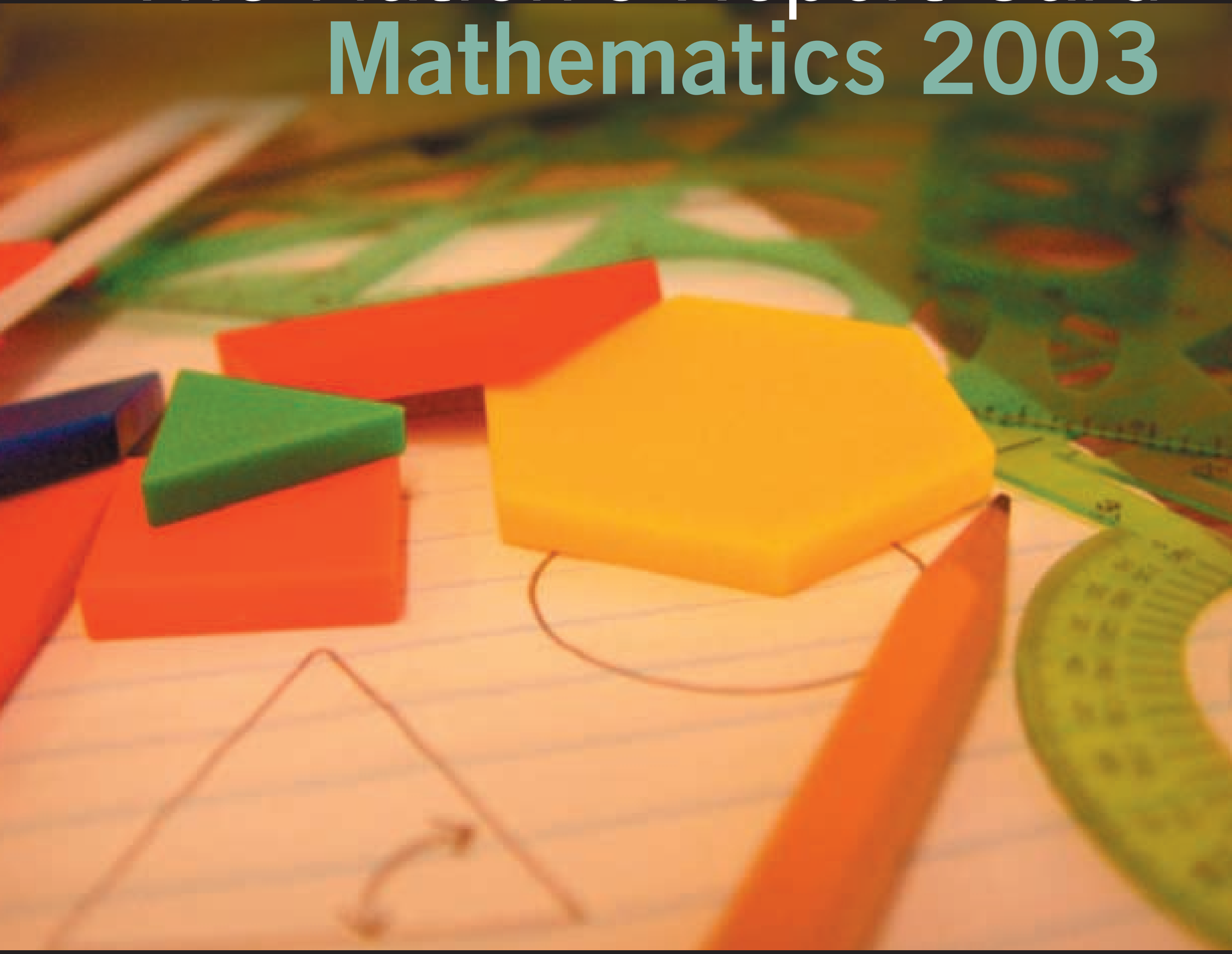
NCES
National Center for Education Statistics

The Nation's Report Card NAEP™

U.S. Department of Education
Institute of Education Sciences
NCES 2005–451

# The Nation's Report Card™
# Mathematics 2003



The National Assessment of Educational Progress

## What is The Nation's Report Card™?

THE NATION'S REPORT CARD™, the National Assessment of Educational Progress (NAEP), is a 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 within the Institute of Education Sciences of 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.

In 1988, Congress established the National Assessment Governing Board (NAGB) to oversee and set policy for NAEP. The Board is responsible for: selecting the subject areas to be assessed; setting appropriate student achievement levels; developing assessment objectives and test specifications; developing a process for the review of the assessment; designing the assessment methodology; developing guidelines for reporting and disseminating NAEP results; developing standards and procedures for interstate, regional, and national comparisons; determining the appropriateness of all assessment items and ensuring the assessment items are free from bias and are secular, neutral, and non-ideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of NAEP reports.

## The National Assessment Governing Board

National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2005-451

# The Nation's Report Card™
# Mathematics 2003

James S. Braswell
Gloria S. Dion
Mary C. Daane
Ying Jin

*In collaboration with*
Scott Davis
Amy Dresher
Frank Jenkins
Edward Kulick
Brenda Tay-Lim
Mei Jang Lin
Yuxin Tang
**Educational Testing Service**

Arnold A. Goldstein
*Project Officer*
**National Center for Education Statistics**

**U.S. Department of Education**
Margaret Spellings
*Secretary*

**Institute of Education Sciences**
Grover J. Whitehurst
*Director*

**National Center for Education Statistics**
Grover J. Whitehurst
*Acting Commissioner*

**July 2005**

SUGGESTED CITATION
Braswell, J.S., Dion, G.S., Daane, M.C., and Jin, Y. (2005). *The Nation's Report Card: Mathematics 2003.* (NCES 2005–451). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

FOR MORE INFORMATION
Content contact:
Arnold Goldstein
202-502-7344
Arnold.Goldstein@ed.gov

To obtain single copies of this report, or for ordering information on other U.S. Department of Education products, call toll free 1-877-4ED-PUBS (877-433-7827), or write:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794-1398

TTY/TDD 1-877-576-7734
FAX 1-301-470-1244

Online ordering via the Internet: http://www.edpubs.org
Copies also are available in alternate formats upon request.
This report also is available on the World Wide Web: http://nces.ed.gov/pubsearch

The work upon which this publication is based was performed for
the National Center for Education Statistics by Educational Testing Service,
Pearson Educational Measurement, and Westat.

# Table of Contents

### Chapter 1: Figures



### Chapter 2: Tables and Figures

**Chapter 3: Tables and Figures**

**Chapter 4: Tables and Figures**



**Chapter 5: Tables and Figures**

**Appendix A: Tables and Figures**

**Appendix B: Tables**

**Appendix C: Tables**

# Executive Summary

The National Assessment of Educational Progress (NAEP) is an ongoing nationally representative sample survey of student achievement in core subject areas. Authorized by Congress and administered by the National Center for Education Statistics (NCES), within the Institute of Education Sciences of the U.S. Department of Education, NAEP regularly reports to the public on the educational progress of fourth-, eighth-, and twelfth-grade students.

