NCES
National Center for Education Statistics

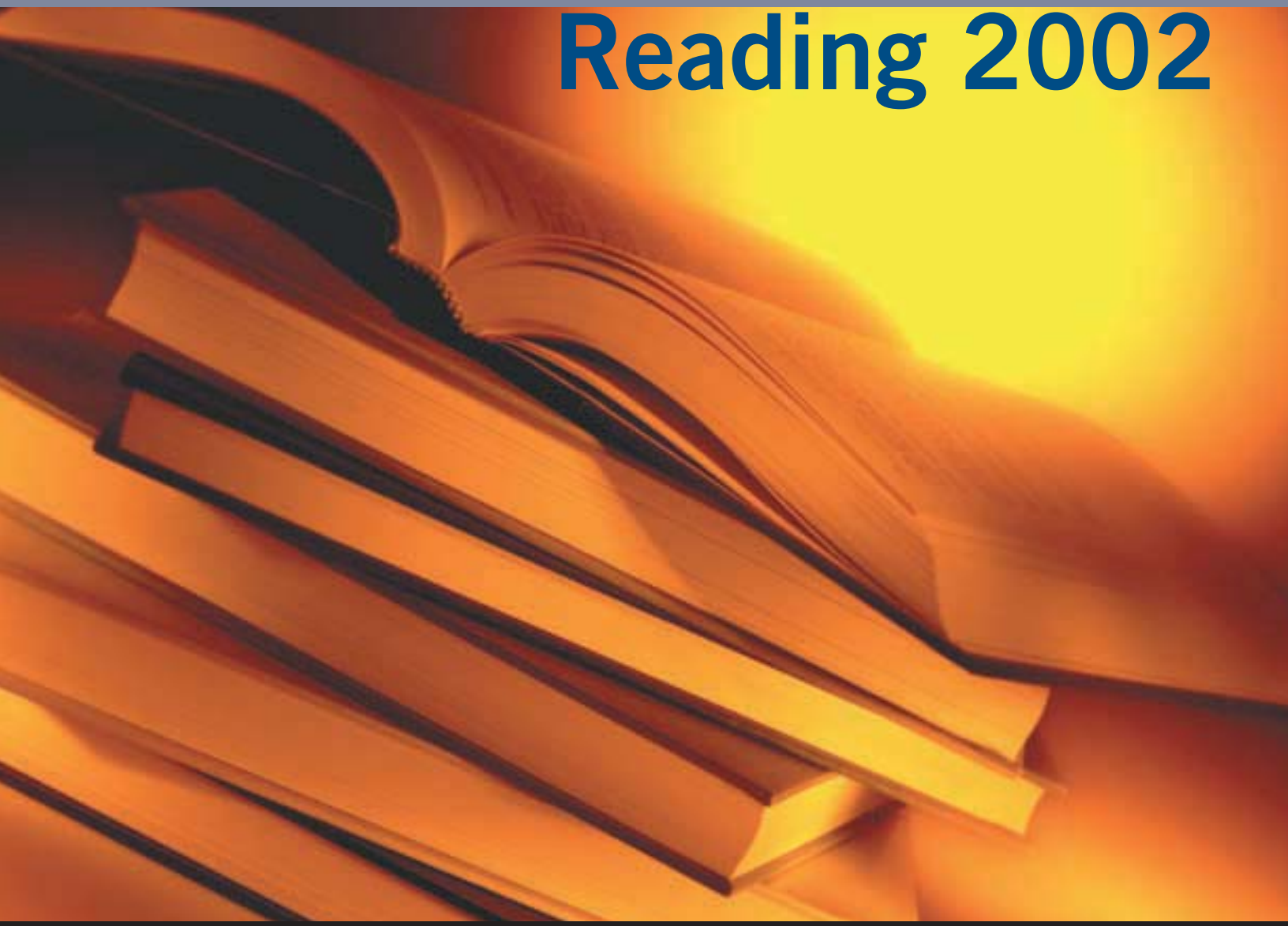
The Nation's Report Card NAEP

U.S. Department of Education
Institute of Education Sciences
NCES 2003-523

The Nation's Report Card

Trial Urban District Assessment

Reading 2002



The National Assessment of Educational Progress

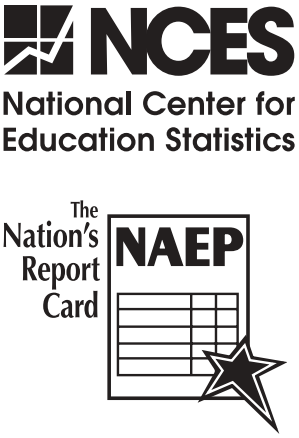
What is The Nation's Report Card?

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 National Assessment of Educational Progress reports.

The National Assessment Governing Board



U.S. Department of Education
Institute of Education Sciences
NCES 2003–523

The Nation's Report Card

Reading 2002

TRIAL URBAN DISTRICT ASSESSMENT

Anthony D. Lutkus
Arlene W. Weiner
Mary C. Daane
Ying Jin

in collaboration with
Laura Jerry
Gerry Kokolis
Educational Testing Service

Arnold A. Goldstein
Project Officer
National Center for Education Statistics

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Val Plisko
Associate Commissioner

July 2003

SUGGESTED CITATION
U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. *The Nation's Report Card: Reading 2002, Trial Urban District Assessment,* NCES 2003–523, by A. D. Lutkus, A. W. Weiner, M. C. Daane, and Y. Jin. Washington, DC: 2003.

FOR MORE INFORMATION
Content contact:
Arnold Goldstein
202–502–7344

To obtain single copies of this report, limited number of copies available, or 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 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.

Table of Contents

Chapter 1: Tables and Figures

Chapter 2: Tables and Figures

Chapter 3: Tables and Figures

Executive Summary

The National Assessment of Educational Progress (NAEP) is the nation's ongoing representative sample survey of student achievement in core subject areas. NAEP, known as the Nation's Report Card, is authorized by Congress and administered by the National Center for Education Statistics (NCES) of the Institute of Education Sciences in the U.S. Department of Education. NAEP regularly reports to the public on the educational progress of students in grades 4, 8, and 12.

In 2002, NAEP assessed the reading and writing performance of the nation's fourth-, eighth-, and twelfth-grade students. NAEP also conducted assessments of fourth- and eighth-graders' reading and writing in most of the states.

In 2001, after discussion among NCES, the National Assessment Governing Board (NAGB), and the leadership of the Council of the Great City Schools, Congress appropriated funds for a trial district-level assessment and NAGB passed a resolution approving the selection of five large urban districts for participation in the Trial Urban District Assessment, a special project within NAEP. Thus, this report presents, for the first time, district-level results of NAEP reading assessments in five urban public-school districts: Atlanta City, City of Chicago, Houston ISD, Los Angeles Unified, and New York City Public Schools. Throughout this report, the districts are referred to simply as Atlanta, Chicago, Houston, Los Angeles, and New York City. The five districts participated voluntarily in the NAEP 2002

Trial Urban District Assessment in reading at grades 4 and 8. Results are also included in this report for the District of Columbia, which, in 2002 and past years, has been included in the main assessments with states and other jurisdictions. Data for public schools across the nation and for central city public schools are provided for comparison purposes.[1] The public schools also included charter schools, which in some cases were not managed by the urban school district.

NAEP does not provide scores for individual students or schools. It reports results for groups of students (e.g., fourth-graders). For each group in each table in the report, assessment results are described in one of two ways: the group's average reading score on a scale from 0 to 500 or the percentage of students in the group who reached each of three achievement levels: *Basic*, *Proficient*, and *Advanced*. The *Proficient* level for each grade is defined by the Governing Board as representing "solid academic performance," which demonstrates "competency over challenging subject matter" for the grade assessed. *Basic* indicates partial mastery of skills that are fundamental for proficient work. *Advanced* denotes superior performance.

The achievement levels are performance standards adopted by 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. As provided by law, NCES, upon review of a congressionally mandated evaluation of NAEP, determined that the achievement levels are to be used on a trial basis and should be interpreted with caution. However, both NCES and NAGB believe that the 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.

The results are based on representative samples of students for the nation and for participating districts. In order to obtain reliable and representative data, a large proportion of the selected schools and students must participate. All six districts met the NCES statistical participation criteria for NAEP samples at grade 4, but New York City data will not be reported for grade 8 because eighth-grade participation did not meet the criteria.

Some students are identified by the school districts as students with disabilities and/or limited English proficient students. Some of these students are excluded from the assessment, and others are tested with accommodations related to their status. Because the percentages of students identified, excluded, and assessed with accommodations vary across the urban districts, that variability should be taken into consideration in interpreting the results and making comparisons (see appendix A, table A.1). For example, in the case of fourth-grade students, the percentages of students identified as having disabilities or limited English proficiency ranged from 8 to 51 percent, the percentages of fourth-grade students excluded for these reasons ranged from 2 to 17 percent, and the percentages assessed with accommodations ranged from 1 to 8 percent. At the eighth grade, the

[1] "Central city" includes a nationally representative sample of public schools located in central cities within metropolitan statistical areas as defined by the federal Office of Management and Budget. "Central City" is not synonymous with "inner city."

percentages of students identified with disabilities or limited English proficiency ranged from 6 to 35 percent, the percentage of eighth-grade students excluded for these reasons ranged from 2 to 7 percent, and the percentages assessed with accommodations ranged from 0 to 8 percent.

Throughout this report, differences between scores and between percentages are discussed in the text only when they are significant from a statistical perspective. All differences reported are significant at the 0.05 level with appropriate adjustments for multiple comparisons.

Overall Reading Results for the Urban Districts

Average Scores

Results for Grade 4

- The average scale scores for fourth-graders ranged from 191 in the District of Columbia and Los Angeles to 206 in Houston and New York City.
- The average score for public-school students in the nation as a whole was higher than the average score in each of the urban districts, and the national average score in central city public schools was higher than the average score in each of the urban districts except Houston and New York City.
- The average scale scores in Houston and New York City were higher than those of the other urban districts and were not found to differ significantly from each other.

Results for Grade 8

Results for New York City schools at grade 8 are not reported because they did not meet participation rates.

- The average scale scores for eighth-graders ranged from 236 in Atlanta to 249 in Chicago.
- The national average scores for public school students and for students in the central city public schools were higher than the average score in any of the urban districts.
- The average scale scores in Chicago and Houston were higher than those of the other urban districts, but were not found to differ significantly from each other.

Reading Achievement Levels

Results for Grade 4

- The percentage of fourth-graders performing at or above *Proficient* ranged from 10 percent in the District of Columbia to 19 percent in New York City. The percentage of students performing at or above *Proficient* in public schools in the nation was 30 percent, and for students in central city public schools it was 21 percent.
- Any apparent differences between the percentages of students performing at or above *Proficient* in the urban districts were not found to be statistically significant.
- The percentage of students performing at or above *Basic* ranged from 31 percent in the District of Columbia to 48 percent in Houston. In public schools across the nation, 62 percent of students performed at or above *Basic*. In central city schools in the national sample, 51 percent performed at or above *Basic*.
- The percentages of students performing at or above *Basic* were higher in Houston and New York City than in the other urban districts.

Results for Grade 8

- The percentages of eighth-graders performing at or above *Proficient* ranged from 8 percent in Atlanta to 17 percent in Houston. Thirty-one percent of students in public schools in the nation and 23 percent in central city public schools performed at or above *Proficient*.
- The percentages of students performing at or above *Proficient* in national public schools and central city public schools were higher than the percentages in each of the urban districts.
- The percentages of students performing at or above *Basic* ranged from 42 percent in Atlanta to 62 percent in Chicago. Seventy-four percent of public school students in the nation and 64 percent in central cities performed at or above *Basic*.
- The percentages of students performing at or above *Basic* in Chicago and Houston did not differ significantly from each other and both were higher than the comparable percentages in Atlanta, the District of Columbia, and Los Angeles.

Results for Student Subgroups

In addition to providing average scores and achievement levels for the nation, for states, and, in this report, for urban districts, NAEP reports provide results for subgroups of students defined by various background and contextual characteristics (e.g., gender, eligibility for free or reduced-price lunch, and level of parents' education).

Gender

Results for Grade 4

- Both male and female fourth-grade students in Atlanta, Chicago, the District of Columbia, and Los Angeles had average scores that were below the national average score for their counterparts in central city public schools.
- Female students in the urban districts outscored male students, on average. The exception was Houston, where the apparent difference was not statistically significant.

Results for Grade 8

- The average score for male eighth-graders in central city schools across the nation was higher than the average for male eighth-graders in each of the urban districts. The average score for female students in all urban districts except Chicago was below the national average for female students in central city public schools.
- In all urban districts, female students had higher average scores than male students.

Race/Ethnicity

Whereas White students constitute 60 percent of the national public sample at grade 4 and 64 percent at grade 8, in the urban districts, White students make up a maximum of 15 percent of the samples at grade 4 (New York City) and 11 percent of the samples at grade 8 (Chicago). Black or Hispanic students or both constitute majorities in the urban districts in the trial assessment. Hispanic students made up half or more of the sample in Houston and Los Angeles at both grades 4 and 8. Black fourth- and eighth-grade students made up more than 80 percent of the sample in both Atlanta and the District of Columbia.

Results for Grade 4

- In five of the six urban districts in which a reliable comparison could be made, White fourth-graders had higher average scores than their Black and Hispanic peers. In Chicago, Hispanic students had higher average scores than Black students.
- The average scores for Black students in Chicago, the District of Columbia, and Los Angeles were lower than the national average for Black students in central city public schools.
- The average scores for Hispanic students in Chicago, Houston, and Los Angeles were lower than the national average for Hispanic students in central city public schools.
- The average score for Asian/Pacific Islander students in New York City was higher than the national average for Asian/Pacific Islander students in central city public schools.

Results for Grade 8

- White eighth-graders had higher average scores than Black eighth-graders in Atlanta, Houston, and Los Angeles. The apparent difference in Chicago was not found to be statistically significant, and the sample size in the District of Columbia was insufficient to permit a reliable comparison.
- The average scores for Black students in Atlanta, the District of Columbia, and Los Angeles were lower than the national average for Black students in central city public schools.
- The average score for Hispanic students in Los Angeles was lower than the national average for Hispanic students in central city public schools.
- The average score for White students in Houston was higher than the national average for White students in central city public schools.

