | 0 (0.1) | 2 (0.6) | $\ddagger$ | $\ddagger$ | $\ddagger$ | $\ddagger$ |
| New Mexico | 22 (1.9) | 12 (0.9) | 2 (0.4) | 8 (1.5) | 18 (0.9) | 12 (0.8) | 1 (0.3) | 5 (0.5) |
| New York | 15 (1.1) | $9(0.8)$ | 1 (0.2) | 5 (0.8) | 14 (0.8) | 9 (0.8) | 0 (0.1) | 4 (0.7) |
| North Carolina | 14 (0.7) | 13 (0.7) | 0 (0.1) | 1 (0.3) | $9(0.6)$ | 9 (0.6) | 0 (0.1) | 1 (0.2) |
| North Dakota | 11 (0.8) | 10 (0.8) | 0 (0.0) | 0 (0.1) | 10 (0.6) | 9 (0.5) | 0 (0.2) | 0 (0.2) |
| Oregon | 19(1.0) | 13 (0.6) | 0 (0.1) | 5 (0.9) | 12 (0.9) | 10 (0.6) | 0 (0.1) | 2 (0.4) |
| Pennsylvania | 9 (0.7) | 8 (0.6) | 0 (0.1) | 1 (0.2) | $\ddagger$ | $\ddagger$ | $\ddagger$ | $\ddagger$ |
| Rhode Island | 18 (1.0) | 13 (0.9) | 0 (0.1) | 5 (0.6) | 17 (0.7) | 12 (0.6) | 0 (0.1) | 4 (0.4) |
| South Carolina | 12 (0.8) | 12 (0.7) | 0 (0.0) | 0 (0.1) | 10 (0.7) | 10 (0.7) | $0(0.0)$ | 0 (0.1) |
| Tennessee | 13 (0.8) | 12 (0.8) | 0 (0.1) | 1 (0.3) | 11 (0.8) | 11 (0.8) | 0 (0.1) | 1 (0.5) |
| Texas | 24 (1.8) | 11 (0.7) | 1 (0.4) | 12 (1.7) | 17 (1.1) | 11 (0.6) | 1 (0.3) | 6 (0.9) |
| Utah | 13 (0.8) | 11 (0.7) | 1 (0.2) | 2 (0.4) | 11 (0.7) | 10 (0.7) | 0 (0.1) | 1 (0.2) |
| Vermont | 14 (0.8) | 13 (0.8) | 0 (0.1) | 1 (0.2) | 12 (0.7) | $12(0.7)$ | 0 (0.0) | 1 (0.2) |
| Virginia | 14 (0.9) | 12 (0.8) | 0 (0.1) | 2 (0.4) | 13 (0.7) | 11 (0.6) | 0 (0.1) | 2 (0.3) |
| Washington | 13 (0.7) | 10 (0.6) | 0 (0.1) | 3 (0.4) | 13 (0.9) | 11 (0.7) | 0 (0.1) | 2 (0.6) |
| West Virginia | 13 (0.9) | 13 (0.9) | 0 (0.2) | 0 (0.1) | 13 (0.7) | 13 (0.7) | 0 (0.0) | 0 (0.1) |
| Wisconsin | 12 (1.1) | 10 (0.7) | 0 (0.2) | 1 (0.5) | 12 (0.8) | 11 (0.7) | 0 (0.1) | 1 (0.4) |
| Wyoming | 13 (0.8) | 12 (0.8) | 0 (0.2) | 1 (0.3) | 10 (0.5) | 10 (0.5) | 0 (0.1) | 0 (0.1) |
| DDESS | 9 (0.8) | 8 (0.6) | 0 (0.1) | 1 (0.3) | 12 (1.1) | 10 (1.0) | 0 (0.0) | 2 (0.5) |
| DoDDS | 10 (0.6) | 8 (0.5) | 0 (0.1) | 2 (0.2) | 7 (0.4) | 6 (0.4) | 0 (0.1) | 1 (0.3) |
| Guam | 16 (0.8) | 6 (0.6) | 1 (0.3) | 9 (0.6) | 7 (0.8) | 4 (0.6) | 0 (0.0) | 3 (0.6) |

The standard errors of the estimated percentages appear in parentheses.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A). DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Table D. 2 <br> Percentages of Students Excluded From the Assessment by State, Grades 4 and 8, Public Schools Only