This report presents results of the NAEP 2003 fourth- and eighth-grade mathematics assessments for the nation, for regions of the country, for participating states and other jurisdictions, and for participating urban districts. Assessment results are described in terms of students' average mathematics score on a 0–500 scale and in terms of the percentage of students attaining each of three achievement levels: *Basic, Proficient,* and *Advanced.*

The achievement levels are performance standards adopted by the National Assessment Governing Board (NAGB) as part of its statutory responsibilities. The achievement levels are a collective judgment of what students should know and be able to do for each grade tested. The law requires that the achievement levels are to be used on a trial basis until the Commissioner of Education Statistics determines "that such levels are reasonable, valid, and informative to the public."[1] Until that determination is made, the law requires the Commissioner and the Board to state clearly the trial status of the achievement levels in all NAEP reports. However, both NCES and NAGB believe these performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials and others as a common yardstick of academic performance.

Approximately 190,000 fourth-graders from 7,500 schools and 153,000 eighth-graders from 6,100 schools were assessed in 2003. The national results reflect the performance of students attending both public and nonpublic schools, while the results for participating states and jurisdictions, and for participating urban districts, reflect the performance of students attending public schools. In addition to providing average scores and achievement-level percentages in mathematics for the nation, states and other jurisdictions, and selected urban districts, this report provides results for subgroups of students defined by various background characteristics.

A summary of major findings from the NAEP 2003 Mathematics Assessment is presented on the following pages. Comparisons are made to results from previous years in which the assessment was administered. In addition to the 2003 results, national results are reported from the 1990, 1992, 1996, and 2000 assessments. Results for states and other jurisdictions are also reported from the 1990 (eighth grade only), 1992, 1996, 2000, and 2003 assessments. Results for participating urban districts are reported for 2003.

The more recent results, from 2000 and 2003, are based on more inclusive samples using administration procedures in which testing accommodations were permitted for students with disabilities and limited-English-proficient students. Accommodations were not permitted in earlier assessments. Comparisons between results from 2003 and those from 2000, in which both types of administration procedures were used, are discussed in this executive summary based on the results when accommodations were permitted.

Changes in student performance across years or differences between groups of students in 2003 are discussed only if they have been determined to be statistically significant at the .05 level based on t-tests adjusted using the False Discovery Rate (FDR) multiple comparison procedure. Beginning with the reading sample in 2002, the NAEP national samples were obtained by aggregating the samples from each state, rather than obtaining an independently selected national sample. As a result, the size of the national sample increased and smaller differences between years or between subgroups of students were found to be statistically significant than would have been detected in previous assessment years.

[1] No Child Left Behind Act of 2001, P. L. 107-110, 115 Stat. 1425 (2002).

## Overall Mathematics Results for the Nation, Regions of the Country, and States and Other Jurisdictions

### Mathematics Results for the Nation

#### *At grade 4*

- The average fourth-grade mathematics score was higher in 2003 than in all the previous assessment years.
- Scores at the 10th, 25th, 50th, 75th, and 90th percentiles were higher in 2003 than in any of the previous assessment years, indicating improvement for lower-, middle-, and higher-performing students. Gains detected between 2000 and 2003 ranged from approximately 5 scale score points for students performing at the 90th percentile to 13 points for students at the 10th percentile.
- In 2003, 32 percent of fourth-graders performed at or above the *Proficient* level. The percentages of fourth-graders performing at or above *Basic*, at or above *Proficient*, and at *Advanced* increased between 2000 and 2003, and were higher in 2003 than in 1990. The percentage at or above *Proficient* increased by approximately 19 points between 1990 and 2003.

#### *At grade 8*

- The average eighth-grade mathematics score was higher in 2003 than in all previous assessment years.
- Scores at the 10th, 25th, 50th, 75th, and 90th percentiles were higher in 2003 than in any of the previous assessment years, indicating improvement for lower-, middle-, and higher-performing students. Increases detected between 2000 and 2003 ranged from approximately 3 scale score points at the 90th percentile to 7 points at the 10th percentile.
- In 2003, 29 percent of eighth-graders performed at or above the *Proficient* level. The percentages of eighth-graders performing at or above *Basic* and at or above *Proficient* increased between 2000 and 2003, and were higher in 2003 than in 1990. The percentage at or above *Proficient* increased by approximately 14 points between 1990 and 2003.

### Mathematics Results for Regions of the Country

Prior to 2003, NAEP results were reported for four NAEP-defined regions of the nation: Northeast, Southeast, Central, and West. As of 2003, to align NAEP with other federal data collections, NAEP analysis and reports have used the U.S. Census Bureau's definition of "region." The four regions defined by the U.S. Census Bureau are Northeast, South, Midwest, and West.

#### *At grade 4*

- The average fourth-grade mathematics score was higher for students in the Northeast and Midwest than for students in the South and West. The average score for students in the South was higher than for students in the West.
- The percentages of fourth-graders performing at or above the *Basic* and *Proficient* levels were higher in the Northeast and Midwest than in the South and West. Higher percentages of students performed at or above *Basic* and *Proficient* in the South than in the West.

### *At grade 8*

- The average eighth-grade mathematics score was higher for students in the Northeast and Midwest than for students in the South and West. The average score was higher for students in the South than for students in the West.
- Higher percentages of eighth-grade students performed at or above *Basic* and *Proficient* in the Northeast and Midwest than in the South and West. A higher percentage of eighth-graders performed at or above *Basic* in the South than in the West.

## Mathematics Results for the States and Other Jurisdictions

Results from the 2003 assessment are reported for fourth- and eighth-grade students attending public schools only in 50 states and 3 other jurisdictions that participated in the assessment. (Throughout this report, the term "jurisdiction" is used to refer to the states, the District of Columbia, and the two Department of Defense system schools that participated in the NAEP mathematics assessment.)

### *At grade 4*

- All 43 jurisdictions that participated in both the 2000 and 2003 fourth-grade assessments showed increases in average scores. Similarly, each of the 42 jurisdictions that participated in the 1992 and 2003 assessments had a higher average score in 2003.
- Kansas, Massachusetts, Minnesota, New Hampshire, North Carolina, Vermont, and Wyoming were among the jurisdictions with the highest average scores. Average fourth-grade scores in Connecticut and Virginia were lower only in comparison with New Hampshire.
- The percentage of fourth-graders performing at or above *Proficient* was higher in 2003 than in 2000 for all 43 jurisdictions that participated in both years. The percentage of fourth-graders at or above *Proficient* was higher in 2003 than in 1992 for all 42 jurisdictions that participated in both years.