Eligibility for Free/Reduced-Price Lunch

The federal program providing free/reduced-price school lunch is administered by the U.S. Department of Agriculture (USDA) for children near or below the poverty line. Eligibility is determined by the USDA's Income Eligibility Guidelines (*http://www.fns.usda.gov/cnd/IEGs&NAPs/IEGs.htm*). At grade 4, the percentages of students in the urban districts eligible for free/reduced-price lunch ranged from 72 percent in Houston to 88 percent in Chicago. By comparison, 43 percent of fourth-graders in public schools nationally were eligible. At grade 8, the percentages of students eligible for free/reduced-price lunch in four of the urban districts ranged from 68 percent to 84 percent. By comparison, 34 percent of eighth-graders in public schools nationally were eligible. (Information on the free/reduced-price lunch data for eighth-graders in Los Angeles is not reported because these data did not meet reporting standards.)

Results for Grade 4

- In each of the urban districts, fourth-grade students not eligible for free/reduced-price lunch had higher average scores than students who were eligible.
- The average scores for eligible students in Atlanta, Chicago, the District of Columbia, and Los Angeles were lower than the national average for eligible students in central city public schools.

- The average scores for students in Atlanta, the District of Columbia, and Los Angeles who were not eligible for the program were lower than the national average for students in central city public schools who were not eligible.

Results for Grade 8

- Eighth-grade students not eligible for free/reduced-price lunch had higher average scores than eligible students in each of the urban districts except Chicago, where the apparent difference was not statistically significant.
- The average scores for eligible students in Atlanta and the District of Columbia were lower than the national average for eligible students in central city public schools.
- The average scores for students in Atlanta, the District of Columbia, and Houston who were not eligible for the program were lower than the national average for students in central city public schools who were not eligible.

Parents' Highest Level of Education

Eighth-grade students who participated in the Trial Urban District Assessment were asked to indicate the highest level of education they thought that their parents had completed. Five response options were offered: did not finish high school, graduated from high school, some education after high school, graduated from college, and "I don't know."

- In comparison with the other urban districts, the District of Columbia had the highest percentage of eighth-graders (40 percent) who reported that at least one parent had graduated from college.
- In each of the urban districts, the percentage of students who reported that at least one parent had graduated from college was lower than that of public schools nationally.
- For students who reported that at least one parent graduated from college, the average scores for students in Atlanta, Chicago, the District of Columbia, and Los Angeles were lower than the national average for students in central city public schools.

1 Introduction

Overview of the Trial Urban District Assessment in Reading 2002

How are America's schools doing? The answer is of deep concern to our nation. So it is very important to have reliable information. With this report, some of the largest urban public-school districts in the United States will have a new source of reliable data about their students with comparisons to public schools nationally and central city schools.[1] This report presents results from the 2002 National Assessment of Educational Progress (NAEP) Trial Urban District Assessment of their fourth- and eighth-grade students' reading.

Brief History of the National Assessment of Educational Progress

For more than thirty years NAEP has been providing reliable information about American students' achievement. The purpose of NAEP reports is to inform educators, policy makers, parents, and the public. NAEP (also known as the Nation's Report Card) regularly and systematically collects, analyzes, and reports valid and reliable information about what American students know and can do in a variety of subject areas. NAEP assesses representative national samples of fourth-, eighth-, and twelfth-grade students, and representative samples of students in a number of states and other jurisdictions such as the

[1] Central city is defined in chapter 2 and more completely in the "Type of Location" section of Appendix C. Central city includes nationally representative public schools located in central cities within metropolitan statistical areas as defined by the federal Office of Management and Budget. It is not synonymous with "inner city."

District of Columbia and the Department of Defense schools (domestic and overseas).

NAEP is a continuing expression of the nation's concern with students' learning. Congress authorized NAEP in 1969. The National Center for Education Statistics (NCES), one of three centers within the U.S. Department of Education's Institute of Education Sciences (IES), is responsible for conducting the NAEP. An independent body, the National Assessment Governing Board (NAGB), provides policy oversight for NAEP.

NAEP does not report results for individual students or schools. In fact, for its first two decades, NAEP reported results only for the nation as a whole and for subgroups within the nation (e.g., for female students and male students). Then, in 1988, Congress authorized a trial of state-level assessment. So, in 1990, NAEP began collecting and providing data for states on mathematics performance on representative samples of students in participating states.

The trial of state-level assessment was clearly successful. By 1994, NAEP was reporting to many states on their students' performance in mathematics, reading, writing, and science. In 2002, NAEP assessed representative samples of students in most of the states and also in several U.S. jurisdictions. What is new about this report is that it presents the results of assessments in each of six participating urban districts.

The national- and state-level reports of the NAEP 2002 assessment are available on the NAEP web site: *http://nces.ed.gov/nationsreportcard*. They present nationwide results for grades 4, 8, and 12, and results for grades 4 and 8 within the states or jurisdictions that participated in the state-level assessment. The national and state reports also compare the results to the results of previous NAEP reading assessments.

Background of the NAEP Trial Urban District Assessment in 2002

There has been an ongoing interest in NAEP reports by school district and even by school building, but NAEP does not assess large enough numbers of students to provide reliable results at those levels. The District of Columbia is an exception. Since it is not within a state, NAEP has sampled and analyzed its school district as the equivalent of a state.

Interest in reporting results by district increased with the passage of the No Child Left Behind Act.[2] Some large urban school districts, such as Los Angeles and New York, have enough students to meet NAEP requirements for sample size in reporting. In 2001, after discussions among the National Center for Education Statistics, the National Assessment Governing Board, and the leadership of the Council of the Great City Schools, Congress appropriated funds for the Trial Urban District Assessment. NAGB passed a resolution approving the selection of five large urban districts for participation in the Trial Urban District Assessment.[3]

[2] No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

[3] National Assessment Governing Board. (2001, November 20). Minutes of the November 15–17, 2001 NAGB Meeting: Committee Reports and Board Actions. Presented at the November 2001 NAGB meeting, Washington, DC.

The trial design calls for larger-than-usual sample sizes within the districts, making reliable district-level data possible. The assessment will allow these districts to make valid comparisons among themselves for the first time. Although individual states' own assessments may report data by district and school, their results are based on different scores, scales, and test designs, so up until now, districts have not been able to compare themselves to districts in other states.

By undertaking the Trial Urban District Assessment in reading and writing in 2002, NAEP continues a tradition of carefully extending its service to education, while preserving the rigorous sampling, scoring, and reporting procedures that have characterized the national and state assessments. The samples were large enough to provide data on subgroups within the districts, such as female students or Hispanic students. The 2002 data can serve the districts as a benchmark for study of changes in the performance of all their students and of particular subgroups of students.

NAEP gathers contextual data about in- and out-of-school experiences and socioeconomic factors from background questionnaires given to students, teachers, and school administrators. These data for the reporting districts are available on the NAEP web site (*http://nces.ed.gov/nationsreportcard/naepdata*).

Selection of Trial Urban Districts

Representatives of the Council of the Great City Schools worked with the staff of NAGB to identify districts for the trial assessment. Districts were selected that permitted testing of the feasibility of conducting NAEP over a range of characteristics, such as district size, minority concentrations, federal program participation, socioeconomic conditions, and percentages of students with disabilities and limited English proficient students. This report presents data for the following participating urban districts: Atlanta City, Chicago School District 299, Houston Independent School District, Los Angeles Unified, New York City Public Schools, and Washington, DC. All these participating districts are located in central cities.

Overview of the NAEP 2002 Reading Assessment

What Was Assessed?

Each NAEP assessment has objectives described in a "framework"—a document that specifies the important content and process areas to be measured and the types of questions to be included.[4] NAGB directs a process for specifying these frameworks. Teachers, curriculum specialists, subject-matter specialists, local school administrators, parents, and members of the general public participated actively in the development of the reading framework.

The NAEP 2002 Trial Urban District Assessment used the same reading framework as the national and state assessments. This framework has guided the NAEP reading assessments since 1992.

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

The framework is founded on a body of educational research that defines reading as an interactive and constructive process involving the reader, the text, and the context of the reading experience. Reading involves the development of an understanding of text, thinking about the text in different ways, and using a variety of text types for different purposes.

The framework calls for assessing reading in three contexts: reading for literary experience, reading to gain information, and reading to perform a task. Each context for reading is associated with a range of different types of texts that are included in the NAEP reading assessment. All three contexts for reading are assessed at grade 8, but reading to perform a task is not assessed at grade 4.

As readers attempt to develop understanding of text, they focus on general topics or themes, interpret and integrate ideas, make connections to background knowledge and experiences, and examine the content and structure of the text. The framework accounts for these different approaches to understanding text by specifying four "aspects of reading" that represent the types of comprehension questions asked of students.

The NAGB framework for the NAEP reading assessment lists the four aspects of reading as follows:

- *Forming a general understanding:* To form a general understanding, the reader must consider the text as a whole and provide a global understanding of it.
- *Developing interpretation:* To develop an interpretation, the reader must extend initial impressions to develop a more complete understanding of what was read.
- *Making reader/text connections:* To make reader/text connections, the reader must connect information in the text with knowledge and experience.
- *Examining content and structure:* Examining text content and structure requires critically evaluating, comparing and contrasting, and understanding the effect of such features as irony, humor, and organization.

All four aspects of reading are assessed at all three grades within each context of reading. Further details on the reading framework are available at *http://www.nagb.org/pubs/pubs.html*.

The assessment contains reading materials that were drawn from sources commonly available to students both in and out of the school environment. These materials were considered to be representative of the types of reading experiences typically encountered by students. Each student in the Trial Urban District Assessment (as well as the national and state assessments) was asked to complete two 25-minute sets of questions, each consisting of a reading passage and comprehension questions associated with it. A combination of multiple-choice and constructed-response (short and extended written responses) questions was used to assess students' understanding of the passages. The passages and associated questions followed a distribution specified by the framework. Example NAEP reading passages and questions that have been released to the public, along with student performance data by state, are available on the NAEP web site (*http://nces.ed.gov/nationsreportcard/itmrls*).

Sampling: Who Was Assessed?

The NAEP 2002 reading assessment was administered to fourth-, eighth-, and twelfth-graders at the national level and to fourth- and eighth-graders at the state level. For the NAEP state assessments, a target for each jurisdiction is a sample of 100 schools and 3,000 students, except in small or sparsely populated jurisdictions. The sample of schools and students is chosen in a two-stage sampling process. First, the sample of schools is selected using probability sampling methods. Then, within the participating schools, random samples of students are chosen. Only public schools are reported in state and jurisdiction reports. In order to obtain a representative sample of students for reporting national and state or jurisdictional results, approximately 140,000 fourth-graders from 5,500 schools, 115,000 eighth-graders from 4,700 schools, and 15,000 twelfth-graders from 700 schools were sampled and assessed. The public schools also included charter schools, which in some cases were not managed by the urban school district. Details of sampling procedures are available in appendix C. The national and state results in 2002 draw on common samples. The national results include the results from the states and trial urban districts, weighted appropriately to represent the student population.

Sampling for the Trial Urban District Assessment was modeled on the procedure for sampling states. The number of participating schools ranged from 38 to 76 per district in the fourth grade and from 15 to 69 per district at the eighth grade. The number of participating students per district ranged from 947 to 2,100 at the fourth grade and from 1,110 to 1,778 at the eighth grade.

The overall participation rates for schools and students in the national, state, and trial urban district assessments must meet statistical guidelines established by NCES and NAGB in order for assessment results to be reported publicly. Data are not reported to the public for a state, jurisdiction, or urban district that participates but does not meet minimum participation rate guidelines. For more information about participation guidelines, see standards for sample participation and reporting of results in appendix C.

As with the national samples, the urban district samples were weighted to allow for valid inferences about the populations of interest. Participation rates for the jurisdictions and urban districts were calculated the same way as rates were computed for the nation.

NAEP endeavors to assess all students selected in the random sampling process, including students with disabilities (SD) and students who are classified by their schools as limited English proficient (LEP). The percentages of students classified as SD or LEP in all participating states and jurisdictions are available in an interactive database at the NAEP web site at *http://nces.ed.gov/nationsreportcard/naepdata*. Information on SD and LEP students for the trial urban districts appears in appendix A. Percentages of students identified as limited English proficient, particularly at grade 4, appear much higher in some districts (Houston and Los Angeles) than in the nation.

Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. Students who cannot be meaningfully assessed, or students for whom NAEP cannot supply the locally required testing accommodation, can be excused from the assessment. Such students are assumed not to be part of the population groups from which NAEP reports assessment results. It is important to note that, guided by the student's Individualized Education Program (IEP), as well as eligibility for "section 504" services, school personnel make decisions on whether students with disabilities should be included in the assessment.[5] They also decide, based on NAEP's guidelines, whether to include an LEP student. The guidelines ask them to judge the student's ability to participate in the assessment in English as well as to consider the number of years the student has been receiving instruction in English.

Percentages of students excluded from NAEP may vary considerably across states/districts and, within a state or district, from one year to another. If there are great differences between groups in the rates of exclusion, comparisons between those groups may be affected. If a higher or lower percentage of students in one district or one state or one year did not take the test for reasons of disability or limited English proficiency, or if a higher proportion required accommodations for disabilities, comparisons of student performance results with similar groups (states, districts, or years) should be interpreted with caution.

For example, the percentage of fourth-grade students identified as having disabilities and or limited English proficiency ranged from 19 to 30 percent in public schools in the nation, central cities and three of the six districts. Atlanta was below the national average, with only 8 percent of its fourth-grade students identified in these groups; and Houston and Los Angeles were above the national average, with 43 and 51 percent, respectively, of their fourth-grade students identified in these groups (in both of these cases the majority of the identified students had limited English proficiency). Given that many students identified in these groups participate in the assessment, the percentage of fourth-grade students who were excluded ranged between 7 and 9

[5] Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

in public schools across the nation, in central cities, and in four of the six public school districts included in this report. In contrast to this, in Atlanta only 2 percent of the fourth-grade students were excluded from the assessment because they had disabilities (1 percent) or were limited English proficient (1 percent). At the same time, 17 percent of the fourth-grade students in Houston were excluded from the assessment, with 16 percent excluded because they had limited English proficiency.

In the eighth grade, the percentage of students identified with disabilities and/or limited English proficiency ranged from 18 to 22 percent in public schools across the nation, central cities, and two of the five districts. Atlanta was below the national average, with only 6 percent of its eighth-grade students identified in these groups; and Houston and Los Angeles were above the national average, with 27 and 35 percent, respectively, of their eighth-grade students identified in these groups (in both of these cases the majority of the identified students had limited English proficiency). The percentages of eighth-grade students excluded from the assessment were more similar across jurisdictions. The exclusion rates for the nation, central city public schools, and four of the five districts ranged from 5 to 7 percent. In contrast, Atlanta excluded only 2 percent of its eighth-graders.

In both grades the percentage of students assessed with accommodations ranged from 0 or 1 percent to 8 percent across jurisdictions. The variability in the identification, exclusion, and accommodation rates should be taken into consideration in interpreting the results and making comparisons (see appendix A, table A.1).