|  | Grade 4 |  | Grade 8 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | S1: Using Original Inclusion Criteria | S2: Using Revised Inclusion Criteria | S1: Using Original Inclusion Criteria | S2: Using Revised Inclusion Criteria |
| Nation | 6 (0.9) | 9 (1.1)* | 5 (0.7) | 5 (0.6) |
| Alabama | 7 (0.9) | 7 (0.9) | 7 (1.0) | 9 (1.1) |
| Alaska | 4 (0.5) | 5 (0.7) | 5 (0.7) | 4 (0.8) |
| Arizona | 13 (1.9) | 12 (1.4) | 8 (1.2) | 8 (1.1) |
| Arkansas | 7 (0.9) | 6 (0.8) | 7 (1.0) | 8 (1.3) |
| California | 15 (2.0) | 20 (2.4) | 10 (1.6) | 10 (0.8) |
| Colorado | 8 (1.0) | 6 (1.0) | 5 (0.7) | 5 (0.7) |
| Connecticut | 9 (0.8) | 9 (1.0) | 8 (0.7) | 9 (0.7) |
| Delaware | 7 (0.5) | 6 (0.6) | 10 (0.7) | 3 (0.3)* |
| District of Columbia | 12 (0.7) | 11 (0.6) | 9 (0.6) | 9 (0.9) |
| Florida | 10 (0.9) | 9 (1.0) | 10 (1.0) | 9 (0.7) |
| Georgia | 7 (1.0) | 8 (1.0) | 7 (0.8) | 7 (0.8) |
| Hawaii | 6 (0.8) | 6 (0.6) | 5 (0.6) | 5 (0.5) |
| Indiana | 5 (0.8) | 5 (0.9) | 6 (0.7) | 6 (0.7) |
| lowa | 6 (1.4) | 7 (0.9) | 5 (0.8) | 4 (0.6) |
| Kentucky | 6 (0.7) | 6 (1.1) | 5 (0.6) | 3 (0.6) |
| Lovisiana | 8 (1.2) | 8 (1.0) | 6 (0.9) | 6 (0.7) |
| Maine | 7 (0.8) | 9 (1.1) | 5 (0.7) | 6 (0.6) |
| Maryland | 7 (0.9) | 8 (0.9) | 7 (0.9) | 7 (0.8) |
| Massachusetts | 9 (1.3) | 9 (1.0) | 8 (1.3) | 8 (0.9) |
| Michigan | 6 (0.7) | 7 (1.1) | 5 (0.7) | 6 (0.9) |
| Minnesota | 6 (0.9) | 5 (0.6) | 3 (0.5) | 5 (0.6) |
| Mississippi | 7 (0.8) | 5 (0.6) | 7 (0.8) | 7 (0.7) |
| Missouri | 5 (0.7) | 7 (0.8) | 7 (0.7) | 5 (0.8) |
| Montana | 4 (0.7) | 5 (0.8) | 3 (0.5) | 3 (0.5) |
| Nebraska | 5 (0.9) | 6 (0.9) | 4 (1.1) | 3 (0.6) |
| Nevada | 9 (1.3) | 9 (1.0) | $\ddagger$ | $\ddagger$ |
| New Jersey | 7 (1.1) | 6 (1.0) | $\stackrel{\ddagger}{\square}$ | $\stackrel{\ddagger}{\text { ¢ }}$ |
| New Mexico | 12 (1.6) | 9 (1.0) | 8 (0.8) | 9 (0.8) |
| New York | 7 (1.0) | 10 (1.3) | 7 (1.1) | 8 (1.3) |
| North Carolina | 7 (0.9) | 8 (1.0) | 5 (0.6) | 5 (0.5) |
| North Dakota | 3 (0.5) | 4 (0.7) | 3 (0.5) | 2 (0.5) |
| Oregon | 9 (0.8) | 8 (1.1) | 4 (0.7) | 5 (0.7) |
| Pennsylvania | 5 (0.7) | 4 (0.7) | $\stackrel{\ddagger}{\ddagger}$ | $\stackrel{\ddagger}{\square}$ |
| Rhode Island | 6 (0.9) | 7 (0.9) | 7 (0.6) | 6 (0.6) |
| South Carolina | 6 (0.7) | 7 (0.9) | 6 (0.8) | 6 (0.8) |
| Tennessee | 6 (1.0) | 5 (1.1) | 4 (0.8) | 4 (1.0) |
| Texas | 10 (1.2) | 10 (1.5) | 9 (1.1) | 6 (0.9) |
| Utah | 6 (0.9) | 5 (0.8) | 6 (0.7) | 6 (0.6) |
| Vermont | 6 (0.9) | 8 (0.9) | 4 (0.6) | 6 (0.8) |
| Virginia | 7 (0.8) | 9 (0.9) | 8 (0.8) | 6 (0.8) |
| Washington | 5 (0.6) | 6 (0.8) | 6 (0.7) | 4 (0.6) |
| West Virginia | 8 (0.9) | 8 (1.1) | 9 (0.9) | 8 (0.8) |
| Wisconsin | 8 (1.0) | 7 (1.1) | 7 (0.9) | 8 (0.9) |
| Wyoming | 4 (0.7) | 3 (0.4) | 2 (0.5) | 4 (0.4) |
| DDESS | 3 (0.5) | 5 (0.7) | 3 (0.8) | 9 (1.5) |
| DoDDS | 5 (0.7) | 5 (0.6) | 3 (0.4) | 2 (0.4) |
| Guam | 12 (0.8) | 8 (0.9) | 3 (0.5) | 6 (0.8) |

The standard errors of the estimated percentages appear in parentheses.

* Indicates a significant difference between S1 and S2 results.
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
*** Sample size insufficient to permit a reliable estimate.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents Schools (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

| Table D. 3 | Percentrage of Grades 4 and 8 Public School Students with Disabiblifies Included in the Assessment | + |
| :---: | :---: | :---: |


| Nation | Grade 4 |  | Grade 8 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | S1: Using Original Inclusion Criteria | S2: Using Revised Inclusion Criteria | S1: Using Original Inclusion Criteria | S2: Using Revised Inclusion Criteria |
|  | 58 (5.5) | 46 (4.2) | 54 (4.0) | 57 (4.0) |
| Alabama | 46 (5.7) | 41 (7.0) | 47 (5.5) | 34 (5.9) |
| Alaska | 73 (3.4) | 62 (4.3) | 55 (5.8) | 72 (5.7) |
| Arizona | 32 (6.4) | 33 (5.7) | 45 (5.3) | 47 (6.1) |
| Arkansas | 31 (6.2) | 32 (5.4) | 36 (5.1) | 30 (9.0) |
| California | 40 (7.2) | 28 (7.4) | 45 (6.4) | 40 (5.9) |
| Colorado | 44 (5.0) | 61 (7.2) | 63 (4.9) | 68 (5.1) |
| Connecticut | 50 (5.1) | 42 (4.7) | 48 (4.6) | 40 (4.3) |
| Delaware | 51 (6.6) | 55 (3.5) | 32 (6.8) | 80 (3.0)* |
| District of Columbia | 17 (3.9) | 13 (2.5) | 20 (8.0) | 27 (6.8) |
| Florida | 48 (5.3) | 49 (5.3) | 41 (4.8) | 46 (3.8) |
| Georgia | 48 (6.2) | 43 (5.6) | 34 (6.5) | 39 (5.4) |
| Hawaii | 53 (4.3) | 55 (4.2) | 53 (7.0) | 59 (6.1) |
| Indiana | 54 (6.2) | 49 (6.9) | 54 (5.0) | 53 (4.4) |
| lowa | 56 (8.6) | 46 (4.3) | 59 (5.1) | 65 (4.0) |
| Kentucky | 44 (6.3) | 37 (7.8) | 51 (5.6) | 62 (6.7) |
| Louisiana | 45 (6.3) | 36 (7.7) | 36 (7.5) | 42 (6.1) |
| Maine | 49 (4.2) | 39 (6.0) | 57 (4.9) | 48 (5.0) |
| Maryland | 48 (5.6) | 44 (5.0) | 48 (6.7) | 45 (5.7) |
| Massachusetts | 51 (6.2) | 48 (5.2) | 56 (6.8) | 56 (4.2) |
| Michigan | 43 (5.4) | 40 (6.6) | 39 (6.4) | 37 (7.9) |
| Minnesota | 55 (5.5) | 58 (4.9) | 73 (4.6) | 60 (4.8) |
| Mississippi | 28 (6.4) | 37 (5.0) | 40 (5.1) | 35 (3.7) |
| Missouri | 65 (4.9) | 49 (5.3) | 41 (4.4) | 56 (5.4) |
| Montana | 51 (5.6) | 56 (6.7) | 65 (6.8) | 71 (4.5) |
| Nebraska | 69 (5.9) | 64 (5.1) | 65 (6.3) | 74 (4.0) |
| Nevada | 44 (5.2) | 40 (7.0) | $\ddagger$ | $\ddagger$ |
| New Jersey | 43 (7.2) | 44 (7.7) | $\ddagger$ | $\ddagger$ |
| New Mexico | 40 (5.7) | 50 (5.2) | 64 (4.8) | 54 (4.9) |
| New York | 46 (6.3) | 38 (5.1) | 45 (6.4) | 55 (10.1) |
| North Carolina | 48 (5.1) | 46 (6.2) | 55 (4.1) | 52 (4.3) |
| North Dakota | 67 (5.3) | 62 (5.6) | 66 (4.5) | 83 (5.1) |
| Oregon | 53 (4.6) | 53 (5.4) | 67 (4.8) | 59 (5.7) |
| Pennsylvania | 51 (6.7) | 46 (6.2) | $\ddagger$ | $\ddagger$ |
| Rhode Island | 64 (4.6) | 64 (4.2) | 59 (3.4) | 69 (3.6) |
| South Carolina | 55 (4.8) | 45 (5.6) | 43 (5.3) | 33 (4.6) |
| Tennessee | 53 (5.9) | 62 (7.2) | 62 (5.7) | 69 (6.4) |
| Texas | 43 (5.8) | 48 (5.2) | 43 (5.2) | 49 (6.1) |
| Utah | 57 (6.8) | 58 (5.3) | 51 (4.3) | 46 (4.5) |
| Vermont | 57 (4.6) | 45 (4.8) | 65 (5.3) | 50 (5.9) |
| Virginia | 49 (4.6) | 31 (4.5) | 44 (4.9) | 51 (6.3) |
| Washington | 56 (6.0) | 52 (5.1) | 55 (4.9) | 63 (5.4) |
| West Virginia | 38 (6.0) | 39 (5.6) | 33 (4.8) | 41 (5.7) |
| Wisconsin | 29 (4.3) | 42 (6.5) | 39 (6.3) | 36 (6.6) |
| Wyoming | 68 (4.6) | 70 (3.7) | 82 (4.2) | 61 (5.5) |
| DDESS | 61 (8.1) | 54 (5.4) | 71 (9.8) | 42 (15.9) |
| DoDDS | 52 (6.1) | 45 (6.7) | 65 (4.4) | 69 (7.8) |
| Guam | 70 (6.9) | 42 (14.6) | *** (***) | *** (***) |