### *At grade 8*

- Of the 42 jurisdictions that participated in both the 2000 and 2003 eighth-grade mathematics assessments, 28 had a higher average score in 2003. Each of the 38 jurisdictions that participated in both the 1990 and 2003 assessments had a higher average score in 2003.
- In 2003, Minnesota had the highest average mathematics score at grade 8. Eighth-graders in Department of Defense Overseas schools, Kansas, Massachusetts, Montana, New Hampshire, North Dakota, South Dakota, and Vermont all had higher average scores than the remaining jurisdictions except Minnesota.
- Among the 42 jurisdictions that participated in both the 2000 and 2003 eighth-grade assessments, 18 showed an increase in the percentage of students performing at or above *Proficient*. The percentage of eighth-graders at or above *Proficient* was higher in 2003 than in 1990 for all 38 jurisdictions that participated in both years.

## Mathematics Results for Student Subgroups in the Nation and in the States and Other Jurisdictions

In addition to overall results, NAEP reports on the performance of various subgroups of students. In interpreting these data, readers are reminded that the relationship between contextual variables and student performance is not necessarily causal. There are many factors that play a role in student achievement in a particular subject area.

### National Results

#### *Gender*

- At both grades 4 and 8, the average scores for both male students and female students were higher in 2003 than in any of the previous assessment years.
- In 2003, male students outperformed female students by 3 points on average at grade 4 and by 2 points on average at grade 8. The male-female gap in 2003 was not measurably different from the gap in any of the previous assessment years since 1990 for either grade.
- At both grades 4 and 8, the percentages of male students and female students performing at or above *Proficient* were higher in 2003 than in any previous assessment year.

#### *Race/Ethnicity*

Based on information obtained from school records, students who took the NAEP mathematics assessment were identified as belonging to one of five mutually exclusive racial/ethnic subgroups: White, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, or Other.

- At both grades 4 and 8, Asian/Pacific Islander students scored higher on average than White students in 2003. Both White students and Asian/Pacific Islander students had higher average scores than Black, Hispanic, and American Indian/Alaska Native students at both grades. Hispanic students and American Indian/Alaska Native students also scored higher on average than Black students at both grades.
- At grade 4, White, Black, and Hispanic students all had higher average scores in 2003 than in any of the previous assessment years. American Indian/Alaska Native fourth-graders had a higher average score in 2003 than in 2000. The average score for Asian/Pacific Islander fourth-graders was higher in 2003 than in 1990.
- At grade 8, White, Black, and Hispanic students all showed increases in average scores between 2000 and 2003. The average score for Asian/Pacific Islander eighth-graders was higher in 2003 than in 1990.

- At grade 4, the score gap between White students and Black students decreased between 2000 and 2003, and was smaller in 2003 than in 1990. The gap between White fourth-graders and Hispanic fourth-graders also narrowed between 2000 and 2003, but there was no measurable difference between the gap in 2003 and the gap in 1990.
- At grade 8, the score gap between White students and Black students was narrower in 2003 than in 2000, but the gap in 2003 was not measurably different from that in 1990.
- At both grades 4 and 8, the percentage of students performing at or above the *Proficient* level was higher in 2003 than in any of the previous assessment years for White, Black, and Hispanic students. The percentage of Asian/Pacific Islander students performing at or above *Proficient* was higher in 2003 than in 1990.

### Eligibility for Free/Reduced-Price School Lunch

NAEP collects data on students' eligibility for free/reduced-price lunch as an indicator of family economic status. Eligibility for free/reduced-price lunch is determined by students' family income in relation to the federally established poverty level. The mathematics results since 1996 are reported for students classified by their eligibility.

- In 2003, the average mathematics scores for fourth- and eighth-graders who were eligible for free/reduced-price lunch were lower than that for students who were not eligible.
- For students who were eligible and those who were not eligible, the average mathematics scores for fourth-grade and eighth-grade students increased between 2000 and 2003 and were higher in 2003 than in 1996.
- At both grades 4 and 8, the percentage of students at or above *Proficient* was higher in 2003 than in 2000 and 1996 for both students who were eligible and those who were not eligible.

### Parents' Level of Education

Eighth-grade students who participated in the NAEP mathematics assessment were asked to indicate the highest level of education completed by each parent. Information about parental education was not collected at grade 4. Results are reported based on the highest level of education reported for either parent.

- Overall, in 2003, there was a positive relationship between student-reported parental education and student achievement: the higher the parental education level, the higher the average mathematics score.
- Average scores for eighth-grade students increased from 2000 to 2003 and were higher in 2003 than in 1990 for each level of parental education reported.
- The percentage of eighth-graders performing at or above *Proficient* was higher in 2003 than in 1990 regardless of the level of parental education students reported.

### *Type of School*

The schools that participate in the NAEP assessment are classified as either public or nonpublic. A further distinction is then made between nonpublic schools that are Catholic schools and those that are some other type of nonpublic school.

- In 2003, fourth- and eighth-grade students in nonpublic schools had higher average scores than students in public schools. Eighth-grade students in Catholic schools had lower average scores than eighth-graders in other nonpublic schools.
- At both grades 4 and 8, the average mathematics scores for students in public and nonpublic schools (including Catholic and other nonpublic schools) increased from 2000 to 2003 and were higher in 2003 than in 1990.
- The percentages of fourth- and eighth-graders performing at or above *Proficient* were higher in 2003 than in 1990 for students in public schools, Catholic schools, and other nonpublic schools.

### *Type of Location*

The schools from which NAEP draws its samples of students are classified according to their type of location (central city, rural/small town, or urban fringe/large town). The methods used to identify the type of school location in 2000 and 2003 were different from those used for prior assessment years; therefore, only the data from the 2000 and 2003 assessments are reported.

- In 2003, fourth- and eighth-grade students in schools located in urban fringe/large town and rural/small town locations had higher average mathematics scores than those in central city locations, and students in urban fringe/large town locations scored higher on average than students in rural/small town locations.
- The average mathematics scores in all three location types were higher in 2003 than in 2000 for both grades 4 and 8.
- The percentage of students at or above *Proficient* increased between 2000 and 2003 in all three types of locations at grade 4.

## State and Other Jurisdiction Results

### *Gender*

- In 2003, male fourth-graders scored higher on average than female fourth-graders in 24 jurisdictions. At grade 8, the average score for male students was higher than for female students in Massachusetts, South Carolina, and Department of Defense Overseas schools.
- The average scores increased between 1992 and 2003 for both male and female fourth-graders in all 42 of the jurisdictions that participated in both assessments. For the 38 jurisdictions that participated in both the 1990 and 2003 eighth-grade assessments, 36 showed increases for both male and female students and Montana and North Dakota showed increases only for female students.

### *Race/Ethnicity*

- At grade 4, average scores were higher in 2003 than in 1992 for White students in 42 jurisdictions, for Black students in 35 jurisdictions, for Hispanic students in 20 jurisdictions, for Asian/Pacific Islander students in 11 jurisdictions, and for American Indian/Alaska Native students in 3 jurisdictions.
- At grade 8, average scores were higher in 2003 than in 1990 for White students in 37 jurisdictions, for Black students in 25 jurisdictions, for Hispanic students in 12 jurisdictions, for Asian/Pacific Islander students in 7 jurisdictions, and for American Indian/Alaska Native students in 5 jurisdictions.