How Is Student Performance Reported?

Results from the NAEP reading assessment are presented in two ways: as scale scores on a scale of 0–500 and as percentages of students performing at different achievement levels (*Basic*, *Proficient*, and *Advanced*). The results of student performance are reported for various groups of students (e.g., fourth-grade female students or students who took the assessment in different years).

Throughout this report, comparisons are made among districts and between districts and public schools in the nation, as well as between districts and central city public schools in the nation. (See "NAEP Reporting Groups" in appendix C for details on how central city public schools were defined.) The significance of differences in performance between groups of students that are reported here are based on statistical tests. The tests consider both the size of differences between averages or percentages and the standard error of those statistics. Every test score has a standard error—a range of a few points plus or minus the score—due to sampling error and measurement error. Statistical tests are used to determine whether the differences between average scores are significant. Only statistically significant differences are cited in this report.

The reader is cautioned to rely on the reported differences in the text and tables, which are statistically significant, rather than on the apparent magnitude of any difference. The standard errors are available on the NAEP web site (*http://www.nces.ed.gov/nationsreportcard/naepdata*).

Scale Scores

Student performance is reported as an average score based on the NAEP reading scale, which ranges from 0 to 500 and is linked to the corresponding scales in 1992, 1994, and 1998. The average scale score reflects the overall reading performance of a particular group of students. The overall composite scale was developed by weighting each of the separate reading subscales (two for fourth grade and three for eighth grade, reflecting each of the three above-mentioned reading contexts) based on its relative importance in the NAEP reading framework. This composite scale is the metric used to present the average scale scores and selected percentiles used in NAEP reports. More information on NAEP scales is available in appendix C.

Achievement Levels

Student reading performance is also reported in terms of three achievement levels: *Basic, Proficient*, and *Advance*d. Results based on achievement levels are expressed in terms of the percentage of students who attained each level. The three achievement levels are defined as follows:

- *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.

The achievement levels are performance standards adopted by NAGB as part of its statutory responsibilities. The levels represent collective judgments of what students should know and be able to do for each grade tested. They are based on recommendations of broadly representative panels of classroom teachers, education specialists, and members of the general public. As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that the achievement levels are to be used on a trial basis until it is determined that the achievement levels are "reasonable, valid, and informative to the public."[6] 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 as a common yardstick for academic performance. The reading achievement level descriptions for grades 4 and 8 are summarized in figures 1.1 and 1.2. The score ranges for the NAGB achievement levels on the NAEP scale are as follows: Grade 4, *Basic*, 208–237; *Proficient* 238–267; *Advanced*, 268 and above; Grade 8, *Basic*, 243–80; *Proficient*, 281–322; and *Advanced*, 323 and above.

[6] No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Cautions in Interpretations

The NAEP reading 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 reflect 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.

Figure 1.1 Descriptions of NAEP reading achievement levels, grade 4

Grade 4 Achievement Levels	
Basic (208)	Fourth-grade students performing at the *Basic* level should demonstrate an understanding of the overall meaning of what they read. When reading text appropriate for fourth graders, they should be able to make relatively obvious connections between the text and their own experiences, and extend the ideas in the text by making simple inferences. For example, when reading **literary** text, they should be able to tell what the story is generally about—providing details to support their understanding—and be able to connect aspects of the stories to their own experiences. When reading **informational** text, *Basic*-level fourth graders should be able to tell what the selection is generally about or identify the purpose for reading it, provide details to support their understanding, and connect ideas from the text to their background knowledge and experiences.
Proficient (238)	Fourth-grade students performing at the *Proficient* level should be able to demonstrate an overall understanding of the text, providing inferential as well as literal information. When reading text appropriate to fourth grade, they should be able to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences. The connections between the text and what the student infers should be clear. For example, when reading **literary** text, *Proficient*-level fourth graders should be able to summarize the story, draw conclusions about the characters or plot, and recognize relationships such as cause and effect. When reading **informational** text, *Proficient*-level students should be able to summarize the information and identify the author's intent or purpose. They should be able to draw reasonable conclusions from the text, recognize relationships such as cause and effect or similarities and differences, and identify the meaning of the selection's key concepts.
Advanced (268)	Fourth-grade students performing at the *Advanced* level should be able to generalize about topics in the reading selection and demonstrate an awareness of how authors compose and use literary devices. When reading text appropriate to fourth grade, they should be able to judge texts critically and, in general, give thorough answers that indicate careful thought. For example, when reading **literary** text, *Advanced*-level students should be able to make generalizations about the point of the story and extend its meaning by integrating personal experiences and other readings with ideas suggested by the text. They should be able to identify literary devices such as figurative language. When reading **informational** text, *Advanced*-level fourth graders should be able to explain the author's intent by using supporting material from the text. They should be able to make critical judgments of the form and content of the text and explain their judgments clearly.

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

Figure 1.2 Descriptions of NAEP reading achievement levels, grade 8

Grade 8 Achievement Levels	
Basic (243)	Eighth-grade students performing at the *Basic* level should demonstrate a literal understanding of what they read and be able to make some interpretations. When reading text appropriate to eighth grade, they should be able to identify specific aspects of the text that reflect the overall meaning, extend the ideas in the text by making simple inferences, recognize and relate interpretations and connections among ideas in the text to personal experience, and draw conclusions based on the text. For example, when reading **literary** text, *Basic*-level eighth graders should be able to identify themes and make inferences and logical predictions about aspects such as plot and characters. When reading **informational** text, they should be able to identify the main idea and the author's purpose. They should make inferences and draw conclusions supported by information in the text. They should recognize the relationships among the facts, ideas, events, and concepts of the text (e.g., cause and effect, order). When reading **practical** text, they should be able to identify the main purpose and make predictions about the relatively obvious outcomes of procedures in the text.
Proficient (281)	Eighth-grade students performing at the *Proficient* level should be able to show an overall understanding of the text, including inferential as well as literal information. When reading text appropriate to eighth grade, they should be able to extend the ideas in the text by making clear inferences from it, by drawing conclusions, and by making connections to their own experiences—including other reading experiences. *Proficient* eighth graders should be able to identify some of the devices authors use in composing text. For example, when reading **literary** text, students at the *Proficient* level should be able to give details and examples to support themes that they identify. They should be able to use implied as well as explicit information in articulating themes; to interpret the actions, behaviors, and motives of characters; and to identify the use of literary devices such as personification and foreshadowing. When reading **informational** text, they should be able to summarize the text using explicit and implied information and support conclusions with inferences based on the text. When reading **practical** text, *Proficient*-level students should be able to describe its purpose and support their views with examples and details. They should be able to judge the importance of certain steps and procedures.
Advanced (323)	Eighth-grade students performing at the *Advanced* level should be able to describe the more abstract themes and ideas of the overall text. When reading text appropriate to eighth grade, they should be able to analyze both meaning and form and support their analyses explicitly with examples from the text, and they should be able to extend text information by relating it to their experiences and to world events. At this level, student responses should be thorough, thoughtful, and extensive. For example, when reading **literary** text, *Advanced*-level eighth graders should be able to make complex, abstract summaries and theme statements. They should be able to describe the interactions of various literary elements (i.e., setting, plot, characters, and theme) and explain how the use of literary devices affects both the meaning of the text and their response to the author's style. They should be able to critically analyze and evaluate the composition of the text. When reading **informational** text, they should be able to analyze the author's purpose and point of view. They should be able to use cultural and historical background information to develop perspectives on the text and be able to apply text information to broad issues and world situations. When reading **practical** text, *Advanced*-level students should be able to synthesize information that will guide their performance, apply text information to new situations, and critique the usefulness of the form and content.

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

2 Average Scale Score and Achievement Level Results for the Trial Urban District Assessment

This chapter presents the NAEP 2002 reading results for the five urban districts that participated in the Trial Urban District Assessment and for the District of Columbia at grades 4 and 8. Average scores are reported on the NAEP reading scale, which ranges from 0 to 500, and in terms of percentages of students reaching the three reading achievement levels—*Basic*, *Proficient,* and *Advanced*. All comparisons in this report have been tested for statistical significance.

Urban District Scale Score and Percentile Results

Table 2.1 shows the overall performance of fourth- and eighth-grade students in the urban districts that participated in the 2002 reading assessment. In order to provide a context for these data, table 2.1 also displays the results for students attending public schools in the nation as a whole, as well as for public schools located in central cities across the nation. In table 2.1 and subsequent tables and figures in this report, a double asterisk (**) marks district statistics (average scores or percentages) that were found to be significantly different from the comparable statistic in public schools in the nation, and a single asterisk (*) marks district statistics that were significantly different from those of public schools in central cities. Following standards established by the Federal Office of Management and Budget, the U.S. Census Bureau defines a central city as a city of 50,000 people or more that is the largest in its metropolitan area, or can otherwise be regarded as "central," taking into account such characteristics as

commuting patterns (see "Type of Location" in appendix C for more detailed definitions of geographical areas, and see *http://www.census.gov/population/www/estimates/metroarea.html* for more information on Metropolitan Statistical Areas). The term means "a city that is central," not "the central part of a city" or "the inner city." Central cities are defined for this report as including large and midsize cities. Note that central cities encompass wider areas than what is commonly referred to as "the inner city." The five participating urban districts, Atlanta City, Chicago School District 299, Houston ISD, Los Angeles Unified, and New York City Public Schools, and Washington, DC are all located in central cities.

All students in these districts attended schools in central cities except in the Houston and Los Angeles districts, where some students included in the study attended schools located in the urban fringe. These included 6 percent and 19 percent of fourth-grade students in Houston and Los Angeles respectively, and 24 percent of eighth-grade students in Los Angeles.

The first column in table 2.1 presents the average score on the NAEP reading scale. At grade 4, the average scores ranged from 191 for the District of Columbia and Los Angeles to 206 for Houston and New York City. The average score for fourth-grade students in public schools in the nation was 217; for students in central city public schools nationally, the average was 208. At grade 8, the average scale scores ranged from 236 for students in Atlanta to 249 for students in Chicago. The average score for eighth-grade students in public schools in the nation was 263; for students in central city public schools, the average was 254. Data for New York City at grade 8 do not appear because the district did not meet the required 70 percent school participation rate (see appendix C, Standards for Sample Participation and Reporting of Results).

The remaining columns in table 2.1 show the scores at selected percentiles. Each percentile indicates the percentage of students in a given jurisdiction whose scores fell below a particular point on the NAEP reading scale. For example, for Atlanta at grade 4, the score at the 75th percentile was 219, indicating that 75 percent of fourth-grade students scored at or below 219. Looking at different percentiles makes it possible to examine the scale score gaps between higher and lower performing students within a district. In addition, comparing the scores for percentiles across districts gives a rough indication of how their score distributions may differ. In subsequent years, when additional results are available, comparing the score values for percentiles year to year will also be useful in pinpointing where changes may occur within a district's score distribution. The corresponding standard errors for these percentile scores are displayed in appendix C in table C.8.

Table 2.1 **Average reading scale scores and selected percentiles, grades 4 and 8 public schools: By urban district, 2002**

	Average scale score	Scale score distribution		
		25th percentile	50th percentile	75th percentile
Grade 4				
Nation (Public)	217	194	219	242
Central city (Public) [1]	208 **	183 **	209 **	234 **
Atlanta	195 *,**	171 *,**	194 *,**	219 *,**
Chicago	193 *,**	170 *,**	194 *,**	217 *,**
District of Columbia	191 *,**	167 *,**	191 *,**	215 *,**
Houston	206 **	183 **	206 **	229 **
Los Angeles	191 *,**	165 *,**	190 *,**	217 *,**
New York City ‡	206 **	182 **	206 **	230 **
Grade 8				
Nation (Public)	263	242	265	286
Central city (Public) [2]	254 **	232 **	256 **	278 **
Atlanta	236 *,**	214 *,**	236 *,**	259 *,**
Chicago	249 *,**	231 **	251 *,**	270 *,**
District of Columbia	240 *,**	219 *,**	241 *,**	262 *,**
Houston	248 *,**	226 **	251 *,**	273 *,**
Los Angeles	237 *,**	213 *,**	238 *,**	261 *,**

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.

* Significantly different from central city public schools.
** Significantly different from nation (public schools).

[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.

[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Comparisons Among Districts by Average Scale Scores

Figures 2.1 and 2.2 display the results of statistical significance tests of differences in the NAEP 2002 average reading scale scores of participating districts at grades 4 and 8. These figures are similar to mileage charts on travel maps: To read them, find the name of the district of interest ("target") in the rows on the left side of the figure. Then follow that row across the columns until you reach the district ("comparison district") whose average score you wish to compare to the target. If the cell comparing the target district (the row variable) to the comparison district is lightly shaded (with an upward arrow), the average scale score of the target district was higher than that of the "comparison district" named at the top

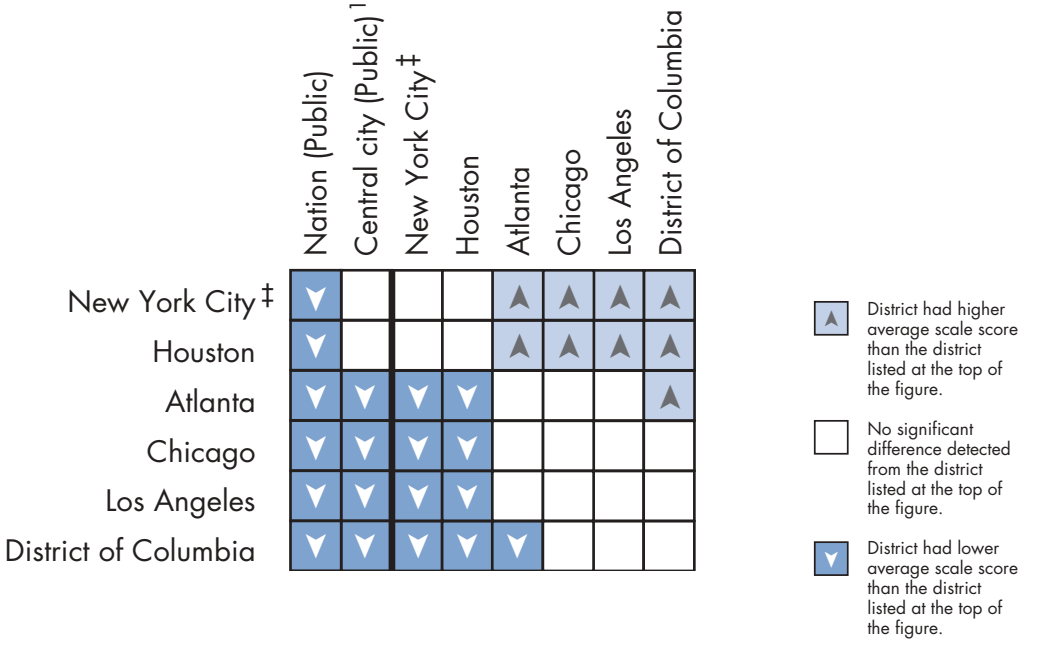
of the column. The darkly shaded cells (with a downward arrow) indicate that the average scale score of the target district was lower than that of the comparison district. Cells that are not shaded indicate that a statistically significant difference between the two districts was not detected. For example, in figure 2.1, the third cell in the third row compares the score at grade 4 in Atlanta to the average score in New York City. The shading in this cell indicates that the results of the statistical significance test showed the average score in Atlanta to be lower than that in New York City.