The standard errors of the estimated percentages appear in parentheses.

* Indicates a significant difference between S1 and S2 results.
*** Sample size insufficient to permit a reliable estimate.
$\ddagger$ Indicates that the jurisdiction did not satisty one or more of the guidelines for school participation rates (see Appendix A). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.


| Grade 4 |  |  |  |
| :---: | :---: | :---: | :---: |
| Nation | 61 (8.2) | 41 | (5.3) |
| Alaska | 82 (4.3) | 84 | (6.8) |
| Arizona | 46 (7.1) | 45 | (6.7) |
| California | 53 (6.5) | 40 | (5.7) |
| District of Columbia | 24 (7.9) | 28 | (7.9) |
| Florida | 46 (9.3) | 59 | (5.7) |
| Hawaii | 71 (7.4) | 70 | (6.4) |
| Nevada | 46 (9.2) | 53 | (5.9) |
| New Mexico | 48(12.0) | 63 | (6.0) |
| Rhode Island | 69 (8.5) | 57 | (7.6) |
| Texas | 66 (6.8) | 67 | (8.3) |
| Other Jurisdictions Guam | 26(14.0) | 62 (10.3) |  |
| Grade 8 |  |  |  |
| Nation | 58 (9.8) | 63 | (5.4) |
| Arizona | 53 (9.2) | 61 | (7.0) |
| California | 51 (7.9) | 55 | (4.4) |
| New Mexico | 35(11.3) | 34 | (5.4) |
| Texas | 55 (5.4) | 79 | (6.7)* |

The standard errors of the estimated percentages appear in parentheses.

* Indicates a significant difference between S 1 and S 2 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

## Appendix E

# Discussion of the Grade 8 Asian/Pacific Islander Sample 

As noted in Chapters 2 and 3, scale score and achievement level results for eighth grade Asian/ Pacific Islander students are not included in the main body of the NAEP 1996 Mathematics Report Card. The decision to present these results in a separate appendix was made following a thorough investigation by the current NAEP grantees (Westat and ETS) ${ }^{1.2}$ into the quality and credibility of these results, as well as an independent review by a committee of statisticians from the National Institute of Statistical Sciences (NISS). ${ }^{3}$ Collateral results from the grade 8 state assessment program in mathematics suggested that the 1996 national results may substantially underestimate actual achievement of the Asian/Pacific Islander group. Because of its potential to misinform, NCES decided to omit the national grade 8 Asian/Pacific Islander results from the body of the report. The results are, however, included in this appendix along with a description of the findings that led to this decision.

Concerns about the accuracy of the grade 8 Asian/Pacific Islander results were initially noted during routine quality control of the NAEP 1996 mathematics results. Despite statistically significant gains from 1992 to 1996 in average scale scores for the nation as a whole at all three grade levels, a large apparent decline in average scores was observed for the grade 8 Asian/Pacific Islander subgroup. Table E. 1 contains average mathematics scale score estimates, and their standard errors, for the Asian/Pacific Islander subgroup for the 1990, 1992 and 1996 assessment years. From 1992 to 1996, the estimated decline in average scores for this subgroup was approximately 14 scale score points (about .4 within-grade standard deviation units) on the NAEP 500-point scale. Despite the large magnitude of this apparent decline, it is not statistically significant at the . 05 level, after controlling for multiple comparisons.

[^50]Table E. 1
Average Mathematics Scale Scores for the Grade 8 Asian/Pacific Islander Subgroup


The standard errors of the estimated percentages and average scale scores appear in parentheses.

* Indicates a significant difference from 1990.
† Indicates a significant difference from 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

The data from the NAEP state assessment program in mathematics provided an independent data source to aid in evaluating the accuracy of the national grade 8 NAEP results for Asian/Pacific Islander students as well as for other subgroups. Forty states and the District of Columbia participated in the state assessment. Results based on the combined data from these jurisdictions are quite stable in that they are based on a sample of approximately 4,000 schools and over 100,000 students. Because of the voluntary nature of the state assessment program, these aggregated state results are not nationally representative. They can, however, be compared to restricted national results, calculated using public-school data from only those states participating in the state assessment, to obtain valuable insight into the quality of the national estimates for the grade 8 race/ethnicity subgroups.

Table E. 2 contains restricted national results. Results are presented separately for four of the race/ethnicity subgroups: White, Black, Hispanic, and Asian/Pacific Islander. Aggregated state results are also presented for these same four subgroups. For three of the four subgroups, the difference between the restricted national estimates and aggregated state estimates are quite small. However, for the Asian/Pacific Islander subgroup, the difference between the two estimates, though again within reasonable bounds of sampling variability, is of considerably greater magnitude and the restricted national estimates are substantially lower than those obtained from the aggregated state data. These results suggest that the national grade 8 Asian/Pacific Islander results may substantially underestimate the performance of this subgroup. NCES was concerned that publishing the national results in the absence of the kind of discussion included in this appendix was potentially misinforming. Hence, they made the decision to omit the results from the body of the report and to include them in this appendix.


SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

It is important to note that all NAEP results are estimates and are subject to some degree of sampling variability. If different samples of schools or students had been obtained, results for some subgroups would be higher than reported here and some would be lower. In most subgroups, particularly large subgroups or subgroups for which special sampling procedures are employed, estimates of performance are likely to remain similar from one sample to another. However, the national population of Asian/Pacific Islander students is small (about 3 percent of the national population), heterogeneous with respect to academic achievement, and highly clustered in certain locations and schools - factors which are associated with large sampling variability in survey results and reflected in the large standard errors associated with performance estimates for this subgroup. Furthermore, the sampling plan for the national assessment does not include explicit stratification procedures designed to mitigate these factors. It was the judgment of all three organizations (ETS, Westat, and NISS) that investigated these results that the occurrence of this large, but statistically nonsignificant, change in the grade 8 Asian/Pacific Islander results was a consequence of these three factors: (1) the heterogeneous nature of the Asian/Pacific Islander population, (2) the current NAEP sampling design, and, (3) the sample sizes that were assessed.

NCES, working with its current NAEP contractors and other advisory groups, will continue to investigate cost-effective ways of improving the accuracy and stability of NAEP results beginning with the 1998 assessment. They will also continue to seek improvements as part of an ongoing redesign of NAEP for the year 2000 and beyond.

## Appendix $\mathbf{F}$

## Standaral Erroors

The comparisons presented in this report are based on statistical tests that consider the magnitude of the difference between group averages or percentages and the standard errors of those statistics. The following appendix contains the standard errors for the averages and percentages discussed in Chapters 2, 3, and 4.

## Chapter 2

Table F.la Scale Score Standard Errors - Grade 4
Table F.1b Scale Score Standard Errors - Grade 8
Table F.lc Scale Score Standard Errors - Grade 12
Table F.2a Scale Score Standard Errors by State - Grade 4 Public Schools
Table F.2b Scale Score Standard Errors by State - Grade 8 Public Schools

## Chapter 3

Table F.3a Achievement Level Standard Errors - Grade 4
Table F.3b Achievement Level Standard Errors - Grade 8
Table F.3c Achievement Level Standard Errors - Grade 12
Table F.4a Achievement Level Standard Errors by State - Grade 4 Public Schools
Table F.4b Achievement Level Standard Errors by State - Grade 8 Public Schools

## Chapter 4

Table F. 5 Standard Errors for the National Population Identified as SD, LEP or Both, Grades 4, 8, and 12, Public and Nonpublic Schools

Table F. 6 Standard Errors for the National Population Excluded From the Assessment, Grades 4, 8, and 12, Public and Nonpublic Schools

Table F. 7 Standard Errors for the National Population of Students with Disabilities and Limited English Proficient Students Included in the Assessment, Grades 4, 8, and 12, Public and Nonpublic Schools

| Table F. 1 a | Scale Score Standard Errors - Grade 4 |  |  |  |  | (eater |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1990 |  | 1992 |  | 1996 |  |
| Grade 4 | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average Scale Score |
| All Students |  | 0.9 |  | 0.7 |  | 0.9 |
| Region |  |  |  |  |  |  |
| Northeast | 1.0 | 2.9 | 0.9 | 2.0 | 1.2 | 2.2 |
| Southeast | 1.1 | 2.1 | 0.9 | 1.6 | 1.6 | 2.1 |
| Central | 0.8 | 1.7 | 0.5 | 1.8 | 0.7 | 1.6 |
| West | 0.8 | 2.4 | 0.7 | 1.5 | 1.8 | 2.0 |
| Gender |  |  |  |  |  |  |
| Male | 1.0 | 1.2 | 0.6 | 0.8 | 0.7 | 1.1 |
| Female | 1.0 | 1.1 | 0.6 | 1.0 | 0.7 | 1.0 |
| Race/Ethnicity |  |  |  |  |  |  |
| White | 0.2 | 1.1 | 0.2 | 0.9 | 0.4 | 0.9 |
| Black | 0.1 | 1.8 | 0.1 | 1.3 | 0.2 | 2.3 |
| Hispanic | 0.2 | 2.0 | 0.2 | 1.4 | 0.4 | 2.1 |
| Asian/Pacific Islander | 0.2 | 3.5 | 0.2 | 2.3 | 0.2 | 4.1 |
| American Indian | 0.2 | 3.9 | 0.2 | 3.1 | 0.2 | 2.3 |
| Parents' Highest Education Level |  |  |  |  |  |  |
| Did Not Finish High School Graduated from HS Some Education After HS Graduated From College I Don't Know | 0.4 | 3.7 | 0.3 | 2.5 | 0.3 | 2.5 |
|  | 0.9 | 1.5 | 0.5 | 1.5 | 0.6 | 1.6 |
|  | 0.5 | 2.5 | 0.4 | 1.5 | 0.4 | 1.5 |
|  | 1.2 | 1.5 | 1.0 | 1.0 | 1.1 | 1.3 |
|  | 1.3 | 1.2 | 0.7 | 0.8 | 1.0 | 1.4 |
| Type of School |  |  |  |  |  |  |
| Public | 1.4 | 1.1 | 1.0 | 0.8 | 1.6 | 1.0 |
| Nonpublic: | 1.4 | 2.6 | 0.8 | 1.1 | 1.6 | 1.9 |
| Catholic | 1.2 | 3.0 | 0.7 | 1.2 | 1.2 | 2.2 |
| Other Private Schools | 0.91 | 3.6 ! | 0.6 | 2.8 | $0.8!$ | 2.8 ! |
| Title I Participation | DATA NOT COLLEC |  | TED PRIOR | TO 1996 |  |  |
| Participated Did Not Participate |  |  | 1.4 1.4 |  | 1.8 0.9 |
| Free/Reduced-Price Lanch Participation |  |  |  |  |  |  |  |
| Eligible Not Eligible Information Not Available | DATA NO | T COLLECT | TED PRIOR | TO 1996 | 1.4 2.5 3.0 | 1.9 1.0 3.1 |

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.


-     - Quality control activities and special analyses involving state assessment data raised concerns about the accuracy and precision of national grade 8 Asian/Pacific results. As a result, they are omitted from the body of this report. See Appendix E for a more detailed discussion.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.