### *Eligibility for Free/Reduced-Price School Lunch*

- In 2003, students who were eligible for free/reduced-price lunch scored lower on average than students who were not eligible in all 52 jurisdictions for which data are available at grade 4 and in 51 of the 52 jurisdictions for which data are available at grade 8.
- The average fourth-grade mathematics score increased between 1996 and 2003 both for students who were eligible and students who were not eligible for free/reduced-price lunch in 44 jurisdictions and for students who were not eligible in North Dakota. The average eighth-grade mathematics scores increased between 1996 and 2003 for both students who were eligible and students who were not eligible in 22 jurisdictions, for eligible students in Montana, and for students who were not eligible in 10 jurisdictions.

## Urban District Results

The 2003 Trial Urban District Assessment (TUDA) included nine urban public-school districts (Atlanta City School District, Boston Public School District, Charlotte-Mecklenburg Schools, City of Chicago School District 299, Cleveland Municipal School District, Houston Independent School District, Los Angeles Unified School District, New York City Public Schools, and San Diego City Unified School District) plus the District of Columbia. Results for the urban districts are compared with results for public schools in the nation and public schools in large central cities.

### Overall Mathematics Results for the Urban Districts

### *At grade 4*

- Fourth-graders in all the participating districts except Charlotte scored lower on average than fourth-graders in the nation. Fourth-graders in Charlotte had a higher average score than public school students in the nation, large central cities, and the other participating districts.
- With the exception of Charlotte, fourth-grade scores at the 25th, 50th, 75th, and 90th percentiles were lower in each of the districts than in the nation. Scores at the 10th, 25th, 50th, 75th, and 90th percentiles were higher in Charlotte than in the nation and in large central cities.
- The percentage of fourth-graders in Charlotte performing at or above *Proficient* was higher than the corresponding percentages in both large central cities and the nation.

### *At grade 8*

- Eighth-graders in all the participating districts except Charlotte scored lower on average than eighth-graders in the nation. Eighth-graders in Charlotte had a higher average score than public school students in the nation, large central cities, and the other participating districts.
- Scores at the 25th, 50th, 75th, and 90th percentiles in all the districts except Charlotte were lower than in the nation. In Charlotte, eighth-grade scores at the 10th, 25th, 50th, 75th, and 90th percentiles were higher than the scores in large central cities, and the scores at the 75th and 90th percentiles were higher than the corresponding national scores.
- The percentage of eighth-graders in Charlotte at or above *Proficient* was higher than the corresponding percentages in both large central cities and in the nation.

## Results for Student Subgroups in Urban Districts

### *Gender*

- At grade 4, the average scores for both male and female students in Charlotte were higher than those for their counterparts in the nation and in large central cities. Male and female fourth-graders in Atlanta, Boston, Chicago, Cleveland, the District of Columbia, and Los Angeles had lower average scores than their counterparts in large central cities and in the nation.
- At grade 8, the average scores for both male and female students in Charlotte were higher than the corresponding average scores for male and female students in large central cities. Both male and female eighth-graders in Atlanta, Chicago, Cleveland, the District of Columbia, and Los Angeles had lower average scores than their counterparts in large central cities and in the nation.

### *Race/Ethnicity*

- At grade 4, the average scores for White students in Charlotte, the District of Columbia, and Houston; Black students in Boston, Charlotte, Houston, and New York City; and Hispanic students in Charlotte and Houston were higher than the corresponding scores in large central cities. The average scores for fourth-grade White students in Boston, Chicago, and Cleveland; Black students in Chicago and the District of Columbia; and Hispanic students in Boston, the District of Columbia, Los Angeles, and San Diego were lower than the corresponding scores in large central cities.
- At grade 8, the average scores for White students in Atlanta, Charlotte, and Houston; Black students in Charlotte, Houston, and New York City; and Hispanic students in Houston were higher than the corresponding scores in large central cities. The average scores for eighth-grade White students in Cleveland; Black students in Atlanta, the District of Columbia, and Los Angeles; and Hispanic students in the District of Columbia, Los Angeles, and San Diego were lower than the corresponding scores in large central cities.

### *Eligibility for Free/Reduced-Price Lunch*

- At grade 4, the average scores for students eligible for free/reduced-price lunch in Charlotte, Houston, and New York City were higher than the average score in large central cities. The average scores for eligible students in Atlanta, Chicago, the District of Columbia, and Los Angeles were lower than the average score for eligible students in large central cities.
- At grade 8, the average scores for students eligible for free/reduced-price lunch in Boston, Houston, and New York City were higher than the average score in large central cities. The average scores for eligible students in Atlanta, the District of Columbia, and Los Angeles were lower than the average score in large central cities.

### *Parents' Level of Education*

- In 2003, the average score for eighth-graders who indicated that at least one parent graduated from college was lower in Atlanta, Chicago, Cleveland, the District of Columbia, and Los Angeles than the average score for students in the same parental education category in public schools in large central cities and in the nation. The average score for eighth-graders who reported at least one parent graduated from college was higher in Charlotte and San Diego than for students in large central cities.

# 1 Introduction

Mathematics provides the basic processes for quantifying information. Using quantities is essential everywhere in our society, in every aspect of our daily lives—at home and in school, for commerce, travel, communications, entertainment, and medicine. Even if mathematics were not important as a key to understanding the structure of our world and universe, it would still be one of the key competencies for personal, civic, and economic engagement. Students need to understand and be able to apply mathematical skills and concepts in order to function effectively in daily activities such as understanding financial information and evaluating product pricing.

Great importance has long been placed on ensuring that students acquire mathematical skills and concepts and these skills have increasingly come to be expected of all students. This report presents major results from the National Assessment of Educational Progress (NAEP) 2003 mathematics assessment of the nation's fourth- and eighth-grade students. In addition, the report provides results for fourth- and eighth-grade students in 53 states and other jurisdictions and for the nine urban school districts that participated in the Trial Urban District Assessment. This report is intended to inform educators, policymakers, parents, and the general public about students' achievement in mathematics.

## Overview of the 2003 National Assessment of Educational Progress in Mathematics

For more than 30 years, NAEP has regularly collected, analyzed, and reported valid and reliable information about what students know and can do in a variety of subject areas. As authorized by the U.S. Congress, NAEP assesses representative national samples of fourth-, eighth-, and twelfth-grade students. Since 1990, NAEP has also assessed representative samples of fourth- and eighth-grade students in states and other jurisdictions that participate in the NAEP state-by-state assessments. NAEP is administered and overseen by the National Center for Education Statistics (NCES), within the U.S. Department of Education's Institute of Education Sciences.