At grade 4, the average scale score for students in public schools in the nation was higher than the average score in each urban district. The average score for students in central city public schools across the nation was higher than the average score in Atlanta, Chicago, Los Angeles, and the District of Columbia. The average scores for students in Houston and New York City were higher than those in the other districts and were not found to differ significantly from each other. Finally, the average score in Atlanta was higher than the average score in the District of Columbia. Other apparent differences were not statistically significant.

Figure 2.1 Cross-district comparisons of average reading scale scores, grade 4 public schools: By urban district, 2002

Grade 4

Instructions: Read across the row corresponding to a district listed to the left of the chart. Match the shading intensity to the key below to determine whether the average reading scale score of this district was found to be higher than, not significantly different from, or lower than the district in the column heading. For example, in the row for Atlanta: Atlanta's score was lower than that of the nation, the central city sample, and New York City and Houston, not significantly different from Chicago and Los Angeles, and was higher than the District of Columbia.



‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

At grade 8, the average score for students in the nation as a whole and in central cities was higher than the average score in any of the districts. The average scores for students in Chicago and Houston were higher than those in the other districts and were not found to differ significantly from one another. Finally, grade 8 students in the District of Columbia outscored those in Los Angeles and Atlanta. Other apparent differences were not statistically significant.

Figure 2.2 Cross-district comparisons of average reading scale scores, grade 8 public schools: By urban district, 2002

Grade 8

Instructions: Read across the row corresponding to a district listed to the left of the chart. Match the shading intensity to the key below to determine whether the average reading scale score of this district was found to be higher than, not significantly different from, or lower than the district in the column heading. For example, in the row for the District of Columbia: The District of Columbia's score was lower than the nation, the central city sample, and Chicago and Houston, and higher than Los Angeles and Atlanta.

[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Reading Achievement Level Results

In addition to reporting average reading scale scores, NAEP reports reading performance in terms of percentages of students reaching achievement levels. The reading achievement levels are *Basic*, *Proficient*, and *Advanced*. The setting of achievement levels is discussed in chapter 1.

Table 2.2 presents the percentages of students at grades 4 and 8 who performed below the *Basic* level, at or above the *Basic* level, at or above the *Proficient* level, and at the *Advanced* level. At grade 4, the percentage of students performing at or above *Proficient* ranged from 10 percent in the District of Columbia to 19 percent in New York City. In contrast, 30 percent of the public-school students across the nation performed at or above the *Proficient* level, while 21 percent of all central city public-school students reached this level of achievement. The percentages of fourth-grade students reaching at least the *Basic* achievement level ranged from 31 percent in the District of Columbia to 48 percent in Houston. In the nation as a whole, 62 percent of public-school students and 51 percent of central-city students performed at or above the *Basic* level.

At grade 8, the percentages of students performing at or above the *Proficient* level ranged from a low of 8 percent in Atlanta to a high of 17 percent in Houston. In contrast, 31 percent of students in public schools in the nation and 23 percent of students in central-city schools performed at or above this level. The percentages of eighth-grade students performing at or above *Basic* ranged from 42 percent in Atlanta to 62 percent in Chicago. Across the nation, 74 percent of public-school students and 64 percent of public-school students in central cities performed at or above the *Basic* level.

Table 2.2 Percentage of students at or above each reading achievement level, grades 4 and 8 public schools: By urban district, 2002

	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Grade 4				
Nation (Public)	38	62	30	6
Central city public [1]	49 **	51 **	21 **	4 **
Atlanta	65 *,**	35 *,**	12 *,**	3 *,**
Chicago	66 *,**	34 *,**	11 *,**	2 *,**
District of Columbia	69 *,**	31 *,**	10 *,**	2 *,**
Houston	52 **	48 **	18 **	3 **
Los Angeles	67 *,**	33 *,**	11 *,**	2 *,**
New York City ‡	53 **	47 **	19 **	5
Grade 8				
Nation (Public)	26	74	31	2
Central city public [2]	36 **	64 **	23 **	2 **
Atlanta	58 *,**	42 *,**	8 *,**	# *,**
Chicago	38 **	62 **	15 *,**	1
District of Columbia	52 *,**	48 *,**	10 *,**	# *,**
Houston	41 *,**	59 *,**	17 *,**	1 *,**
Los Angeles	56 *,**	44 *,**	10 *,**	# *,**

Percentage rounds to zero.
‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

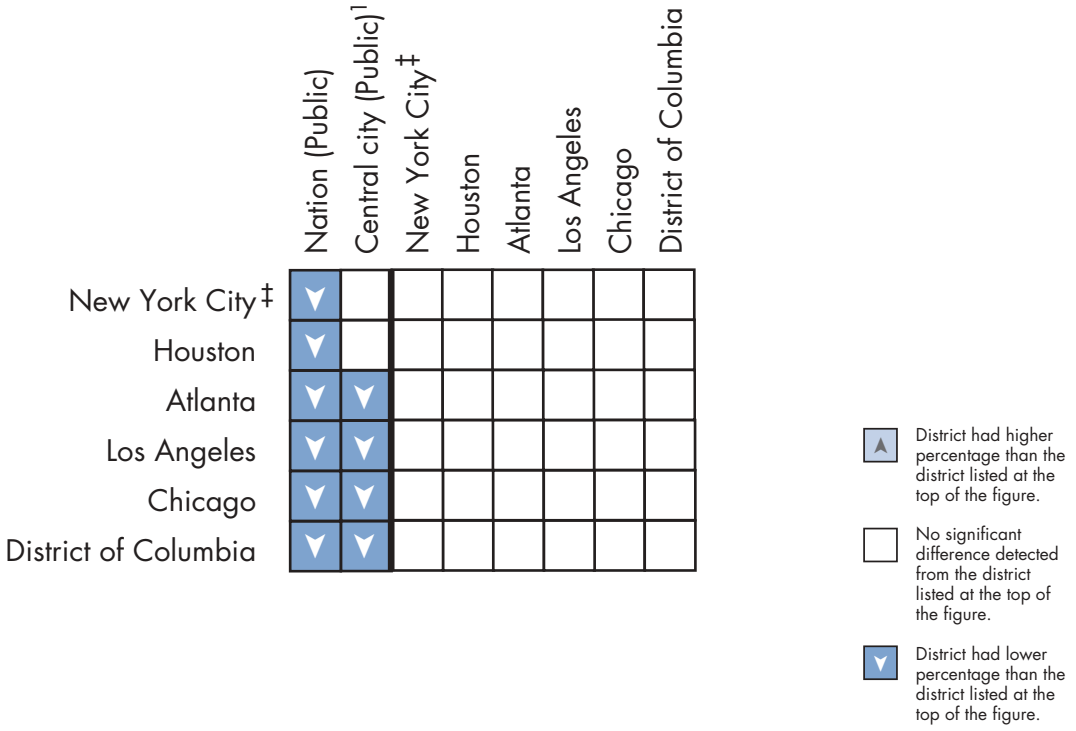
Comparisons Among Districts by Achievement Levels

Figures 2.3 and 2.4 are "mileage charts" similar to figures 2.1 and 2.2; in this case the data compared represent percentages of students performing at or above *Proficient*. At grade 4, the percentage of students performing at or above *Proficient* in the public schools nationally was higher than the percentages in each of the urban districts. The percentage of students performing at or above *Proficient* in central city public schools was not found to be different from the percentages in Houston and New York City, but was higher than the other four districts. The percentages of fourth-grade students performing at or above *Proficient* was not found to differ significantly across districts.

Figure 2.3 Cross-district comparisons of percentage of students at or above *Proficient* in reading, grade 4 public schools: By urban district, 2002

Grade 4

Instructions: Read across the row corresponding to a district listed to the left of the chart. Match the shading intensity to the key below to determine whether the percentage of students at or above *Proficient* in this district was found to be higher than, not significantly different from, or lower than the district in the column heading. For example, in the row for Chicago: The percentage of students at or above *Proficient* in Chicago was lower than the nation and the central city sample, and not significantly different from any of the other districts.



‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

At grade 8, the percentages of students performing at or above *Proficient* in the national public schools and in central city public schools were higher than the percentages in each of the participating urban districts. Houston had higher percentages of students performing at or above *Proficient* than Atlanta, the District of Columbia, and Los Angeles.

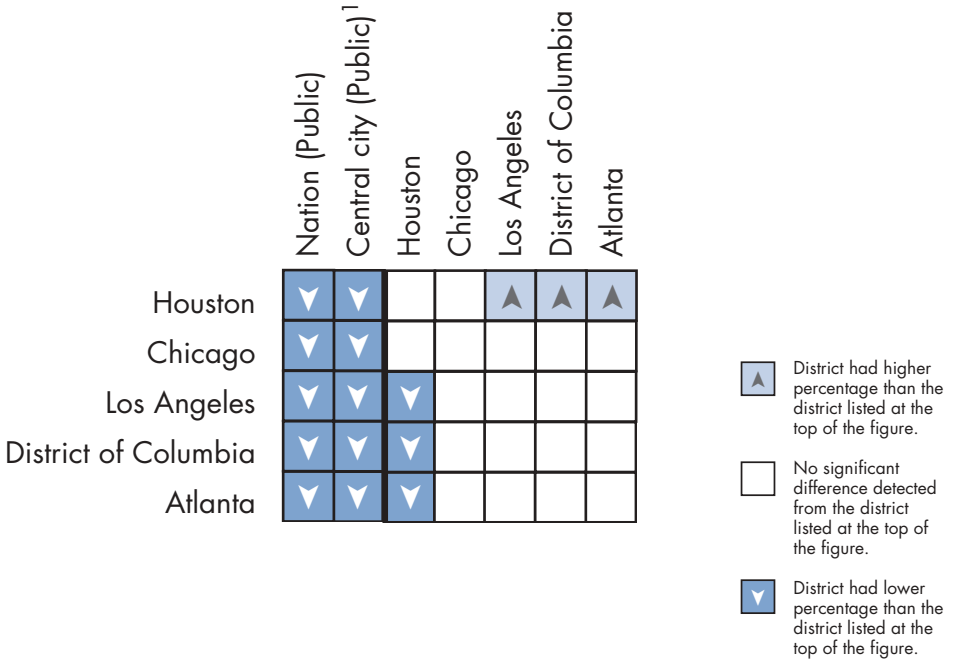
While they are not displayed in figures, there were also interesting patterns in the percentages of students performing at or above *Basic*. At grade 4, students in participating urban districts were less likely than students in public schools in the nation as a whole to perform at *Basic* or higher. However, only four of the urban districts—Atlanta, Chicago, the District of Columbia, and Los Angeles—had lower percentages of students performing at or above *Basic* than central-city schools across the nation. Houston and New York had higher percentages of students performing at or above *Basic* than the other four districts.

At grade 8, a higher percentage of public school students across the nation performed at or above *Basic* than those in the participating urban districts. A higher percentage of students in central city schools across the nation performed at or above *Basic* than did students in Houston, Los Angeles, the District of Columbia, and Atlanta. Students in Chicago and Houston were more likely than those in the District of Columbia, Atlanta, and Los Angeles to perform at or above *Basic*. Finally, students in the District of Columbia were more likely to perform at or above *Basic* than were those in Atlanta and Los Angeles. Other apparent differences were not found to be statistically significant.

Figure 2.4 **Cross-district comparisons of percentage of students at or above *Proficient* in reading, grade 8 public schools: By urban district, 2002**

Grade 8

Instructions: Read across the row corresponding to a district listed to the left of the chart. Match the shading intensity to the key below to determine whether the percentage of students at or above *Proficient* in this district was found to be higher than, not significantly different from, or lower than the district in the column heading. For example, in the row for Houston: The percentage of students at or above *Proficient* in Houston was lower than the national and central city samples, not significantly different from Chicago, and higher than Los Angeles, the District of Columbia, and Atlanta.



[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

3 Results for Subgroups

In addition to reporting on the performance of all students, NAEP provides results for a variety of subgroups of students for each grade level assessed. The subgroup results show how these groups of students performed in comparison with one another. When additional years of urban district assessments are completed, these results will provide a basis for examining each subgroup's progress over time.

This chapter includes the percentage of students, average reading scores, and achievement level results at grades 4 and 8 for subgroups of students in the five urban districts that participated in the Trial Urban District Assessment, for the District of Columbia, and for public schools in the nation and central cities. Results are reported by gender, race/ethnicity, students' eligibility for free/reduced-price lunch, and parents' highest level of education.

Throughout this chapter, the results for students in different subgroups within a district are compared to each other. Student subgroup results in each district are also compared to the central city public-school results. In most cases, the average scores or achievement level results for central city schools were below those for the national public schools. All differences noted as such in this chapter are statistically significant. In interpreting the results, it is important to bear in mind that the estimated average score for a particular group does not include the whole range of performance within that group. Differences in subgroup performance cannot be ascribed solely to students'

membership in an identified subgroup. Average student performance is affected by the interaction of a complex set of factors not discussed in this report or addressed by NAEP assessments.

Performance of Selected Subgroups in the Urban Districts

Gender

Educators and government agencies have produced a body of research that is rich in data documenting gender differences in language arts achievement.[1] National results for the NAEP 1998 and 2002 reading assessments indicated that female students outperformed male students at grades 4, 8, and 12.

Table 3.1 shows average reading scores and achievement level results by gender at grade 4. For five of the six urban districts at grade 4, female students had higher average scores than male students. The apparent four point difference between average scores for male and female students in Houston was not found to be statistically significant. Average scores for male students at grade 4 ranged from 185 in the District of Columbia to 204 in Houston. The central city public school average score for male students was 204. Average scores for male fourth-grade students in Houston and New York City were not found to differ significantly from the average score for central city public schools, but male students in Atlanta, Chicago, the District of Columbia, and Los Angeles scored lower than the average for central city public schools.