| Table F.1c | Scole Score Stondord Errors - Grode 12 |  |  |  |  | $996$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grade 12 | 1990 |  | 1992 |  |  |  |
|  | Percentage | Average Scale Score | Percentage | Average Scale Score | Percentage | Average <br> Scale Score |
| All Students |  | 1.1 |  | 0.9 |  | 1.0 |
| Region |  |  |  |  |  |  |
| Northeast | 1.2 | 2.3 | 0.6 | 1.5 | 1.3 | 2.0 |
| Southeast | 1.1 | 2.2 | 0.6 | 1.4 | 1.9 | 1.9 |
| Central | 0.8 | 2.6 | 0.6 | 1.8 | 0.8 | 2.9 |
| West | 1.2 | 2.6 | 0.9 | 1.7 | 2.0 | 1.7 |
| Gender |  |  |  |  |  |  |
| Male | 1.0 | 1.4 | 0.8 | 1.1 | 0.9 | 1.1 |
| Female | 1.0 | 1.3 | 0.8 | 1.0 | 0.9 | 1.1 |
| Race/Ethnicity |  |  |  |  |  |  |
| White | 0.6 | 1.2 | 0.6 | 0.9 | 0.5 | 1.0 |
| Black | 0.5 | 1.9 | 0.4 | 1.7 | 0.4 | 2.2 |
| Hispanic | 0.2 | 2.8 | 0.5 | 1.7 | 0.4 | 1.8 |
| Asian/Pacific Islander | 0.3 | 5.2 | 0.2 | 3.5 | 0.4 | 4.8 |
| American Indian | 0.3 ! | *** | 0.1 | *** | $0.6!$ | 8.9! |
| Parents' Highest Education Level |  |  |  |  |  |  |
| Did Not Finish High School | 0.7 | 2.1 | 0.4 | 1.7 | 0.5 | 1.8 |
| Graduated from HS | 1.1 | 2.0 | 0.8 | 1.4 | 0.8 | 1.3 |
| Some Education After HS | 1.0 | 1.2 | 0.7 | 1.0 | 0.8 | 0.8 |
| Graduated From College | 1.4 | 1.6 | 1.1 | 1.2 | 1.5 | 1.3 |
| I Don't Know | 0.3 | 4.9 | 0.3 | 3.0 | 0.2 | 2.4 |
| Type of School |  |  |  |  |  |  |
| Public | 2.0 | 1.2 | 1.2 | 1.0 | 1.6 | 0.9 |
| Nonpublic: | 2.0! | 3.6! | 1.2 | 2.3 | 1.5 | 2.2 |
| Catholic | 1.6! | $4.6!$ | 1.3 | 2.5 | 1.3 | 2.1 |
| Other Private Schools | $1.4!$ | 5.1! | 1.0! | 4.2! | 0.8 | 4.2 |
| Title I Participation | DATA NOT COLLECT |  | E PRIOR | TO 1996 |  |  |
| Participated Did Not Particpate |  |  | $0.6!$ 0.6 |  | $\begin{gathered} 3.4! \\ 1.1 \end{gathered}$ |  |
| Free/Reduced-Price Lanch Participation | DATA NOT |  |  |  |  |  |  |
| Eligible <br> Not Eligible |  | T COLLECT | ED PRIOR | TO 1996 | 1.3 3.7 | 1.6 1.3 |
| Information Not Available |  |  |  |  | 3.8 | 1.9 |

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
***Sample size insufficient to permit a reliable estimate.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.


|  | 1996 | Change from 1992 |
| :---: | :---: | :---: |
| Grade 4 |  |  |
| Nation | 1.0 | 1.3 |
| Alabama | 1.2 | 2.0 |
| Alaska $\ddagger$ | 1.3 | - |
| Arizona | 1.7 | 2.0 |
| Arkansas $\ddagger$ | 1.5 | 1.7 |
| California | 1.8 | 2.4 |
| Colorado | 1.0 | 1.4 |
| Connecticut | 1.1 | 1.6 |
| Delaware | 0.6 | 1.0 |
| District of Columbia | 1.1 | 1.2 |
| Florida | 1.2 | 1.9 |
| Georgia | 1.5 | 1.9 |
| Hawaii | 1.5 | 2.0 |
| Indiana | 1.0 | 1.5 |
| lowa $\ddagger$ | 1.1 | 1.5 |
| Kentucky | 1.1 | 1.5 |
| Louisiana | 1.1 | 1.8 |
| Maine | 1.0 | 1.4 |
| Maryland | 1.6 | 2.0 |
| Massachusetts | 1.4 | 1.8 |
| Michigan $\ddagger$ | 1.3 | 2.1 |
| Minnesota | 1.1 | 1.4 |
| Mississippi | 1.2 | 1.6 |
| Missouri | 1.1 | 1.6 |
| Montana $\ddagger$ | 1.2 | - |
| Nebraska | 1.2 | 1.7 |
| Nevada $\ddagger$ | 1.3 | - |
| New Jersey $\ddagger$ | 1.5 | 2.1 |
| New Mexico | 1.8 | 2.3 |
| New York $\ddagger$ | 1.2 | 1.8 |
| North Carolina | 1.2 | 1.6 |
| North Dakota | 1.2 | 1.4 |
| Oregon | 1.4 | - |
| Pennsylvania $\ddagger$ | 1.2 | 1.8 |
| Rhode Island | 1.4 | 2.1 |
| South Carolina $\ddagger$ | 1.3 | 1.7 |
| Tennessee | 1.4 | 1.9 |
| Texas | 1.4 | 1.8 |
| Utah | 1.2 | 1.5 |
| Vermont $\ddagger$ | 1.2 | - |
| Virginia | 1.4 | 1.9 |
| Washington | 1.2 | - |
| West Virginia | 1.0 | 1.5 |
| Wisconsin | 1.0 | 1.4 |
| Wyoming | 1.4 1.0 | 1.7 |
| DoDDS | 0.7 | - |
| Guam | 1.3 | 1.5 |

$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

- Indicates that the jurisdiction did not participate in 1992.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992 and 1996
Mathematics Assessments.


|  | 1996 | Change from 1992 | Change from 1990 |
| :---: | :---: | :---: | :---: |
| Grade 8 |  |  |  |
| Nation | 1.2 | 1.6 | 1.8 |
| Alabama | 2.1 | 2.7 | 2.4 |
| Alaska $\ddagger$ | 1.8 | - | - |
| Arizona | 1.6 | 2.0 | 2.1 |
| Arkansas $\ddagger$ | 1.5 | 1.9 | 1.8 |
| California | 1.9 | 2.5 | 2.3 |
| Colorado | 1.1 | 1.5 | 1.4 |
| Connecticut | 1.1 | 1.6 | 1.5 |
| Delaware | 0.9 | 1.4 | 1.3 |
| District of Columbia | 1.3 | 1.6 | 1.6 |
| Florida | 1.8 | 2.3 | 2.2 |
| Georgia | 1.6 | 2.0 | 2.1 |
| Hawaii | 1.0 | 1.3 | 1.3 |
| Indiana | 1.4 | 1.8 | 1.8 |
| lowa $\ddagger$ | 1.3 | 1.7 | 1.7 |
| Kentucky | 1.1 | 1.5 | 1.6 |
| Lovisiana | 1.6 | 2.3 | 2.0 |
| Maine | 1.3 | 1.6 | , |
| Maryland $\ddagger$ | 2.1 | 2.5 | 2.6 |
| Massachusetts | 1.7 | 2.0 | - |
| Michigan $\ddagger$ | 1.8 | 2.3 | 2.2 |
| Minnesota | 1.3 | 1.7 | 1.6 |
| Mississippi | 1.2 | 1.7 | - |
| Missouri | 1.4 | 1.8 | - |
| Montana $\ddagger$ | 1.3 | - | 1.6 |
| Nebraska | 1.0 | 1.5 | 1.5 |
| New Mexico | 1.2 | 1.5 | 1.4 |
| New York $\ddagger$ | 1.7 | 2.7 | 2.2 |
| North Carolina | 1.4 | 1.8 | 1.8 |
| North Dakota | 0.9 | 1.5 | 1.5 |
| Oregon | 1.5 | - | 1.8 |
| Rhode Island | 0.9 | 1.2 | 1.1 |
| South Carolina $\ddagger$ | 1.5 | 1.8 | - |
| Tennessee | 1.4 | 2.0 | - |
| Texas | 1.4 | 1.9 | 2.0 |
| Utah | 1.0 | 1.3 | - |
| Vermont $\ddagger$ | 1.0 | - | - |
| Virginia | 1.6 | 1.9 | 2.2 |
| Washington | 1.3 | - | - |
| West Virginia | 1.0 | 1.4 | 1.4 |
| Wisconsin $\ddagger$ | 1.5 | 2.1 | 2.0 |
| Wyoming | 0.9 | 1.2 | 1.1 |
| DDESS | 2.3 | - | - |
| DoDDS | 0.9 | - | - |
| Guam | 1.7 | 2.0 | 1.8 |