The content of all NAEP assessments is determined by subject-area frameworks that are developed by the National Assessment Governing Board (NAGB) in a comprehensive process involving a broad spectrum of interested parties, including teachers, curriculum specialists, subject-matter specialists, school administrators, parents, and members of the general public. The framework for the NAEP 2003 mathematics assessment, which was updated in 1996, is essentially the same framework that has guided development of the NAEP mathematics assessments since 1990.

This report describes the results of the NAEP 2003 mathematics assessment at grades 4 and 8. National results for 2003 are compared to those from 1990, 1992, 1996, and 2000. Using the same test as that used nationally, state-level assessments were conducted at grade 4 in 1992, 1996, 2000, and 2003. At grade 8, state-level assessments were conducted in 1990, 1992, 1996, 2000, and 2003. Results for the nine districts that participated in the Trial Urban District Assessment (TUDA) are reported for 2003 only. Comparisons across assessment years are possible because the assessments were developed under the same basic framework and share a common set of mathematics questions.

Prior to 1996, administration procedures for the NAEP mathematics assessments did not permit the use of accommodations (e.g., extra time; individual rather than group administration) for students with special needs who could not participate without them. For the 1996 national assessment, however, administrative procedures were introduced that allowed expanding participation in NAEP through the use of accommodations by students with disabilities (SD) and limited-English-proficient (LEP) students (see appendix A). A split-sample design was used at the national level in 1996 and 2000 and at the state level in 2000, so that both administration procedures could be used during the same assessment, but with different samples of students. This made it possible to report trends in students' mathematics achievement across all the assessment years and, at the same time, examine the effects of including students assessed with accommodations on overall assessment results. Based on an examination of how permitting accommodations affected overall population results, it was decided that, beginning with the 2003 assessment, NAEP would use only one set of procedures—permitting the use of accommodations.

During the period in which accommodations were not permitted, students with special needs could only be included in the assessment if it was determined by school staff that they could be assessed meaningfully without accommodations. The change in administration procedures makes it possible for more students to be included in the assessments; however, it also represents an important altering of procedures from previous assessments. (See the section on Students with Disabilities and/or Limited-English-Proficient Students in appendix A for a more detailed discussion.) The reader is encouraged to consider the difference in accommodation procedures when interpreting comparisons between the two sets of results.

The charts and tables throughout this report distinguish between results from assessment years in which accommodations were not permitted and results from assessment years in which accommodations were permitted. In the tables and charts that display results across assessment years, all previous assessment results that were found to be significantly different (at the .05 level based on t-tests adjusted using the False Discovery Rate (FDR) multiple comparison procedures) from 2003 results are marked with an asterisk (*). Two sets of results are presented for assessment years in which both administration procedures were used (accommodations not permitted and accommodations permitted). Both sets of results may also be notated, if found to be significantly different from 2003. The text that accompanies these tables and charts indicates which previous assessment results were significantly different from 2003. Comparisons between the 2003 results, when accommodations were permitted, and the 1990 and 1992 results, when they were not permitted, are discussed in the text. However, for previous assessment years with both accommodations-not-permitted results and accommodations-permitted results, the text describes comparisons only between the accommodations-permitted results and 2003.

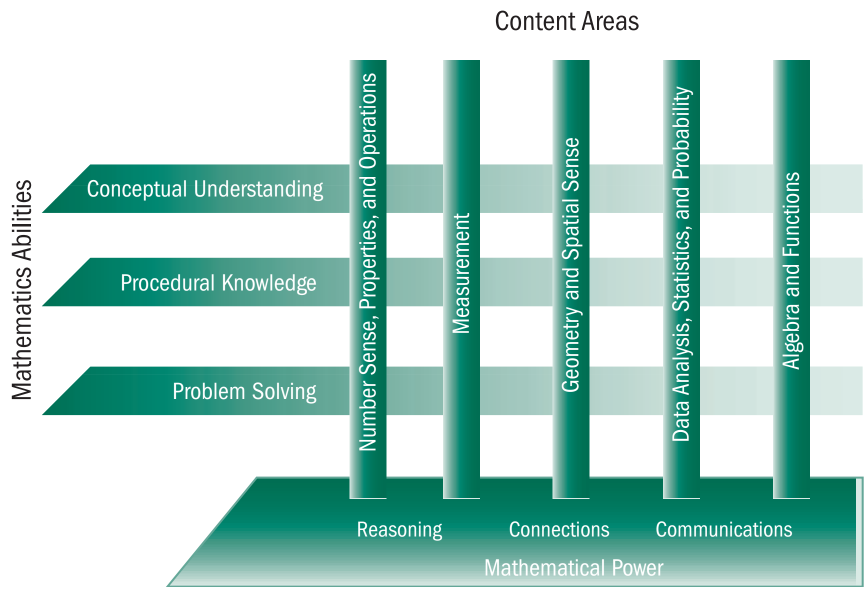
### Framework for the 2003 Mathematics Assessment Instrument

The NAEP Mathematics Framework is the blueprint that has specified the content and guided the development of each NAEP mathematics assessment since 1990. The framework resulted from a national process involving many organizations and individuals concerned with mathematics education. This cooperative effort was directed by the National Assessment Governing Board (NAGB) and managed by the Council of Chief State School Officers (CCSSO). In 1996, the framework was refined so that the 1996, 2000, and 2003 assessments could better reflect recent curricular emphases in mathematics, while maintaining the connection to the 1990 and 1992 assessments in order to measure trends in student performance.[1]

The framework calls for questions based on five mathematics content areas: 1) number sense, properties, and operations; 2) measurement; 3) geometry and spatial sense; 4) data analysis, statistics, and probability; and 5) algebra and functions. Questions were categorized according to two additional domains: mathematical abilities and mathematical power. The first domain, mathematical abilities, describes three types of knowledge or processes required for a student to successfully respond to a question: conceptual understanding; procedural knowledge; and problem solving, the ability to synthesize several processes when confronting a mathematical situation. The second domain, mathematical power, reflects the three processes stressed as major goals of the mathematics curriculum: the ability to reason, to communicate, and to make connections between concepts and skills either across the mathematics content areas, or from mathematics to other curricular areas. Figure 1.1 summarizes the structure of the 2003 assessment.

[1] National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

**Figure 1.1** **Structure of the NAEP 2003 mathematics assessment**




SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

A breakdown of the percentage of questions in each content area prescribed by the framework for the 1990, 1992, 1996, 2000, and 2003 assessments is provided in appendix A (see table A.1). The framework also incorporates the use of calculators (four-function at grade 4 and scientific at grade 8), rulers (at grades 4 and 8), protractors (at grade 8), and manipulatives such as spinners and geometric shapes (at grades 4 and 8). The use of these ancillary materials and the use of calculators were incorporated into some parts of the assessment, but not all. Calculator use was permitted on approximately one-third of the test questions.