Average scores for female students at grade 4 ranged from 194 in Los Angeles to 213 in New York City. The central city public school average score for female fourth-graders was 211. Average scores for female fourth-grade students in Houston and New York City were not found to differ significantly from the average score for central city public schools, while the average score for female students in each of the other districts was lower than the average for central city schools.

[1] U.S. Department of Education. (2001). *Reading for Understanding: Towards an R & D Program in Reading Comprehension.* Washington, DC: Author.

MacMillan, P. (2000). Simultaneous Measurement of Reading Growth, Gender, and Relative-Age Effects: Many Faceted Rasch Applied to CBM Reading Scores. *Journal of Applied Measurement 1*(4), 393–408.

Table 3.1 **Average reading scale scores and percentage of students at or above each achievement level, by gender, grade 4 public schools: By urban district, 2002**

Grade 4	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male						
Nation (Public)	51	214	41	59	26	5
Central city (Public) [1]	50	204 **	53 **	47 **	19 **	3 **
Atlanta	47	191 *,**	69 *,**	31 *,**	11 *,**	2 **
Chicago	50	189 *,**	70 *,**	30 *,**	9 *,**	1 *,**
District of Columbia	49	185 *,**	74 *,**	26 *,**	8 *,**	1 *,**
Houston	51	204 **	55 **	45 **	16 **	3
Los Angeles	51	188 *,**	70 *,**	30 *,**	10 *,**	1 *,**
New York City ‡	50	199 **	61 *,**	39 *,**	14 **	3 **
Female						
Nation (Public)	49	220	35	65	33	8
Central city (Public) [1]	50	211 **	45 **	55 **	24 **	5 **
Atlanta	53	200 *,**	60 *,**	40 *,**	13 *,**	4 **
Chicago	50	198 *,**	62 *,**	38 *,**	12 *,**	2 *,**
District of Columbia	51	196 *,**	64 *,**	36 *,**	11 *,**	2 *,**
Houston	49	208 **	50 **	50 **	19 **	3 **
Los Angeles	49	194 *,**	64 *,**	36 *,**	12 *,**	2 *,**
New York City ‡	50	213 **	45 **	55 **	23 **	7

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table 3.2 displays average scores and achievement level results by gender for grade 8. In every district, female students scored higher, on average, than male students. Average scores for male eighth-grade students ranged from 231 in Atlanta to 245 in Chicago. The average score for eighth-grade male students in central city schools was 250, which was higher than in each of the five urban districts.

For eighth-grade female students, average scores ranged from 240 in Atlanta to 254 in Chicago. The average score for female eighth-grade students in central city public schools was 258, which was higher than the average score in every district except Chicago.

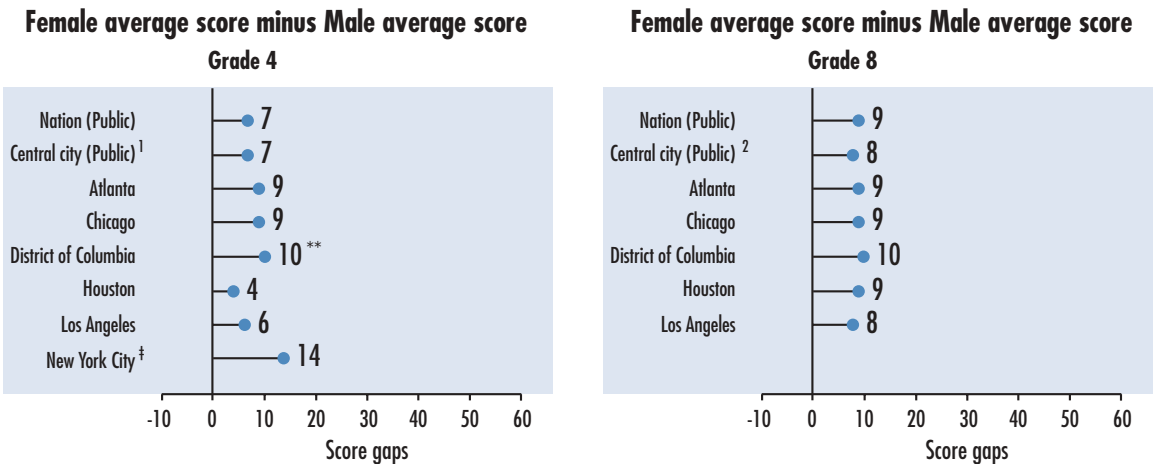
Table 3.2 Average reading scale scores and percentage of students at or above each achievement level, by gender, grade 8 public schools: By urban district, 2002

Grade 8	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male						
Nation (Public)	50	258	30	70	26	2
Central city (Public) [1]	50	250 **	40 **	60 **	19 **	1
Atlanta	49	231 *,**	63 *,**	37 *,**	6 *,**	#
Chicago	50	245 *,**	43 **	57 **	12 *,**	1
District of Columbia	47	235 *,**	58 *,**	42 *,**	9 *,**	#
Houston	51	243 *,**	47 *,**	53 *,**	13 *,**	#
Los Angeles	53	233 *,**	61 *,**	39 *,**	8 *,**	#
Female						
Nation (Public)	50	267	21	79	36	3
Central city (Public) [1]	50	258 **	31 **	69 **	26 **	2 **
Atlanta	51	240 *,**	53 *,**	47 *,**	9 *,**	#
Chicago	50	254 **	33 **	67 **	17 *,**	1
District of Columbia	53	245 *,**	46 *,**	54 *,**	11 *,**	1
Houston	49	253 *,**	35 **	65 **	21 *,**	1 **
Los Angeles	47	241 *,**	51 *,**	49 *,**	12 *,**	1 *,**

Percentage rounds to zero.
[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

An additional way to compare student performance is to focus on the size of the difference in scores or "gap" between the subgroup average scores. Figure 3.1 presents these "gaps" in average reading scores between female and male students in each district as well as in the national and central city public schools. At grade 4, the gap between the average scores of female and male students in the District of Columbia was wider than the gap in public schools in the nation. The average-score gaps between male and female students in all the other participating urban districts at both grades 4 and 8 were not found to differ significantly from the average-score gaps in public schools in central cities or the nation.

Figure 3.1 Gaps in average reading scale scores, by gender, grades 4 and 8: By urban district, 2002



‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
1 For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
2 For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
** Significantly different from nation (public schools).
NOTE: Score gaps are calculated based on differences between unrounded average scale scores.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Race/Ethnicity

For the purpose of studying the progress of subgroups, NAEP collects information from school records on the racial/ethnic identification that best describes each participating student. The six mutually exclusive categories are White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. For further details, see "NAEP Reporting Groups" in appendix C.

Tables 3.3 and 3.4 show average scores and achievement level results by racial/ethnic group membership for public-school students in the urban districts at grades 4 and 8, respectively. Only the race/ethnicity categories with sufficient membership to meet reporting requirements in the urban districts are reported below.

The distribution of students in terms of race/ethnicity in the urban districts differs from that of the national distribution in public schools. Whereas White students comprise 60 percent of the national public sample at grade 4 and 64 percent at grade 8, in the urban districts shown here White students make up a maximum of 15 percent of the district samples at grade 4 (New York City) and 11 percent at grade 8 (Chicago). Black and/or Hispanic students constitute the majority in each of the urban districts in the trial assessment. Hispanic students made up half or more of the sample in Houston and Los Angeles at both grades 4 and 8. Black fourth- and eighth-grade students made up more than 80 percent of the sample in both Atlanta and the District of Columbia.

In five of the six of urban districts in which a reliable comparison could be made, White fourth-graders had higher average scores than their Black and Hispanic peers. In Chicago, Hispanic fourth-graders had higher average scores than Black fourth-graders. Any apparent difference between Hispanic and Black students' average scores in the District of Columbia, Houston, Los Angeles, and New York City was not found to be statistically significant. The White/Hispanic and the Hispanic/Black comparisons in Atlanta could not be tested for statistical significance due to insufficient sample size.

At grade 4, Black students in Chicago, the District of Columbia, and Los Angeles scored lower than their counterparts in central city public schools. Hispanic students in Chicago, Houston, and Los Angeles also scored lower, on average, than their counterparts in central city public schools. White students in Atlanta and the District of Columbia scored higher than their counterparts in central city public schools. Average scores for White students in Chicago, Houston, Los Angeles, and New York City were not found to differ significantly from those of their counterparts in central city public schools. Asian/Pacific Islander fourth-graders in New York City had higher average scores than their counterparts in central city public schools. The average score for Asian/Pacific Islander students in Los Angeles was not found to differ significantly from that of Asian/Pacific Islander students in central city schools. For the other districts, the samples of Asian/Pacific Islander fourth-graders were insufficient for reliable significance testing.

Table 3.3 Average reading scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 4 public schools: By urban district, 2002

Grade 4	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
White						
Nation (Public)	60	227	26	74	39	9
Central city (Public) [1]	35	225 **	29 **	71 **	37 **	9
Atlanta	6	250 *,**	14 *,**	86 *,**	67 *,**	34 *,**
Chicago	10	221	36 *,**	64 *,**	35	9
District of Columbia	3	248 *,**	9 *,**	91 *,**	66 *,**	28 *,**
Houston	10	233	21	79	45	13
Los Angeles	9	223	30	70	38	9
New York City ‡	15	226	29	71	35	10
Black						
Nation (Public)	18	198	61	39	12	1
Central city (Public) [1]	31	194 **	65 **	35 **	10 **	1
Atlanta	90	192 **	68 **	32 **	8 **	1
Chicago	48	185 *,**	75 *,**	25 *,**	5 *,**	#
District of Columbia	88	188 *,**	72 *,**	28 *,**	7 *,**	1 **
Houston	37	200	60	40	12	1
Los Angeles	12	186 *,**	75 *,**	25 *,**	6 **	#
New York City ‡	36	197	63	37	9	2
Hispanic						
Nation (Public)	17	199	57	43	14	2
Central city (Public) [1]	27	199	59	41	13	2
Atlanta	3	***	***	***	***	***
Chicago	37	193 *,**	67 *,**	33 *,**	9 *,**	1
District of Columbia	7	193	66	34	8 **	1
Houston	50	203 *	55	45	14	2
Los Angeles	72	185 *,**	74 *,**	26 *,**	7 *,**	1 *,**
New York City ‡	40	201	58	42	15	3
Asian/Pacific Islander						
Nation (Public)	4	223	31	69	36	9
Central city (Public) [1]	5	217 **	38 **	62 **	29 **	7 **
Atlanta	#	***	***	***	***	***
Chicago	3	***	***	***	***	***
District of Columbia	1	***	***	***	***	***
Houston	3	***	***	***	***	***
Los Angeles	6	218	30 *	70 *	26	3
New York City ‡	8	235 *	22 *	78 *	50 *	20 *
American Indian/Alaska Native						
Nation (Public)	1	207	49	51	22	5
Central city (Public) [1]	1	207	50	50	23	6
Atlanta	#	***	***	***	***	***
Chicago	1	***	***	***	***	***
District of Columbia	#	***	***	***	***	***
Houston	#	***	***	***	***	***
Los Angeles	1	***	***	***	***	***
New York City ‡	#	***	***	***	***	***

Percentage rounds to zero.
‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
*** Sample size is insufficient to permit a reliable estimate.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table 3.4 displays the performance results by race/ethnicity for grade 8. White eighth-graders outperformed Black eighth-graders, on average, in Atlanta, Houston, and Los Angeles. The apparent difference in Chicago was not found to be statistically significant, and the sample size in the District of Columbia was insufficient to permit a reliable comparison. White eighth-graders also had higher average scores than their Hispanic peers in Houston and Los Angeles. The apparent difference in Chicago was not found to be statistically significant, and the sample sizes in Atlanta and the District of Columbia were insufficient to permit a reliable comparison. Black eighth-graders in Los Angeles outperformed their Hispanic peers, on average. The apparent difference in Chicago, the District of Columbia, and Houston between these two groups of students was not statistically significant, and the sample size was insufficient in Atlanta to permit a reliable comparison.

Black eighth-graders in Atlanta, the District of Columbia, and Los Angeles had lower average scores than their counterparts in central city public schools. Los Angeles was the only district where Hispanic eighth-graders scored lower, on average, than Hispanic students in central city public schools. The other apparent differences between districts and central city public schools for Black and Hispanic students' average scores were not found to be statistically significant. The average score for White students in Houston was higher than the average for White students in central city public schools, while all other districts were not found to differ significantly from the central city public schools. Average scores for Asian/Pacific Islander students in Los Angeles were not found to differ significantly from the scores of their counterparts in central city schools. The sample of Asian/Pacific Islander students in each of the other districts was too small to support significance testing.