$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

- Indicates that the jurisdiction did not participate in 1990 and/or 1992.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.


-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.


[^51]

[^52]
$\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

- Indicates that the jurisdiction did not participate in 1992.
! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
-     -         - Standard error estimates can not be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992 and 1996 Mathematics Assessments.

## Table F.4b $\quad \begin{aligned} & \text { Achievement Level Standard Errors } \\ & \text { by State - Grade } 8 \text { Public Schools }\end{aligned}$



[^53]| Table F. 5 | Standard Errors for the National Population of Students Identified as SD, LEP, or Both Grades 4, 8, and 12, Public and Nonpublic Schools |  |
| :---: | :---: | :---: |


|  | Total | SD Only | Both SD and LEP | LEP Only |
| :--- | :---: | :---: | :---: | :---: |
| Grade 4 | 0.5 | 0.5 | 0.1 | 0.3 |
| Grade 8 | 0.6 | 0.5 | 0.1 | 0.3 |
| Grade 12 | 0.4 | 0.4 | 0.0 | 0.2 |

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.


| S1: Using <br> Original <br> Inclusion <br> Criteria | S2: Using <br> Revised <br> Inclusion <br> Criteria | S3: Using Revised <br> Criteria and <br> Providing <br> Accommodations/ <br> Adaptions |  |
| :--- | :---: | :---: | :---: |
| Gracte 4 | 0.9 | 0.9 | 0.7 |
| Grade 8 | 0.6 | 0.5 | 0.4 |
| Grade 12 | 0.4 | 0.4 | 0.4 |

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

| Standard <br> Students <br> Table F. 7 <br> Proficient <br> Grades 4, | Standard Errors for the National Population of Students with Disabilities and Limited English Proficient Students Induded in the Assessment, Grades 4, 8, and 12, Public and Nonpublic Schools |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | SD |  |  | LEP |  |  |
|  | S1: Using Original Incusion Criteria | $\begin{aligned} & \text { S2: Using } \\ & \text { Revised } \\ & \text { Inclusion } \\ & \text { Criteria } \end{aligned}$ | S3: Using Revised Criteria And Providing Accommodations/ Adaptations | S1: Using Original Inclusion Criteria | $\begin{aligned} & \text { S2: Using } \\ & \text { Revised } \\ & \text { Inclusion } \\ & \text { Criteria } \end{aligned}$ | S3: Using Revised Criteria And Providing Accommodations Adaptations |
| Grade 4 |  |  |  |  |  |  |
| Assessed Under Standard Conditions <br> Assessed With Accommodations | 5.4 | 4.0 | 3.9 | 8.2 | 5.3 | 6.7 |
|  |  |  | 5.3 |  |  | 7.2 |
| Total Assessed | 5.4 | 4.0 | 4.3 | 8.2 | 5.3 | 7.0 |
| Grade 8 |  |  |  |  |  |  |
| Assessed Under Standard Conditions <br> Assessed With Accommodations | 4.0 | 4.4 | 4.2 | 9.6 | 5.4 | 4.0 |
|  |  |  | 3.5 |  |  | 4.6 |
| Total Assessed | 4.0 | 4.4 | 3.4 | 9.6 | 5.4 | 4.4 |
| Grade 12 |  |  |  |  |  |  |
| Assessed Under Standard Conditions Assessed With Accommodations | 5.1 | 4.1 | 5.0 | 6.5 | 6.8 | 4.0 |
|  |  |  | 3.8 |  |  | 2.0 |
| Total Assessed | 5.1 | 4.1 | 5.1 | 6.5 | 6.8 | 3.4 |

[^54]
## Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 1996 mathematics assessment. The NAEP 1996 mathematics assessment was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). In addition, the program benefited from the contributions of hundreds of individuals at the state and local levels - governors, chief state school officers, state and district test directors, state coordinators, and district administrators - who tirelessly provided their wisdom, experience, and hard work. Most importantly, NAEP is grateful to the thousands of students and hundreds of teachers and administrators who made the assessment possible.

The NAEP 1996 mathematics assessment was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. The Commissioner of Education Statistics, Pascal D. Forgione, and the NCES staff - Sue Ahmed, Peggy Carr, Arnold Goldstein, Steven Gorman, Larry Ogle, Gary W. Phillips, Sharif Shakrani, Maureen Treacy, and Alan Vanneman - worked closely and collegially with the authors to produce this report. The authors were also provided invaluable advice and guidance by the members of the National Assessment Governing Board and NAGB staff. In particular, the authors are indebted to Arnold Goldstein of NCES for his daily efforts to coordinate the activities of the many people who contributed to this report.

The NAEP project at ETS is housed in the Center for the Assessment of Educational Progress under the direction of Paul Williams. The NAEP 1996 assessments were directed by Stephen Lazer and John Mazzeo. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Diane Walsh. Printing, distribution, scoring, and processing activities were conducted by NCS under the direction of Brad Thayer, Patrick Bourgeacq, Charles Brungardt, Mathilde Kennel, Linda Reynolds, and Connie Smith.

The complex statistical and psychometric activities necessary to report results for the NAEP 1996 mathematics assessment were directed by Nancy Allen, John Barone, James Carlson, and Juliet Shaffer. The analyses presented in this report were led by Frank Jenkins and Edward Kulick with assistance from Hua-hua Chang, David Freund, Shuyi Hua, Bruce Kaplan, Jiahe Qian, Steve Wang, and Xiaohui Wang.