## The NAEP 2003 Mathematics Assessment Instrument

The NAEP mathematics assessment is the only federally authorized, ongoing, nationwide assessment of student mathematics achievement. As such, it is necessary for the assessment to reflect the framework and expert perspectives on the measurement of mathematics performance. During the development process, the assessment undergoes stringent review by teachers and other educators, as well as by state officials and measurement specialists. All components of the assessment are evaluated for curricular relevance, developmental appropriateness, and fairness concerns.

The assessment comprised 50 booklets at each grade. Each booklet contained two separately timed 25-minute sections of mathematics questions. The total numbers of test questions used in the 2003 mathematics assessment at grades 4 and 8 were 181 and 197, respectively. Typically, a section, or block, contained approximately 16–20 questions, but there was considerable variation depending on the balance between multiple-choice and constructed-response questions.

The mathematics blocks include both multiple-choice and constructed-response questions designed to assess the framework objectives. Approximately 50 percent of student assessment time is devoted to constructed-response questions. Two types of constructed-response questions are used: 1) short constructed-response questions that require students to provide answers to computation problems or to describe solutions in one or two sentences, and 2) extended constructed-response questions that require students to give more detailed responses or explanations. Additional information about the design of the 2003 mathematics assessment is presented in appendix A.

In order to ensure reliable and valid scoring of constructed-response questions, a unique scoring guide describing the specific criteria for assigning a score level to each student's response is developed for each question. Expert scorers go through extensive training to understand how to apply these scoring criteria fairly and consistently. During the scoring process, scorers are consistently monitored to ensure that scoring standards are being applied appropriately and to ensure a high degree of scorer agreement (i.e., interrater reliability). In addition, for those constructed-response questions that were used in previous assessments, monitoring of scorers includes checking to make sure that scoring standards remain consistent from year to year.

In order to minimize the burden on any individual student, NAEP uses a procedure referred to as matrix sampling in which an individual student is administered only a small portion of the entire assessment at any grade. For example, at grades 4 and 8, each student is given only one of the 50 different grade-specific test booklets, each containing only two 25-minute blocks. Because each block is administered to a representative sample at each grade, the results can then be combined to produce average group and subgroup results based on the entire assessment. In addition to completing the two 25-minute blocks in each student's test booklet, students are asked to complete two sections of background questions that ask about their home or school experiences related to mathematics achievement. The time required for each student to participate in the NAEP mathematics assessment is approximately one hour.

### Description of School and Student Samples

The NAEP 2003 mathematics assessment was administered to fourth- and eighth-graders at the national and state levels. At the national level, results are reported for both public and nonpublic school samples. At the state or jurisdiction level, results are reported only for public school students. All 50 states and jurisdictions that participated in the 2003 assessment met the minimum guidelines for reporting their results.

In order to obtain a representative sample of students for reporting national and state or jurisdiction results, approximately 190,000 fourth-graders from 7,500 schools and 153,000 eighth-graders from 6,100 schools were sampled and assessed in 2003. Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. The national samples for mathematics were larger in 2003 than in previous assessment years because they were based on the combined sample of students assessed in each participating state, plus an additional sample from nonpublic schools. In the 1990–2000 assessments, the national samples were drawn separately from the state samples and were smaller than the samples resulting from aggregating the state samples.

For information on sample sizes and participation rates for the nation and by state or jurisdiction, see tables A.6–A.9 in appendix A.

Results from the 2003 Trial Urban District Assessment (TUDA) are reported for the participating districts for public-school students at grades 4 and 8. The TUDA employed a larger-than-usual sampling rate within the districts, making reliable district-level data possible. The samples were also large enough to provide reliable estimates on subgroups within the districts, such as female students or Hispanic students.

### Reporting the Assessment Results

Results from the NAEP mathematics assessment are presented in two ways: as scale scores and as percentages of students attaining the various achievement levels. The scale scores, indicating how much students *know and can do* in mathematics, are presented as average scale scores and as scale scores at selected percentiles. The achievement-level results indicate the degree to which student performance meets the standards set for what they *should know and be able to do*. Results are reported only for groups or subgroups of students; individual student performance cannot be reported based on the NAEP assessment.

Average scale score results are based on the NAEP mathematics scale, which ranges from 0 to 500. To calculate students' average scores on the NAEP mathematics assessment, the first step is to determine the percentage of students responding correctly to each multiple-choice question and the percentage of students responding at each score level for both the short and extended constructed-response questions. The determination of average scale scores entails summarizing the results on separate subscales for each of the five content areas in mathematics and then combining the separate scales to form a single composite scale. (See appendix A for more information on scaling procedures.) Results by separate subscales are accessible through the NAEP Data Tool on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

Achievement-level results are presented in terms of mathematics achievement levels as authorized by the NAEP legislation and adopted by NAGB. For each grade assessed, NAGB has adopted three achievement levels: *Basic, Proficient*, and *Advanced*. For reporting purposes, achievement-level cut scores are placed on the mathematics scale, resulting in four ranges: below *Basic, Basic, Proficient*, and *Advanced*. The achievement-level results are then reported as percentages of students scoring within each range, as well as the percentage of students at or above *Basic* and at or above *Proficient*.

### The Setting of Achievement Levels

The 1988 NAEP legislation that created the National Assessment Governing Board directed that the Board establish achievement-level goals for all the subjects assessed by NAEP.[2] The NAEP 2001 reauthorization reaffirmed many of the Board's statutory responsibilities, including "developing appropriate student achievement levels for each grade or age in each subject area to be tested. . . ."[3] In order to follow this directive and to achieve the mandate of the original NAEP legislation, NAGB undertook the development of student performance standards (called "achievement levels"). Since 1990,

[2] National Assessment of Educational Progress Improvement Act, P. L. 100–297, 20 U.S.C. § 1221 *et seq.* (1988).

[3] No Child Left Behind Act of 2001, P. L. 107-110, 115 Stat. 1425 (2002).

the Board has adopted achievement levels in mathematics, reading, U.S. history, world geography, science, writing, and civics.

The Board defined three levels for each grade: *Basic, Proficient*, and *Advanced*. 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 over challenging subject matter. The *Advanced* level presumes mastery of both the *Basic* and *Proficient* levels and represents superior performance. Figure 1.2 presents the policy definitions of the achievement levels that apply across grades and subject areas. The policy definitions guided the development of the achievement levels established in all subject areas. Adopting three levels of achievement for each grade signals the importance of looking at more than one standard of performance. In the Board's view, the overall achievement goal for students is performance at the *Proficient* level or higher as measured by NAEP. The *Basic* level is not the desired goal, but represents partial mastery that is a step toward *Proficient*.

**Figure 1.2 Policy definitions of the three NAEP achievement levels**

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

SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

The achievement levels in this report were adopted by the Board based on a standard-setting process designed and conducted under a contract with ACT. To develop these levels, ACT convened a cross-section of educators and interested citizens from across the nation and asked them to judge what students should know and be able to do relative to a body of content reflected in the mathematics framework. This process of setting achievement levels was reviewed by an array of individuals including policymakers, representatives of professional organizations, teachers, parents, and other members of the general public. Prior to adopting these levels of student achievement, NAGB engaged a large number of people to comment on the recommended levels and to review the results.