Table 3.4 Average reading scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 8 public schools: By urban district, 2002

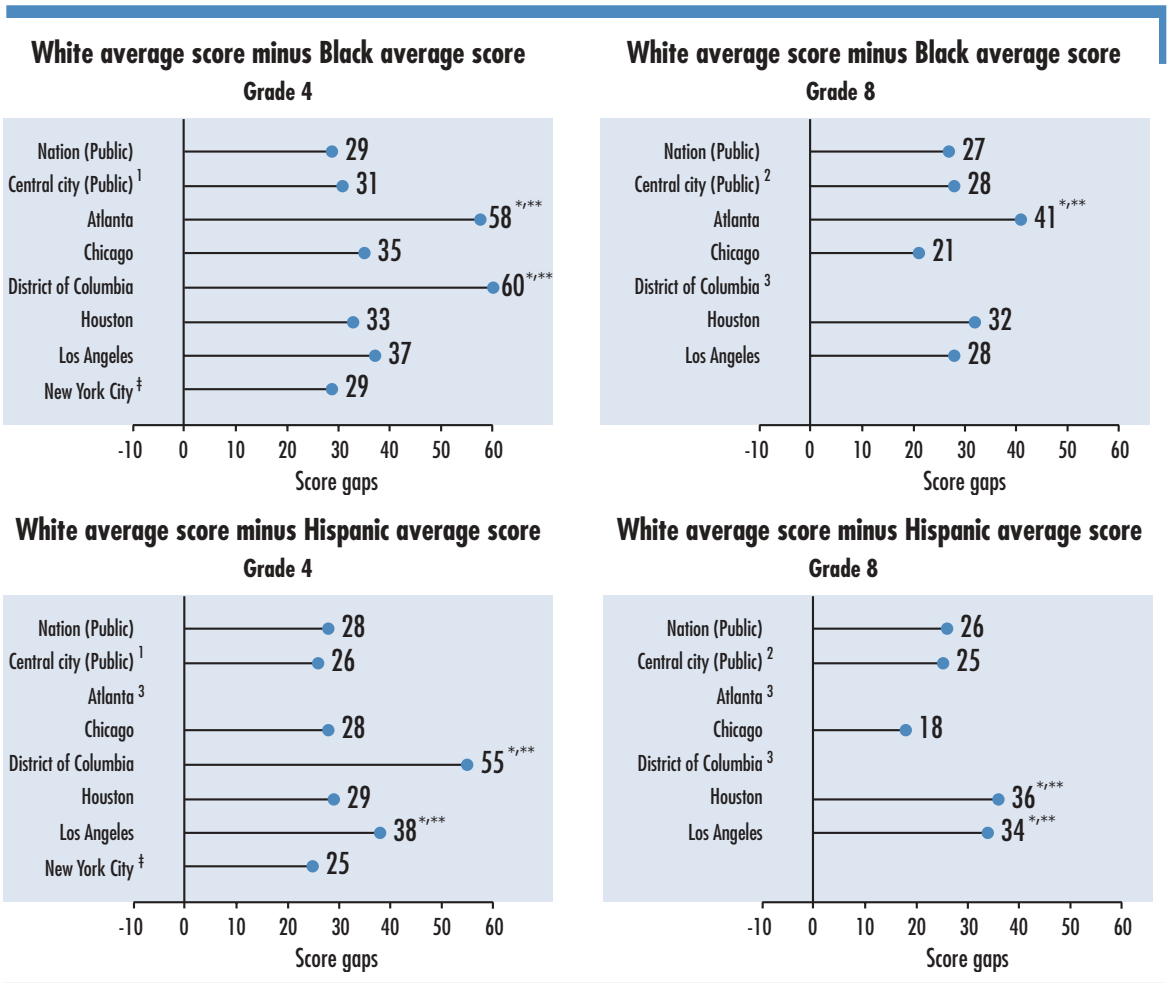
Grade 8	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
White						
Nation (Public)	64	271	17	83	39	3
Central city (Public) [1]	39	269 **	20 **	80 **	37	3
Atlanta	5	275	16	84	47	5
Chicago	11	266	25	75	31	5
District of Columbia	3	***	***	***	***	***
Houston	8	279 *,**	13	87	47	5
Los Angeles	10	264 **	27 **	73 **	33	3
Black						
Nation (Public)	15	244	46	54	13	#
Central city (Public) [1]	28	241 **	50 **	50 **	11 **	#
Atlanta	92	233 *,**	61 *,**	39 *,**	5 *,**	#
Chicago	50	245	43 *	57 *	10	#
District of Columbia	88	238 *,**	54 *,**	46 *,**	8 *,**	#
Houston	31	247	40	60	15	#
Los Angeles	14	236 *,**	57 **	43 **	8	#
Hispanic						
Nation (Public)	15	245	44	56	14	#
Central city (Public) [1]	25	244	46	54	13	#
Atlanta	2	***	***	***	***	***
Chicago	35	248	39	61	12	#
District of Columbia	7	240	47	53	11	#
Houston	58	243	48	52	13	#
Los Angeles	67	230 *,**	64 *,**	36 *,**	5 *,**	#
Asian/Pacific Islander						
Nation (Public)	4	265	25	75	34	3
Central city (Public) [1]	6	258 **	32 **	68 **	27 **	1 **
Atlanta	1	***	***	***	***	***
Chicago	2	***	***	***	***	***
District of Columbia	2	***	***	***	***	***
Houston	3	***	***	***	***	***
Los Angeles	9	259	27	73	26	1
American Indian/Alaska Native						
Nation (Public)	1	252	36	64	18	1
Central city (Public) [1]	1	258	28	72	21	2
Atlanta	#	***	***	***	***	***
Chicago	1	***	***	***	***	***
District of Columbia	#	***	***	***	***	***
Houston	#	***	***	***	***	***
Los Angeles	#	***	***	***	***	***

Percentage rounds to zero.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
*** Sample size is insufficient to permit a reliable estimate.
[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Average scale score differences or "gaps" between White and Black students and between White and Hispanic students are presented in figure 3.2. At grade 4, the gaps between the average scores for White and Black students in Atlanta and the District of Columbia were wider than the corresponding gaps between White and Black students in public schools in the nation and in central cities. The average-score gaps between White and Hispanic fourth-graders in the District of Columbia and Los Angeles were wider than the corresponding gaps in the nation and in central cities.

At grade 8, the average-score gap between White and Black students in Atlanta was wider than the corresponding gaps in public schools in the nation and in central cities. The average-score gaps between White and Hispanic eighth-graders in Houston and Los Angeles were wider than the corresponding gaps in the nation and in central cities.

Figure 3.2 Gaps in average reading scale scores, by race/ethnicity, grades 4 and 8: By urban district, 2002



‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
[3] Sample sizes were insufficient to permit a reliable estimate for fourth-and eighth-grade Hispanic students in Atlanta and for eighth-grade White students in the District of Columbia.
NOTE: Score gaps are calculated based on differences between unrounded average scale scores.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Free/Reduced-Price Lunch Eligibility

NAEP collects data on a variety of factors that may affect the context of learning. One of these is eligibility for the federal program providing free/reduced-price school lunches. 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. This program is available to public schools, nonprofit private schools, and residential childcare institutions. Eligibility is determined through the USDA's Income Eligibility Guidelines, and is an indicator of low income.[2] Tables 3.5 and 3.6 present these data for grades 4 and 8.

At grade 4, the percentages of students from the urban districts eligible for free/reduced-price lunch ranged from 72 percent in Houston to 88 percent in Chicago. At grade 8, the percentages in four of the urban districts ranged from 68 percent to 84 percent. Information on the free/reduced-price lunch data for Los Angeles at grade 8 is not reported because the data received did not meet reporting standards.

NAEP national report cards across all subjects have consistently reported lower performance scores for students eligible for free/reduced-price lunch than for those who were not eligible. This pattern was also generally the case in the urban district data. At grade 4, students not eligible for the free/reduced-price lunch had higher average scores than those eligible in every district.

Fourth-grade students eligible for free/reduced-price lunch in Atlanta, Chicago, the District of Columbia, and Los Angeles had lower average scores than their eligible counterparts in central city public schools. No significant difference was found between the average scores for eligible students in central city public schools and those in Houston and New York City. In comparison with ineligible students in central city public schools, ineligible students in Atlanta, the District of Columbia, and Los Angeles scored lower. No significant difference was found between the central city public schools and the other districts for students who were not eligible.

[2] U.S. General Services Administration. (2001). *Catalog of Federal Domestic Assistance*. Washington, DC: Executive Office of the President, Office of Management and Budget.

Table 3.5 Average reading scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 4 public schools: By urban district, 2002

Grade 4	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Eligible						
Nation (Public)	43	202	54	46	16	2
Central city (Public) [1]	61	198 **	60 **	40 **	13 **	2
Atlanta	74	189 *,**	71 *,**	29 *,**	7 *,**	1 *,**
Chicago	88	190 *,**	70 *,**	30 *,**	8 *,**	1 *,**
District of Columbia	78	185 *,**	75 *,**	25 *,**	5 *,**	# *,**
Houston	72	199	60	40	11 **	1
Los Angeles	79	186 *,**	73 *,**	27 *,**	7 *,**	1 *,**
New York City ‡	73	201	58	42	15	3
Not eligible						
Nation (Public)	50	229	24	76	41	10
Central city (Public) [1]	33	225 **	29 **	71 **	37 **	9 **
Atlanta	16	214 *,**	45 *,**	55 *,**	27 *,**	10
Chicago	8	222	35 **	65 **	33	11
District of Columbia	21	210 *,**	48 *,**	52 *,**	23 *,**	7 **
Houston	24	226	28	72	39	9
Los Angeles	5	199 *,**	58 *,**	42 *,**	14 *,**	1
New York City ‡	16	219 **	38 **	62 **	30	8
Information not available [2]						
Nation (Public)	7	217	38	62	30	7
Central city (Public) [1]	6	210 **	47 **	53 **	24 **	5
Atlanta	11	211	51	49	22	7
Chicago	4	206	52 **	48 **	19	4
District of Columbia	1	***	***	***	***	***
Houston	4	***	***	***	***	***
Los Angeles	16	215	40	60	28	6
New York City ‡	11	221	38	62	28	11

Percentage rounds to zero.
‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
*** Sample size is insufficient to permit a reliable estimate.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
[2] 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."
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table 3.6 shows student performance data at grade 8 by free/reduced-price lunch eligibility. Eighth-graders not eligible for free/reduced-price lunch had higher average scores in comparison to eligible students in every district except Chicago, where the apparent difference was not statistically significant.

In comparison with central city public schools, the average scores for eligible students were lower in Atlanta and the District of Columbia, and were not found to differ significantly in Chicago and Houston. Among students who were not eligible, average scores in Atlanta, the District of Columbia, and Houston were lower than in central city schools. No significant difference was detected between central city schools and Chicago in the average scores of students who were not eligible for free/reduced-price school lunch.

Table 3.6 Average reading scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 8 public schools: By urban district, 2002

Grade 8	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Eligible						
Nation (Public)	34	249	40	60	17	1
Central city (Public) [1]	47	244 **	47 **	53 **	12 **	# **
Atlanta	76	233 *,**	62 *,**	38 *,**	6 *,**	#
Chicago	84	246	41	59	11 **	#
District of Columbia	68	235 *,**	57 *,**	43 *,**	6 *,**	#
Houston	68	243 **	48 **	52 **	13	#
Not eligible						
Nation (Public)	57	271	17	83	40	3
Central city (Public) [1]	41	268 **	21 **	79 **	36 **	3
Atlanta	20	244 *,**	47 *,**	53 *,**	12 *,**	1
Chicago	10	267	24	76	36	4
District of Columbia	31	251 *,**	39 *,**	61 *,**	18 *,**	1 *,**
Houston	29	261 *,**	25	75	26 *,**	2
Information not available [2]						
Nation (Public)	10	264	25	75	32	4
Central city (Public) [1]	12	253 **	36 **	64 **	21 **	2
Atlanta	4	***	***	***	***	***
Chicago	6	268 *	21 *	79 *	34	7
District of Columbia	1	***	***	***	***	***
Houston	3	***	***	***	***	***

Percentage rounds to zero.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
*** Sample size is insufficient to permit a reliable estimate.
[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White.
[2] 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."
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Parents' Highest Level of Education

Eighth-grade students who participated in the NAEP 2002 national reading assessment, including those in the Trial Urban District Assessment, were asked to indicate the highest level of education completed by each parent. Five response options were offered: did not finish high school, graduated from high school, some education after high school, graduated from college, or "I don't know." The highest level of education reported for either parent was used in the analysis of this question. The question was not posed to fourth-graders.

Table 3.7 presents the data for the urban districts at grade 8. NAEP assessments typically find a positive relationship between student reported parental education and student achievement: the higher the parental education level, the higher the average reading score.[3] This pattern was also generally the case for the trial urban districts.

The District of Columbia had the highest percentage of eighth-graders (40 percent) who reported that at least one parent had graduated from college. However, for all of the urban districts, the percentage of students who reported that at least one parent graduated from college was lower than that of public schools nationally. Houston and Los Angeles had the highest percentages of students who reported that neither parent had finished high school (21 and 19 percent, respectively). For these two districts and Chicago, the percentage of students who reported that neither parent finished high school was greater than that for national public schools.

The apparent difference in average scores between students in central city public schools who reported that at least one parent graduated from college and their counterparts in Houston was not found to be statistically significant. However, students in Atlanta, Chicago, the District of Columbia, and Los Angeles who reported that at least one parent graduated from college scored lower, on average, than their counterparts in central city public schools. Students in Atlanta and Los Angeles who reported that neither parent had finished high school had lower scores than their counterparts in central city public schools. Students in Houston who reported that neither parent graduated from high school scored higher, on average, than their counterparts in central city public schools.

[3] Grigg, W. S., Daane, M. C., Jin, Y., and Campbell, J. R. (2003). *The Nation's Report Card: Reading 2002* (NCES 2003–521). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Table 3.7 Average reading scale scores and percentage of students at or above each achievement level, by student-reported parents' highest level of education, grade 8 public schools: By urban district, 2002

Grade 8	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Less than high school						
Nation (Public)	7	247	42	58	14	#
Central city (Public) [1]	9	243 **	48	52	11	#
Atlanta	7	233 *,**	66 *,**	34 *,**	8	#
Chicago	14	246	43	57	10	#
District of Columbia	7	240 **	46	54	6 **	#
Houston	21	251 *	38	62	17	1
Los Angeles	19	234 *,**	61 *,**	39 *,**	7 **	#
Graduated high school						
Nation (Public)	18	256	31	69	21	1
Central city (Public) [1]	18	248 **	42 **	58 **	13 **	#
Atlanta	26	233 *,**	63 *,**	37 *,**	4 *,**	#
Chicago	19	246 **	40	60	9 *,**	#
District of Columbia	21	235 *,**	57 *,**	43 *,**	5 *,**	#
Houston	19	242 **	48 **	52 **	9 **	#
Los Angeles	14	233 *,**	61 *,**	39 *,**	5 *,**	#
Some education after high school						
Nation (Public)	20	267	19	81	33	2
Central city (Public) [1]	19	260 **	27 **	73 **	25 **	1 **
Atlanta	22	241 *,**	50 *,**	50 *,**	8 *,**	#
Chicago	22	260 **	24	76	20 **	1
District of Columbia	18	247 *,**	43 *,**	57 *,**	12 *,**	#
Houston	15	260 **	25	75	24	1
Los Angeles	16	249 *,**	40 *,**	60 *,**	17 *,**	1
Graduated college						
Nation (Public)	46	273	17	83	42	4
Central city (Public) [1]	41	266 **	25 **	75 **	34 **	3
Atlanta	35	243 *,**	49 *,**	51 *,**	13 *,**	1 *,**
Chicago	31	255 *,**	33 *,**	67 *,**	20 *,**	3
District of Columbia	40	247 *,**	45 *,**	55 *,**	15 *,**	1 *,**
Houston	28	262 **	26 **	74 **	29 **	2
Los Angeles	26	251 *,**	40 *,**	60 *,**	21 *,**	1 *,**
Unknown						
Nation (Public)	9	246	44	56	14	#
Central city (Public) [1]	12	241 **	50 **	50 **	11 **	#
Atlanta	10	229 *,**	67 *,**	33 *,**	4 *,**	#
Chicago	15	242	48	52	11	#
District of Columbia	14	231 *,**	65 *,**	35 *,**	5 *,**	#
Houston	17	235 **	57 **	43 **	7	#
Los Angeles	26	228 *,**	67 *,**	33 *,**	4 *,**	#

Percentage rounds to zero.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Appendix A
Identification, Exclusion, and Accommodation Rates for Special-Needs Students

The following appendix presents the percentages of students with disabilities and limited English proficient students who were identified, excluded, and assessed with accommodations in the Trial Urban District Assessment. In interpreting the performance results presented in this report, it is important to consider the magnitude of rates of identification and exclusion of these special-needs students. At grade 4, there was considerable variation in exclusion rates for limited English proficient students among the districts.