The design and production of the report was overseen by Carol Errickson, Karen Miller, and Clyde Reese. The considerable production efforts were completed by Loretta Casalaina, Kelly Gibson, Sharon Davis-Johnson, Alice Kass and Barbette Tardugno. Editorial assistance was provided by John Calderone of Aspen Systems.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in this report: Sue Ahmed, Wanda Chambers, Shi-Chang Wu, Pasquale DeVito, Lawrence Feinberg, Mary Frase, Jim Houser, Andrew Kolstad, Milagros Lanauze, Michelle Leon, Judith Leonard, David Malouf, John Olson, Mary Rollefson, and Sharif Shakrani.
$\qquad$


[^0]:    ${ }^{1}$ The Executive Summary for this report was prepared by Alan Vanneman of the Education Statistics Services Institute.

[^1]:    2 In all discussions of differences in mathematics performance, either over time or between subgroups, only statistically significant differences are reported. Such differences are unlikely to be due to chance factors.

[^2]:    ${ }^{1}$ Executive Office of the President, National goals for education (Washington, DC: Government Printing Office, 1990); Goals 2000: Educate America Act, L. 103-227 (1994).
    ${ }^{2}$ Curriculum and evaluation standards for school mathematics (Reston, VA: National Council of Teachers of Mathematics, 1989).

[^3]:    ${ }^{3}$ Education achievement standards: NAGB's approach yields misleading interpretations (United States General Accounting Office Report to Congressional Requestors. Washington, DC: United States General Accounting Office, 1993). Setting performance standards for student achievement. A report of the National Academy of Education Panel on the evaluation of the NAEP Trial State Assessment: An evaluation of the 1992 achievement levels (Stanford, CA: National Academy of Education, 1993).
    ${ }^{4}$ Cizek, G., Reactions to National Academy of Education report (Washington, DC: National Assessment Governing Board., 1993)

    Kane, M., Comments on the NAEP evaluation of the NAGB achievement levels (Washington, DC: National Assessment Governing Board, 1993).
    NAEP reading revisited: An evaluation of the 1992 achievement levels descriptions (American College Testing, Washington, DC: National Assessment Governing Board, 1993).
    Technical report on setting achievement levels on the 1992 National Assessment of Educational Progress in mathematics, reading, and writing. (American College Testing, Washington, DC: National Assessment Governing Board, 1993).

[^4]:    ${ }^{5}$ The procedures used to develop the item maps are detailed in the forthcoming NAEP 1996 Technical Report.
    ${ }^{6}$ For constructed-response questions a criterion of 65 percent was used. For multiple-choice questions with four or five alternatives, the criteria were 74 and 72 percent, respectively. The use of a higher criteria for multiple-choice questions reflected students ability to "guess" the correct answer from among the alternatives.

[^5]:    *NAEP mathematics composite scale range
    ***Sample size insufficient to permit reliable estimates

[^6]:    * NAEP mathematics composite scale range

[^7]:    *NAEP mathematics composite scale range
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Apppendix A).

[^8]:    *NAEP mathematics composite scale range
    *** Sample size insufficient to permit reliable estimates

[^9]:    7 All differences reported are statistically significant at the 0.05 level with appropriate adjustments for multiple comparisons.

[^10]:    ${ }^{8}$ Campbell, J.R.; Reese, C.M.; O’Sullivan, C.; \& Dossey, J.A., NAEP 1994: Trends in academic progress (Washington, DC: National Center for Education Statistics, 1996).
    ${ }^{9}$ Meyer, M., Gender differences in mathematics, in M.M. Lindquist (Ed.), Results from the fourth mathematics assessment of the NAEP (Reston, VA: National Council of Teachers of Mathematics, 1989).

[^11]:    ${ }^{10}$ Campbell, J.R.; Reese, C.M.; O’Sullivan, C.; \& Dossey, J.A., NAEP 1994: Trends in academic progress (Washington, DC: National Center for Education Statistics, 1996).
    ${ }^{11}$ Mullis, I.V.S.; Jenkins, F.; \& Johnson, E.G., Effective schools in mathematics (Washington, DC: National Center for Education Statistics, 1994).
    ${ }^{12}$ Oakes, J., Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science (Santa Monica, CA: RAND, 1990).
    ${ }^{13}$ Backer, A., \& Akin, S. (Eds.), Every child can succeed: Readings for school improvement (Bloomington, IN: Agency for Instructional Television, 1993).

[^12]:    ${ }^{14}$ Stevenson, H.W., \& Newman, R.S., Long-term prediction of achievement and attitudes in mathematics and reading, Child Development, 57, pp. 646-659 (1986); National Education Longitudinal Study, National Education Longitudinal Study of 1988: Base year student survey (Washington, DC: National Center for Education Statistics, 1995).
    ${ }^{15}$ Looker, E.D., Accuracy of proxy reports of parental status characteristics, in Sociology of Education, 62(4), pp. 257-276 (1989).

[^13]:    ${ }^{16}$ Campbell, J.R.; Reese, C.M.; O'Sullivan, C.; \& Dossey, J.A., NAEP 1994: Trends in Academic Progress (Washington, DC: National Center for Education Statistics, 1996); National Education Longitudinal Study of 1988: Base year student survey (Washington, DC: National Center for Education Statistics, 1995).

[^14]:    ${ }^{17}$ U.S. Department of Education, Office of Elementary and Secondary Compensatory Education Programs, Improving basic programs operated by local education agencies (Washington, DC: U.S. Department of Education, 1996).
    ${ }^{18}$ For a study of mathematics performance of Title I students in 1991-92, see U.S. Department of Education, PROSPECTS: The Congressionally Mandated Study of Educational Growth and Opportunity, Interim Report: Language, Minority and Limited English Proficient Students (Washington, DC: U.S. Department of Education, 1995).

[^15]:    ${ }^{19}$ U.S. General Services Administration, Catalog of federal domestic assistance (Washington, DC: Executive Office of the President, Office of Management and Budget, 1995).

[^16]:    ${ }^{20}$ The 1996 results for eighth-grade Asian/Pacific Islander students are not included in this chapter. Following a thorough investigation into the quality and credibility of these results, NCES decided to omit these results from the body of the report and to include them in an appendix. (See Appendix E for further discussion.)

[^17]:    ${ }^{21}$ The percentages of students at or above Basic did not sifgnificantly change between 1990 and 1996 for the following: did not finish high school (grades 4 and 12); some education after high school (grade 4); and graduated from college (grade 8).

[^18]:    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^19]:    ${ }^{22}$ Kaplan, B.A. \& Leung, P.T. Statistical Summary of the 1994 NAEP Samples. In N. Allen, D. Kline, \& C. Zelenak (eds), The NAEP 1994 Technical Report. (Washington, D.C.: National Center for Education Statistics, 1996).
    ${ }^{23}$ Olson, J.F. \& A.A. Goldstein. Increasing the inclusion of students with disabilities and limited english proficient students in NAEP. Focus on NAEP, 2(1). (Washington, D.C.: National Center for Education Statistics, 1996).
    ${ }^{24}$ National Academy of Education. The Trial State Assessment: Prospects and Realities. The Third Report of the National Academy of Education Panel on the Evaluation of the NAEP 1992 Trial State Assessment. (Stanford, CA: National Academy of Education, 1993).
    ${ }^{25}$ Ysseldyke, J.E., M.L. Thurlow, K.S. McGrew, \& M. Vanderwood. Making decisions about the inclusion of students with disabilities in statewide assessments (Synthesis Report 13). ( Minneapolis: University of Minnesota, National Center on Education Outcomes, 1994).
    ${ }^{26}$ Olson, J.F. \& A.A. Goldstein. Increasing the inclusion of students with disabilities and limited english proficient students in NAEP. Focus on NAEP, 2(1). ( Washington, D.C.: National Center for Education Statistics, 1996).