The results of the achievement-level-setting process, after NAGB's approval, became a set of achievement-level descriptions and a set of achievement-level cut scores on the 0–500 NAEP mathematics scale. The cut scores are the scores that define the boundaries between below *Basic, Basic, Proficient,* and *Advanced* performance levels at each grade.

### Mathematics Achievement-Level Descriptions for Each Grade

Specific definitions of the *Basic, Proficient,* and *Advanced* mathematics achievement levels for grades 4 and 8 are presented in figures 1.3 and 1.4. As noted previously, the achievement levels are cumulative; therefore, students performing at the *Proficient* level also display the competencies associated with the *Basic* level, and students at the *Advanced* level also demonstrate the competencies associated with both the *Basic* and *Proficient* levels. For each achievement level listed in figures 1.3 and 1.4, the scale score that corresponds to the lowest cut score within that level on the NAEP mathematics scale is shown in parentheses. For example, in figure 1.3, the scale score of 249 corresponds to the lowest score in the range defining the grade 4 *Proficient* level of achievement in mathematics.

**Figure 1.3 Descriptions of NAEP mathematics achievement levels, grade 4**

| Grade 4 Achievement Levels | |
|---|---|
| *Basic* (214) | **Fourth-grade students performing at the *Basic* level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content strands.**<br>Fourth graders performing at the *Basic* level should be able to estimate and use basic facts to perform simple computations with whole numbers, show some understanding of fractions and decimals, and solve some simple real-world problems in all NAEP content strands. Students at this level should be able to use — though not always accurately — four-function calculators, rulers, and geometric shapes. Their written responses are often minimal and presented without supporting information. |
| *Proficient* (249) | **Fourth-grade students performing at the *Proficient* level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content strands.**<br>Fourth graders performing at the *Proficient* level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content strands; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the *Proficient* level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved. |
| *Advanced* (282) | **Fourth-grade students performing at the *Advanced* level should apply integrated procedural knowledge and conceptual understanding to complex and nonroutine real-world problem solving in the five NAEP content strands.**<br>Fourth graders performing at the *Advanced* level should be able to solve complex nonroutine real-world problems in all NAEP content strands. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. These students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely. |

NOTE: The scores in parentheses indicate the cut point on the scale at which the achievement-level range begins.
SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

**Figure 1.4** **Descriptions of NAEP mathematics achievement levels, grade 8**

| Grade 8 Achievement Levels | |
|---|---|
| ***Basic*** **(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.**<br>Eighth graders performing at the *Basic* level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content strands through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.<br>As they approach the *Proficient* level, students at the *Basic* level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth graders show limited skill in communicating mathematically. |
| ***Proficient*** **(299)** | **Eighth-grade students performing at the *Proficient* level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content strands.**<br>Eighth graders performing at the *Proficient* level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections among 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.<br>Quantity and spatial relationships in problem 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. |
| ***Advanced*** **(333)** | **Eighth-grade students performing at the *Advanced* level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content strands.**<br>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. |

NOTE: The scores in parentheses indicate the cut point on the scale at which the achievement-level range begins.

SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

### Trial Status of Achievement Levels

The law requires that the achievement levels are to be used on a trial basis until the Commissioner of Education Statistics determines "that such levels are reasonable, valid, and informative to the public."[4] Until that determination is made, the law requires the Commissioner and the Board to state clearly the trial status of the achievement levels in all NAEP reports. In 1993, the first of several congressionally mandated evaluations of the achievement-level-setting process concluded that the procedures used to set the achievement levels were flawed and that the percentage of students at or above any particular achievement-level cut point may be underestimated.[5] Others have critiqued these evaluations, asserting that the weight of the empirical evidence does not support such conclusions.[6]

In response to the evaluations and critiques, NAGB sponsored an additional study of the 1992 reading achievement levels before deciding to use them for reporting NAEP 1994 results.[7] When reviewing the findings of this study, the National Academy of Education (NAE) panel expressed concern about what it saw as a "confirmatory bias" in the study and about the inability of this study to "address the panel's perception that the levels had been set too high."[8] In 1997, the NAE panel summarized its concerns with interpreting NAEP results based on the achievement levels as follows:

> First, the potential instability of the levels may interfere with the accurate portrayal of trends. Second, the perception that few American students are attaining the higher standards we have set for them may deflect attention to the wrong aspects of education reform. The public has indicated its interest in benchmarking against international standards, yet it is noteworthy that when American students performed very well on a 1991 international reading assessment, these results were discounted because they were contradicted by poor performance against the possibly flawed NAEP reading achievement levels in the following year.[9]

NCES and NAGB have sought and continue to seek new and better ways to set performance standards for NAEP.[10]

4 No Child Left Behind Act of 2001, P. L. 107-110, 115 Stat. 1425 (2002).

5 United States General Accounting Office. (1993). *Education Achievement Standards: NAGB's Approach Yields Misleading Interpretations.* U.S. General Accounting Office Report to Congressional Requestors. Washington, DC: Author.

National Academy of Education. (1993). *Setting Performance Standards for Achievement: A Report of the National Academy of Education Panel on the Evaluations of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement Levels.* Stanford, CA: Author.

6 Cizek, G. (1993). *Reactions to National Academy of Education Report.* Washington, DC: National Assessment Governing Board.

Kane, M. (1993). *Comments on the NAE Evaluation of the NAGB Achievement Levels.* Washington, DC: National Assessment Governing Board.

7 American College Testing. (1995). *NAEP Reading Revisited: An Evaluation of the 1992 Achievement Level Descriptions.* Washington, DC: National Assessment Governing Board.

8 National Academy of Education. (1996). Reading Achievement Levels. In *Quality and Utility: The 1994 Trial State Assessment in Reading. The Fourth Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment.* Stanford, CA: Author.

9 National Academy of Education. (1997). *Assessment in Transition: Monitoring the Nation's Educational Progress,* p. 99. Mountain View, CA: Author.

10 Reckase, M. D. (2000). *The Evolution of the NAEP Achievement Levels Setting Process: A Summary of the Research and Development Efforts Conducted by ACT.* Iowa City, IA: ACT, Inc.

For example, NCES and NAGB jointly sponsored a national conference that explored many issues related to standard setting in large-scale assessments.[11] Although new directions were presented and discussed, a proven alternative to the current process has not yet been identified. NCES and NAGB continue to call on the research community to assist in finding ways to improve standard setting for reporting NAEP results.