Students with Disabilities (SD) and/or Limited English Proficient (LEP) Students

It is NAEP's intent to assess all selected students from the target population. 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 were revised in 1996 to communicate more clearly a presumption of inclusion except under special circumstances. According to these criteria, students who have an Individualized Education Program (IEP) or are protected under section 504 of the Rehabilitation Act of 1973 are to be included in the NAEP assessment except in the following cases:

- the school's IEP team determines that the student cannot participate;
- the student's cognitive functioning is so severely impaired that she or he cannot participate;
- the student's IEP requires that the student has to be tested with an accommodation or adaptation that NAEP does not allow and the student cannot demonstrate his or her knowledge without that accommodation.

All LEP students who received academic instruction in English for three years or more were to be included in the assessment. Those LEP students who received instruction in English for fewer than three years were to be included unless school staff judged them to be incapable of participating in the assessment in English.

Participation of SD and/or LEP Students in the NAEP Samples

Testing all sampled students is the best way for NAEP to ensure that the statistics generated by the assessment are as representative as possible of the performance of the entire national population and the populations of participating jurisdictions and districts. However, all groups of students include certain proportions that cannot be tested in large-scale assessments (such as students who have profound mental disabilities) or who can only be tested through the use of "accommodations," such as extra time, one-on-one administration, or use of magnifying equipment. Some students with disabilities and some LEP students cannot show on a written test what they know and can do unless they are provided with accommodations. When such accommodations are not allowed, students requiring such adjustments are often excluded from large-scale assessments such as NAEP. Section 504 of the Rehabilitation Act of 1973 requires that, when students with disabilities are tested, schools must provide them with appropriate accommodations so that the test results accurately reflect students' achievement. The provision of accommodations has become more common with the passage of the 1997 Individuals with Disabilities Education Act (IDEA), which led schools and states to identify increasing proportions of students as needing accommodations on assessments in order to best show what they know and can do.[1]

[1] Office of Special Education Programs. (1997). *Nineteenth Annual Report to Congress on the Implementation of the Individuals With Disabilities Education Act*. Washington, DC: U. S. Department of Education.

In addition, as the proportion of limited English proficient students in the population has increased, some states have started offering accommodations, such as translations of assessments or the use of bilingual dictionaries, as part of assessments.

Before 1996, NAEP did not allow any testing under nonstandard conditions (i.e., accommodations were not permitted). At that time, NAEP samples were able to include almost all sampled students in standard assessment sessions. However, as the influence of IDEA grew more widespread, the failure to provide accommodations led to increasing levels of exclusion in the assessment. NAEP's response was to begin a research program to assess the impact on NAEP's scales and trends of permitting accommodations for those students who normally received them in classroom assessments. Beginning in 2002, NAEP uses only the more inclusive samples in which assessment accommodations are permitted. Consequently the data reported here for urban districts represent samples in which accommodations were permitted.

Percentages of SD and/or LEP students for the Trial Urban District Assessment are presented in table A.1. The data in this table include the percentages of students *identified* as SD and/or LEP, the percentage of students *excluded*, and the percentage of *accommodated* SD and/or LEP students. Different rates of exclusion may influence the meaning of district comparisons. Thus, exclusion data should be carefully reviewed in this context.

The percentage of fourth-graders in participating urban districts excluded from the assessment ranged from two percent in Atlanta to 17 percent in Houston. At grade 4, the percentage of students excluded in the national public school sample was seven percent, and the percentage of students excluded in the national central city public school sample was eight percent. The percentage of eighth-graders in participating urban districts excluded from the assessment ranged from two percent in Atlanta to nine percent in New York City. At grade 8, the percentage of students excluded in the national public school sample was six percent, and the percentage of students excluded in the national central city public school sample was seven percent.

Table A.1 **Percentage of students with disabilities and limited English proficient students identified, excluded, and assessed with accommodations, grades 4 and 8 public schools: By urban district, 2002**

	Students with disabilities and/or limited English proficient students			Students with disabilities			Limited English proficient students		
	Identified	Excluded	Assessed with accommodations	Identified	Excluded	Assessed with accommodations	Identified	Excluded	Assessed with accommodations
Grade 4									
Nation (Public)	21	7	4	13	5	4	9	2	1
Central city (Public) [1]	25	8	4	13	6	3	14	4	1
Atlanta	8	2	1	5	1	1	4	1	0
Chicago	30	9	5	16	4	4	19	7	2
District of Columbia	19	8	5	14	7	4	7	3	2
Houston	43	17	1	12	4	1	36	16	#
Los Angeles	51	8	2	11	3	2	46	6	1
New York City ‡	22	8	8	14	5	6	11	6	3
Grade 8									
Nation (Public)	18	6	4	13	5	4	6	2	1
Central city (Public) [2]	22	7	4	14	5	3	10	3	1
Atlanta	6	2	1	5	1	1	1	#	#
Chicago	21	6	7	15	3	6	8	4	1
District of Columbia	21	7	8	16	6	7	5	2	2
Houston	27	7	0	15	5	0	16	4	0
Los Angeles	35	5	2	12	3	2	30	5	1

Percentage rounds to zero.
‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Trial Urban District Assessment.

In the 2002 national combined sample (public and nonpublic schools), 6 percent of students at grade 4, and 5 percent of students at grade 8 were excluded from the assessment (see table A.10 of the *The Nation's Report Card: Reading 2002*).[2] Across the various states and jurisdictions that participated in the 2002 state assessment, the percentage of students excluded ranged from 3 to 12 percent at grade 4 (see table A.11 in the national report card for reading 2002) and from 2 to 10 percent at grade 8 (see table A.12 in the national report card).

Types of Accommodations Permitted

Table A.2 displays the percentages of SD and/or LEP students assessed with the variety of available accommodations at grade 4. Table A.3 shows the comparable data for grade 8. It should be noted that students assessed with accommodations typically received some combination of accommodations. The numbers and percentages presented in the table reflect only the primary accommodation provided. For example, students assessed in small groups

[2] Grigg, W. S., Daane, M. C., Jin, Y., and Campbell, J. R. (2003). *The Nation's Report Card: Reading 2002* (NCES 2003–521). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

(as compared with standard NAEP sessions of about 30 students) usually received extended time. In one-on-one administrations, students often received assistance in recording answers and were afforded extra time. Extended time was considered the primary accommodation only when it was the sole accommodation provided.

The assessment did not allow some accommodations that were permitted in certain states in past assessments. Some states have allowed questions and, in some cases, reading passages to be read aloud to the students. In designing the reading assessment, reading aloud as an accommodation was viewed as changing the nature of the construct being measured, and hence was not permitted. Because NAEP considers the domain of its reading assessment to be reading in English, no attempt was made to provide an alternate language version of the assessment, and the use of bilingual dictionaries was not permitted.

Table A.2 Students with disabilities and limited English proficient students assessed with accommodations, by type of primary accommodation, grade 4: By urban district, 2002

	Weighted percentage of students sampled					
	Large-print book	Extended time	Small group	One-on-one	Scribe/ computer	Other
SD[1] and/or LEP[2] students						
Nation (Public)	0.04	1.57	2.23	0.08	0.06	0.03
Central city (Public) [3]	0.02	2.09	1.96	0.05	0.04	0.03
Atlanta	0.06	0.77	0.06	#	#	#
Chicago	#	3.67	1.30	#	0.05	0.10
District of Columbia	#	2.80	2.45	#	#	0.07
Houston	#	0.32	0.73	#	#	#
Los Angeles	0.05	1.08	1.23	0.12	#	#
New York City ‡	#	6.61	1.49	#	#	#
SD[1] students						
Nation (Public)	0.04	1.24	2.08	0.08	0.06	0.03
Central city (Public) [3]	0.02	1.53	1.83	0.04	0.04	0.02
Atlanta	0.06	0.77	0.06	#	#	#
Chicago	#	2.51	1.30	#	0.05	0.10
District of Columbia	#	2.11	1.63	#	#	0.07
Houston	#	0.32	0.73	#	#	#
Los Angeles	0.05	0.98	1.05	0.12	#	#
New York City ‡	#	5.18	0.92	#	#	#
LEP[2] students						
Nation (Public)	#	0.44	0.25	0.01	#	0.01
Central city (Public) [3]	#	0.79	0.25	0.01	#	0.01
Atlanta	#	#	#	#	#	#
Chicago	#	1.88	0.35	#	#	0.05
District of Columbia	#	0.77	1.03	#	#	0.03
Houston	#	0.16	#	#	#	#
Los Angeles	#	0.56	0.82	0.07	#	#
New York City ‡	#	1.94	0.66	#	#	#

Percentage rounds to zero.
‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
[1] Students with disabilities.
[2] Limited English proficient students.
[3] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table A.3 Students with disabilities and limited English proficient students assessed with accommodations, by type of primary accommodation, grade 8: By urban district, 2002

	Weighted percentage of students sampled					
	Large-print book	Extended time	Small group	One-on-one	Scribe/ computer	Other
SD[1] and/or LEP[2] students						
Nation (Public)	0.02	2.06	1.69	0.05	0.03	0.04
Central city (Public) [3]	0.01	2.25	1.52	0.05	0.01	0.06
Atlanta	#	#	1.00	#	#	#
Chicago	#	2.50	4.05	#	#	#
District of Columbia	0.06	6.02	1.95	#	#	#
Houston	#	#	#	#	#	#
Los Angeles	#	1.05	1.18	0.19	#	#
SD[1] students						
Nation (Public)	0.02	1.82	1.61	0.05	0.03	0.04
Central city (Public) [3]	0.01	1.80	1.45	0.05	0.01	0.06
Atlanta	#	#	1.00	#	#	#
Chicago	#	2.19	3.95	#	#	#
District of Columbia	0.06	4.58	1.95	#	#	#
Houston	#	#	#	#	#	#
Los Angeles	#	1.05	1.18	0.19	#	#
LEP[2] students						
Nation (Public)	#	0.39	0.14	#	#	#
Central city (Public) [3]	#	0.72	0.21	#	#	#
Atlanta	#	#	#	#	#	#
Chicago	#	0.65	0.38	#	#	#
District of Columbia	#	1.86	0.06	#	#	#
Houston	#	#	#	#	#	#
Los Angeles	#	0.58	0.88	#	#	#

Percentage rounds to zero.
[1] Students with disabilities.
[2] Limited English proficient students.
[3] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Investigating the Potential Effects of Exclusion Rates on Assessment Results

In considering the effects of exclusion rates on assessment results, at least two major issues are evident. First, if exclusion rates vary substantially across assessment years, then the ability to report trends (i.e., compare results between years) may be affected by the fact that the results from different years are based on different proportions of the population. Second, the variation in exclusion rates among states, jurisdictions, and school districts may threaten the comparison of results within a given year, because the results for different districts, states, or jurisdictions are based on different proportions of the populations.

NCES has funded research investigating ways in which excluded students might be included in the estimation of scores for total populations. NCES has also commissioned studies on the impact of assessment accommodations on overall scores. A detailed discussion of some of these estimation procedures is included in appendix A of the full report card.[3]

[3] Grigg, W. S., Daane, M. C., Jin, Y., and Campbell, J. R. (2003). *The Nation's Report Card: Reading 2002* (NCES 2003–521). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

B Appendix B
District-Level Contextual Variables

To help place results from the NAEP 2002 Trial Urban District Assessment into context, this appendix presents selected district-level data from sources other than NAEP.

These data are taken from the *Public Elementary/Secondary School Universe Survey, 2000–01*, and *Local Education Agency Universe Survey, 2000–01*. The interested reader can access most of the data presented in this appendix in the NCES report, *Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States* at *http://nces.ed.gov/pubs2001/100_largest/tab_fig.asp*.

Table B.1 Number of students, high-school completers, teachers, and schools, from non-NAEP sources: By urban district, school year 2000–01

	Number of students[1]	Number of 1999–2000 completers[2]	Number of full-time equivalent teachers	Number of schools
National	47,086,931	2,548,076	2,841,677	93,344
Central city	13,523,126	610,467	808,288	22,310
Atlanta	58,230	2,056	3,950	98
Chicago	435,261	14,875	23,935	602
District of Columbia	68,925	2,916	5,044	165
Houston	208,462	7,735	11,197	289
Los Angeles	721,346	27,439	35,150	659
New York City	1,066,516	40,827	65,242	1,213

[1] Count of students receiving educational services from school district may differ somewhat from the counts in table B.4, which reflect the count of students from the schools aggregated up to the school district.
[2] Includes high school diploma recipients as well as other high school completers (e.g., certificates of attendance), but does not include high school equivalencies (e.g., GEDs).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.2 Poverty rate and federal funding, from non-NAEP sources: By urban district, 1996–97, 1997–98, 2001–02

	5- to 17-year olds, 1999–2000		Revenue, 1999–2000 (in thousands)				Title I allocations, 2001–02 (in thousands)	
	Population	Percent in poverty	Total	Federal	Federal as percent of total	Federal revenue per student	Basic grants	Concentration grants
Atlanta	66,131	34.6	$695,919	$56,558	8.1	$95.17	$752	$20,824
Chicago	540,667	24.8	3,604,873	539,567	15	124.97	963	137,865
District of Columbia	82,456	27.8	881,423	184,825	21	261.19	1,509	21,341
Houston	230,514	26	1,469,074	152,679	10.4	72.8	585	50,084
Los Angeles	863,656	28.3	5,757,819	592,508	10.3	83.45	668	178,088
New York City	1,397,739	28.3	10,945,650	1,119,944	10.2	104.11	812	397,541

NOTE: Federal revenue per student based on fall enrollment collected by the Bureau of the Census. Detail may not sum to totals, because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data survey; and U.S. Department of Commerce, "Survey of Local Government Finances."