[^20]:    ${ }^{27}$ Throughout this chapter, results from the state assessment are limited to public school students. State-level samples of nonpublic school students were relatively modest in size and, for a substantial number of jurisdictions, did not meet minimum NCES participation rate guidelines established for the reporting of results (see Appendix A). Hence they were excluded from this chapter and from Appendix D in the interests of clarity and brevity.

[^21]:    ${ }^{28}$ Appendix D presents comparable results for the small number of jurisdictions participating in the state assessment with sufficient sample sizes of LEP students.

[^22]:    ${ }^{1}$ National Assessment Governing Board. Mathematics Framework for the 1996 National Assessment of Educational Progress. (Washington, DC: NAGB, 1994)
    ${ }^{2}$ National Council of Teachers of Mathematics. Curriculum and Evaluation Standards for School Mathematics. (Reston, VA: NCTM, 1989).

[^23]:    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^24]:    National results are based on the national assessments samples, not on aggregated state assessment program samples.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).
    DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
    DoDDS: Department of Defense Dependents Schools (Overseas)
    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[^25]:    ${ }^{3}$ For additional information about the use of weighting procedures in NAEP, see Johnson, E.G. (1989, December). Journal of Educational Statistics, 14(4), 303-334.

[^26]:    ${ }^{4}$ For theoretical and empirical justification of the procedures employed, see Mislevy, R.J. (1988). Randomization-based inferences about latent variables from complex samples. Psychometrika, 56(2), 177-196.
    For computational details, see National Assessment of Educational Progress. (1990). Focusing the new design: NAEP 1988 technical report, and the 1990 NAEP technical report. Princeton, NJ: Educational Testing Service.
    ${ }^{5}$ Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16(2), 159-176.
    ${ }^{6}$ Donoghue, J.R. (1994). An empirical examination of the IRT information of polytomously scored reading items under the generalized partial credit model. Journal of Educational Measurement, 31(4), 295-311.
    Muraki, E. (1993). Information functions of the generalized partial credit model. Applied Psychological Measurement, 17(4), 351-363.

[^27]:    ${ }^{7}$ For the national assessment, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area.). For the state assessment program, a PSU is most often a single school.

[^28]:    ${ }^{8}$ The procedure for assigning race/ethnicity was modified for Hawaii. See the forthcoming NAEP 1996 Technical Report for the State Assessment Program in Mathematics for details.

[^29]:    ${ }^{9}$ For further details, see Johnson, E.G., \& Rust, K.F. Population inferences and variance estimation for NAEP data, Journal of Educational Statistics 17(2) (1992) 175-190.

[^30]:    ${ }^{10}$ Miller, R.G. Simultaneous Statistical Inference. (New York: Wiley, 1966).
    ${ }^{11}$ Benjamin and Hochberg. False Discovery Rate (FDR) Procedure. Journal of the Royal Statistical Society, Series B, No. 1. (1995, 289-300).
    ${ }^{12}$ Williams, V. S. L., L. V. Jones, and J. W. Tukey. Controlling Error in Multiple Comparisons with Special Attention to the National Assessment of Educational Progress. (Research Triangle Park, NC: National Institute of Statistical Sciences, December 1994).

[^31]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^32]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^33]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^34]:    National results are based on the national assessment samples, not on aggregated state assessment program samples
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A)
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).
    --Quality control activities and special analyses involving state assessement data raised concerns about the accuracy and precision of national grade 8 Asian/Pacific results. As a result, they are omitted from the body of this report. See Appendix A for a more detailed discussion.

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^35]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^36]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates. $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^37]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^38]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^39]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^40]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^41]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^42]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^43]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    ***Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^44]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^45]:    National results are based on the national assessment samples, not on aggregated state assessment program samples.
    *** Sample size insufficient to permit reliable estimates.
    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessments.

[^46]:    ${ }^{1}$ U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-25, No. 1095 at the national level and forthcoming state-level P-25 Reports.
    ${ }^{2}$ U.S. Department of Education, National Center of Education Statistics, Common Core of Data surveys.
    Information reprinted from the Digest of Education Statistics 1996 (NCES 96-133).

[^47]:    ${ }^{3}$ U.S. Department of Commerce, Bureau of the Census, Decennial Census, Minority Economic Profiles, unpublished data, and Current Population Reports, Series P-60, "Poverty in the United States," "Money Income of Households, Families, and Persons in the United States," and "Income, Poverty, and Valuation of Noncash Benefits," various years.
    ${ }^{4}$ U.S. Department of Education, Office of Special Education and Rehabilitative Services, Annual Report to Congress on the Implementation of The Individuals with Disabilities Education Act, various years, and unpublished tabulations.
    Information reprinted from the Digest of Education Statistics 1996 (NCES 96-133).

[^48]:    ${ }^{5}$ U.S. Department of Commerce, Bureau of the Census, Government Division, Government Finances: 1991-92, Series GF/92-5.

[^49]:    ${ }^{1}$ Note that the impact of providing accommodations is not discussed because they were not provided at the state level in 1996.
    ${ }^{2}$ Grade 8 sample sizes in Guam for students with disabilities are not sufficiently large in either the S1 or S2 samples to reliably report on inclusion rates.

[^50]:    ${ }^{1}$ Carlson, J. \& Williams, P. (1996, October 29) ETS/NAEP Technical Memorandum on 1996 Mathematics Grade 8 results for Asian/Pacific Island Subpopulation.
    ${ }^{2}$ Rust, K (1996, November 1) Westat Memorandum to Gary Phillips on 1996 Mathematics Grade 8 Results for Asian and Pacific Islander Students.
    ${ }^{3}$ Letter from Jerome Sacks to Gary Phillips, dated November 21, 1996.

[^51]:    -     - Quality control activities and special analyses involving state assessment data raised concerns about the accuracy and precision of national grade 8 Asian/Pacific results. As a result, they are omitted from the body of this report. See Appendix E for a more detailed discussion.
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.

[^52]:    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A)
    *** Sample size insufficient to permit a reliable estimate.

    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP),1990, 1992 and 1996 Mathematics Assessments.

[^53]:    $\ddagger$ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

    - Indicates that the jurisdiction did not participate in 1990 and/or 1992
    ! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (See Appendix A).
    -     -         - Standard error estimates can not be accurately determined.

    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992 and 1996 Mathematics Assessments.

[^54]:    SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