The most recent congressionally mandated evaluation conducted by the National Academy of Sciences (NAS) relied on prior studies of achievement levels, rather than carrying out new evaluations, on the grounds that the process has not changed substantially since the initial problems were identified. Instead, the NAS panel studied the development of the 1996 science achievement levels. The NAS panel basically concurred with earlier congressionally mandated studies. The panel concluded that "NAEP's current achievement-level-setting procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; raters' judgments of different item types are internally inconsistent; appropriate validity evidence for the cut scores is lacking; and the process has produced unreasonable results."[12]

The NAS panel accepted the continuing use of achievement levels in reporting NAEP results on a trial basis, until such time as better procedures can be developed. Specifically, the NAS panel concluded that ". . . tracking changes in the percentages of students performing at or above those cut scores (or in fact, any selected cut scores) can be of use in describing changes in student performance over time."[13]

NAGB urges all who are concerned about student performance levels to recognize that the use of these achievement levels is a developing process and is subject to various interpretations. NAGB and NCES believe that the achievement levels are useful for reporting trends in the educational achievement of students.[14] In fact, achievement-level results have been used in reports by the President of the United States, the Secretary of Education, state governors, legislators, and members of Congress. Government leaders in the nation and in more than 40 states use these results in their annual reports. However, based on the congressionally mandated evaluations so far, NCES agrees with the NAS panel's recommendation that caution needs to be exercised in the use of the current achievement levels. NCES has concluded

11 National Assessment Governing Board and National Center for Education Statistics. (1995). *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments of the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES).* Washington, DC: U.S. Government Printing Office.

12 Pellegrino, J. W., Jones, L. R., and Mitchell, K. J. (Eds.). (1998). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress.* Committee on the Evaluation of National Assessments of Educational Progress, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press.

13 Ibid., 176.

14 Forsyth, R. A. (2000). A Description of the Standard-Setting Procedures Used by Three Standardized Test Publishers. In *Student Performance Standards on the National Assessment of Educational Progress: Affirmations and Improvements.* Washington, DC: National Assessment Governing Board.

Nellhaus, J. M. (2000). States with NAEP-Like Performance Standards. In *Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvement.* Washington, DC: National Assessment Governing Board.

that these achievement levels should continue to be used on a trial basis and be interpreted with caution.

### Interpreting NAEP Results

The average scores and percentages presented in this report are estimates based on samples of students rather than on entire populations. Moreover, the collection of questions used at each grade level is but a sample of the many questions that could have been asked to assess the skills and abilities described in the NAEP mathematics framework. As such, the results are subject to a measure of uncertainty, reflected in the standard error of the estimates—a range of a few points above or below the score—which accounts for potential score or percentage fluctuation due to sampling and measurement error. The estimated standard errors for the estimated scale scores and percentages in this report are accessible through the NAEP Data Tool on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/). Examples of these estimated standard errors are also provided in appendix A, tables A.23 to A.27, of this report.

The differences between scale scores and between percentages discussed in the following chapters take into account the standard errors associated with the estimates. Comparisons are based on statistical tests that consider both the magnitude of the difference between the group average scores or percentages and the standard errors of those statistics. Estimates based on subgroups with smaller sample sizes are likely to have relatively large standard errors. As a consequence, some seemingly large differences may not be statistically significant. That is, it cannot be determined whether these differences are due to the particular makeup of the samples of students who were selected, or to true differences in the population of interest. When this is the case, the term "apparent difference" or "no measurable difference" is used in this report. Differences between scores or between percentages are discussed in this report only when they are significant from a statistical perspective.

Beginning with the reading sample in 2002, the NAEP national samples were obtained by aggregating the samples from each state, rather than obtaining an independently selected national sample. Consquently, the national sample size increased and smaller differences between years or between subgroups of students were found to be statistically significant than would have been detected in previous assessment years. In keeping with past practice, all statistically significant differences are indicated in this report. All differences reported are significant at the .05 level with appropriate adjustments for multiple comparisons. The term "significant" is not intended to imply a judgment about the absolute magnitude or the educational relevance of the differences. It is intended to identify statistically dependable differences in average scores or percentages to help inform dialogue among policymakers, educators, and the public.

While the score ranges at each grade in mathematics are identical, the scale was derived independently at each grade. Therefore, average scale scores across grades cannot be compared. For example, equal scale scores on the grade 4 and grade 8 scales do not imply equal levels of mathematics achievement.

Comparisons of performance results may be affected by changes in exclusion rates for students with disabilities and limited-English-proficient students in NAEP samples. Percentages of students excluded from NAEP may vary considerably across states or districts, as well as across years. Comparisons of achievement results should be interpreted with caution if the exclusion rates vary widely. The percentages of students who were identified and assessed or excluded based on their disability or limited-English-proficient status are presented in appendix A.

The results presented are meant to describe some aspects of the condition of education. They are best viewed as suggesting various ideas to be further examined in light of other data, including state and local data, and in the context of the large research literature elaborating on the many factors contributing to educational achievement.

However, some readers are tempted to make unwarranted causal inferences from simple cross tabulations. At the risk of sounding dogmatic, it is almost never the case that a simple cross tabulation of any variable with a measure of educational achievement is conclusive proof that differences in that variable are a cause of differential educational achievement. The old adage that "correlation is not causation" is a wise precaution to be kept in mind when viewing the results presented here. Experienced researchers routinely formulate multiple hypotheses to take these possibilities into account and readers of this volume are encouraged to do likewise.

Additional NAEP data are available in the NAEP data tool and in restricted-access research databases. Researchers and policy analysts are free to make use of the data (subject to various confidentiality restrictions) as they wish. However, as part of the Institute for Education Sciences, NCES has a responsibility to try to discourage misleading inferences from the data presented and to educate the public on the difficulty of making valid causal inferences in a field as complex as education.

## Overview of the Remaining Report

This report describes the mathematics performance of fourth- and eighth-graders in the nation, participating states and other jurisdictions, large central city school districts, and selected urban school districts. Chapter 2 presents overall mathematics scale scores and achievement-level results across years for both the nation and participating states and other jurisdictions. Chapter 3 discusses national results for subgroups of students by gender, race/ethnicity, students' eligibility for free/reduced-price school lunch, parents' highest level of education (for grade 8 only), type of school (public and nonpublic), and school's type of location (central city, urban fringe/large town, rural/small town). State and jurisdiction results are reported by gender, race/ethnicity, and eligibility for free/reduced-price lunch. Overall and subgroup results for selected urban districts that were part of the TUDA are presented in chapter 4.

Chapter 5 presents sample assessment questions and student responses at each grade level, including samples of multiple-choice and constructed-response questions. A table showing the percentage of students at each achievement level who answered the questions successfully accompanies each sample question. In addition, item maps for each grade level describe the skill or ability needed to answer particular mathematics questions and show the score points at which individual students had a high probability of successfully answering particular questions, thereby indicating the relative difficulty of each question.

The appendices of this report contain information to expand the results presented in chapters 2–5. Appendix A contains an overview of assessment development, sampling, administration, and analysis procedures. Appendix B presents the percentages of students in each of the subgroups reported for the nation, states and other jurisdictions, and other selected urban districts. Appendix C includes tables with additional state-level and district-level subgroup results. Finally, appendix D shows state-level and district-level contextual data from sources other than NAEP.