Table B.3 Number of public elementary and secondary schools, by type of school, from non-NAEP sources: By urban district, school year 2000–01

	Total number of schools	Type of school [1]			
		Regular	Special education	Vocational education	Alternative
National	93,344	85,493	2,008	1,025	4,818
Central city	22,310	20,141	536	173	1,460
Atlanta	98	95	0	0	3
Chicago	602	578	24	0	0
District of Columbia	165	150	10	0	5
Houston	289	264	1	3	21
Los Angeles	659	582	18	0	59
New York City	1,213	1,106	8	18	81

[1] Type of school is a mutually exclusive category on the Common Core of Data. There are cases in which special education, vocational education, and alternative programs reside in other types of schools.
NOTE: Types of schools are defined in the following way on the Common Core of Data: Regular school—A public elementary/secondary school that does not focus primarily on vocational, special, or alternative education. NAEP is conducted only in regular schools. Special education school—A public elementary/secondary school that (a) focuses primarily on special education, including instruction for any of the following: hard of hearing, deaf, speech-impaired, health-impaired, orthopedically impaired, mentally retarded, seriously emotionally disturbed, multi-handicapped, visually handicapped, deaf and blind; and (b) adapts curriculum, materials, or instruction for students served. Vocational education school—A public elementary/secondary school that focuses primarily on vocational education, and provides education and training in one or more semi-skilled or technical operations. Alternative education school—A public elementary/secondary school that (a) addresses the needs of students that typically cannot be met in a regular school; (b) provides nontraditional education; (c) serves as an adjunct to a regular school; and (d) falls outside of the categories of regular, special education, or vocational education.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a.

Table B.4 Number of students in public elementary and secondary schools, by type of school, from non-NAEP sources: By urban district, school year 2000–01

		Type of school[1]			
	Total number of students[2]	Regular	Special education	Vocational education	Other and alternative
National	47,094,888	46,228,904	174,577	199,669	491,738
Central city	13,522,154	13,152,151	73,387	75,953	220,663
Atlanta	58,230	56,896	†	†	1,334
Chicago	435,261	431,553	3,708	†	†
District of Columbia	68,925	65,285	2,772	†	868
Houston	208,462	204,042	19	906	3,495
Los Angeles	721,346	704,932	4,480	†	11,934
New York City	1,066,945	1,009,319	1,688	25,409	30,529

† No students reported in membership for this type of school.
[1] Type of school is a mutually exclusive category on the Common Core of Data. There are cases in which special education, vocational education, and alternative programs reside in other types of schools.
[2] Student distribution by type of school is based on membership in the schools of the school district. Counts may vary from those in table B.1.
NOTE: Types of schools are defined in the following way on the Common Core of Data: Regular school — A public elementary/secondary school that does not focus primarily on vocational, special, or alternative education. NAEP is conducted only in regular schools. Special education school — A public elementary/secondary school that (a) focuses primarily on special education, including instruction for any of the following: hard of hearing, deaf, speech-impaired, health-impaired, orthopedically impaired, mentally retarded, seriously emotionally disturbed, multi-handicapped, visually handicapped, deaf and blind; and (b) adapts curriculum, materials, or instruction for students served. Vocational education school — A public elementary/secondary school that focuses primarily on vocational education, and provides education and training in one or more semi-skilled or technical operations. Alternative education school — A public elementary/secondary school that (a) addresses the needs of students that typically cannot be met in a regular school; (b) provides nontraditional education; (c) serves as an adjunct to a regular school; and (d) falls outside of the categories of regular, special education, or vocational education.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.5 **Percentage of minority students, and number of public elementary and secondary schools within specified ranges of minority student percentages, from non-NAEP sources: By urban district, school year 2000–01**

	Percentage of minority students enrolled	Number of schools with students [2]	Number of schools with a minority student [1] percentage of				
			0–20 percent	21–40 percent	41–60 percent	61–80 percent	81–100 percent
National	39.0	89,110	43,827	14,023	9,866	7,142	14,252
Central city	63.0	21,513	3,534	3,165	3,370	3,088	8,356
Atlanta	93.2	98	2	2	2	5	87
Chicago	90.4	596	1	20	26	51	498
District of Columbia	95.5	165	0	5	3	5	152
Houston	90.0	289	1	4	17	20	247
Los Angeles	90.1	659	4	16	50	88	501
New York City	84.7	1,207	29	39	114	139	886

[1] Minority students, in this table, includes all race/ethnicity categories except White, non-Hispanic.
[2] Includes only schools for which student membership by race/ethnicity was reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.6 **Number of public elementary and secondary schools, by instructional level, from non-NAEP sources: By urban district, school year 2000–01**

	Number of schools with students[2]	Number of schools by instructional level[1]			
		Primary	Middle	High	Other
National	90,711	52,285	15,584	17,280	5,562
Central city	21,900	13,962	3,318	3,081	1,539
Atlanta	98	70	15	12	1
Chicago	596	472	21	72	31
District of Columbia	165	112	10	16	27
Houston	289	197	43	35	14
Los Angeles	659	456	75	105	23
New York City	1,207	692	210	184	121

[1] Instructional level is based on the lowest and highest grade in a school: Primary schools begin between prekindergarten and grade 3 and may go as high as grade 8. Middle schools have grade spans ranging from as low as grade 4 to as high as grade 9. High schools start at grade 7 or higher and must extend through grade 12. Other schools include all other grade combinations, including prekindergarten, kindergarten, or 1–12, and ungraded schools.
[2] Includes only schools for which student membership was reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.7 Median pupil/teacher ratios in public elementary and secondary schools, by instructional level, from non-NAEP sources: By urban district, school year 2000–01

	Median pupil/teacher ratio[1]				
		By school instructional level[2]			
	Overall	Primary	Middle	High	Other
National	15.9	16.3	15.7	15.1	12.1
Central city	16.6	16.7	16.0	17.1	11.2
Atlanta	14.2	13.7	16.2	16.2	13.2
Chicago	18.4	18.9	17.3	15.5	11.9
District of Columbia	13.5	13.7	13.6	13.6	8.7
Houston	18.5	18.7	18.7	18.0	3.3
Los Angeles	19.6	19.2	23.6	22.7	10.4
New York City	16.1	15.7	16.7	18.7	6.3

[1] Includes only those schools whose student membership was greater than zero.
[2] Instructional level is based on the lowest and highest grade offered in a school. Primary schools begin between prekindergarten and grade 3 and may go as high as grade 8. Middle schools have grade spans ranging from as low as grade 4 to as high as grade 9. High schools start at grade 7 or higher and must extend through grade 12. Other schools include all other grade combinations, including prekindergarten, kindergarten, or 1–12, and ungraded schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.8 Percentage of staff in public elementary and secondary schools, by type of staff, from non-NAEP sources: By urban district, school year 2000–01

		Percent of full-time equivalent staff						
	Total staff	Teachers	Instructional support	Guidance counselors	Library media staff	District administrators	School administrators	Other staff
National	5,397,788	52.6	11.4	1.8	1.5	1.0	2.5	29.1
Central city	1,513,730	53.4	11.2	1.7	1.3	0.8	2.5	29.1
Atlanta	7,552	52.3	12.7	2.0	1.3	0.5	2.5	28.8
Chicago	28,687 [1]	83.4	2.9	2.9	1.7	1.7	3.1	4.4
District of Columbia	10,808	46.7	10.8	1.9	1.3	0.1	2.5	36.8
Houston	24,820	45.1	10.2	1.2	0.9	0.1	2.3	40.2
Los Angeles	66,598	52.8	14.5	1.4	0.1	0.6	1.9	28.7
New York City	100,198	65.1	0.8	1.9	0.7	0.7	2.9	27.9

[1] The non-teaching staff categories may be underrepresented.
NOTE: Percentage may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.9 **Percentage of reported students eligible for free/reduced-price school lunch and percentage of students in each racial/ethnic category, from non-NAEP sources: By urban district, school year 2000–01**

		Students eligible for free/reduced-price lunch [1]		Racial/ethnic composition of district as a percentage of students enrolled				
	Number of schools with students	Percentage of students	Percentage of schools reporting	White	Black	Hispanic	Asian/ Pacific Islander	American Indian/ Alaska Native
National	90,711	38.6	83.8	61.0	17.0	16.6	4.2	1.2
Central city	21,900	53.6	82.9	36.8	29.2	27.7	5.5	0.9
Atlanta	98	76.4	100.0	6.8	89.5	2.8	0.9	0.1
Chicago	596	—	0.0	9.6	52.0	34.9	3.3	0.2
District of Columbia	165	70.0	98.8	4.5	84.6	9.2	1.6	0.1
Houston	289	70.7	99.7	10.0	32.1	55.0	2.9	0.1
Los Angeles	659	73.5	99.8	9.9	12.8	70.8	6.3	0.3
New York City	1,207	71.9	100.0	15.3	34.9	37.8	11.7	0.3

— Not available.
[1] These percentages should be interpreted with caution; jurisdictions may not have reported students eligible for reduced-price meals, and a number of jurisdictions reported participation instead of eligibility data, which may not be strictly comparable. Percentages are based on those schools that reported.
[2] Includes only schools for which student membership was reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Appendix C

Overview of Procedures Used for the NAEP 2002 Trial Urban District Reading Assessment

This appendix provides an overview of the NAEP 2002 reading assessment's primary components—framework, development, administration, scoring, and analysis. A more extensive review of the procedures and methods used in the national and state reading assessments can be found in the assessment procedures sections of the NAEP web site (*http://nces.ed.gov/nationsreportcard*).

The NAEP 2002 Reading Assessment

The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and test specifications. The design of the NAEP 2002 reading assessment follows the guidelines first provided in the framework developed for the 1992 assessment.[1] The framework underlying the 1992, 1994, 1998, 2000 (fourth grade only), and 2002 reading assessments reflects the expert opinions of educators and researchers about reading. Its purpose is to present an overview of the most essential outcomes of students' reading education. The development of this framework and the specifications that guided the development of the assessment involved the critical input of hundreds of individuals across the country, including representatives of national education organizations, teachers, parents, policymakers, business leaders, and the interested general public. The framework development

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

process was managed by the Council of Chief State School Officers (CCSSO) for NAGB.

The framework sets forth a broad definition of "reading literacy"—developing a general understanding of a text, thinking about a text in different ways, and using a variety of text types for different purposes. In addition, the framework views reading as an interactive and constructive process involving the reader, the text, and the context of the reading experience. For example, readers may read stories to enjoy and appreciate the human experience, study science texts to form new hypotheses about knowledge, or use maps to gain information about specific places. NAEP reflects current definitions of literacy by differentiating among three contexts for reading and four aspects of reading. Contexts for reading and aspects of reading make up the foundation of the NAEP reading assessment.

The "contexts for reading" dimension of the NAEP reading framework provides guidance for the types of texts to be included in the assessment. Although many commonalities exist among the different reading texts, they do lead to real differences in what readers do. For example, when *reading for literary experience*, readers make complex, abstract summaries and identify major themes. They describe the interactions of various literary elements (e.g., setting, plot, characters, and theme). When *reading for information*, readers critically judge the form and content of the text and explain their judgments. They also look for specific pieces of information. When *reading to perform a task*, readers search quickly for specific pieces of information.

The "aspects of reading" dimension of the NAEP reading framework provides guidance for the types of comprehension questions to be included in the assessment. The four aspects are 1) *forming a general understanding*, 2) *developing interpretation*, 3) *making reader/text connections*, and 4) *examining content and structure*. These four aspects represent different ways in which readers develop understanding of a text. In *forming a general understanding*, readers must consider the text as a whole and provide a global understanding of it. As readers engage in *developing interpretation*, they must extend initial impressions in order to develop a more complete understanding of what was read. This involves linking information across parts of a text or focusing on specific information. When *making reader/text connections*, the reader must connect information in the text with knowledge and experience. This might include applying ideas in the text to the real world. Finally, *examining content and structure* requires critically evaluating, comparing and contrasting, and understanding the effect of different text features and authorial devices.

Figure C.1 demonstrates the relationship between these reading contexts and aspects of reading in the NAEP reading assessment. Included in the figure are sample questions that illustrate how each aspect of reading is assessed within each reading context. (Note that *reading to perform a task* is not assessed at grade 4.)

Figure C.1 Sample NAEP questions, by aspects of reading and contexts for reading specified in the reading framework

	Aspect of Reading			
Context for Reading	Forming a general understanding	Developing interpretation	Making reader/text connections	Examining content and structure
Reading for literary experience	*What is the story/plot about?*	*How did this character change from the beginning to the end of the story?*	*What other character that you have read about had a similar problem?*	*What is the mood of this story and how does the author use language to achieve it?*
Reading for information	*What point is the author making about this topic?*	*What caused this change?*	*What other event in history or recent news is similar to this one?*	*Is this author biased? Support your answer with information about this article.*
Reading to perform a task	*What time can you get a nonstop flight to X?*	*What must you do before step 3?*	*Describe a situation in which you would omit step 5.*	*Is the information in this brochure easy to use?*

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

The assessment framework specifies not only the particular dimensions of reading literacy to be measured, but also the percentage of assessment questions that should be devoted to each. The target percentage distribution for contexts of reading and aspects of reading as specified in the framework, along with the actual percentage distribution in the assessment, are presented in tables C.1 and C.2.

The actual content of the assessment varied from the targeted distribution, with reading for literary experience falling below the target proportions and reading for information falling above the target proportions specified in the framework. The reading instrument development panel overseeing the development of the assessment recognized this variance but felt strongly that assessment questions must be sensitive to the unique elements of the authentic reading materials being used. Thus, the distribution of question classifications will vary across reading passages and reading purposes.

Table C.1 **Target and actual percentage distribution of questions, by context for reading, grades 4 and 8: 2002**

		Context for Reading		
		Reading for literary experience	Reading for information	Reading to perform a task
Grade 4				
	Target	55	45	†
	Actual	50	50	†
Grade 8				
	Target	40	40	20
	Actual	27	43	30

† Reading to perform a task was not assessed at grade 4.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table C.2 **Target and actual percentage distribution of questions, by aspect of reading, grades 4 and 8: 2002**

		Aspect of Reading		
		Forming a general understanding/ Developing interpretation	Making reader/text connections	Examining content and structure
Grade 4				
	Target	60	15	25
	Actual	59	18	24
Grade 8				
	Target	55	15	30
	Actual	54	18	28

NOTE: Actual percentages are based on the classifications agreed upon by NAEP's Instrument Development Panel. It is recognized that making discrete classifications for these categories is difficult and that independent efforts to classify NAEP questions have led to different results.
Percentages may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

The Assessment Design

Each student who participated in the reading assessment received a booklet containing three or four sections: a set of general background questions, a set of subject-specific background questions, and one or two sets of questions assessing students' comprehension of a text or texts. The sets of questions assessing students' comprehension are referred to as "blocks." Each block contains one or more reading passages and a set of comprehension questions. At grade 8, students were given either two 25-minute blocks or one 50-minute block. At grade 4, however, only 25-minute blocks were used.

The blocks contain a combination of multiple-choice and constructed-response questions. Multiple-choice questions require students to select the best answer from a set of four options. Constructed-response questions require students to provide their own written response to an open-ended question. Short constructed-response questions may require a response of only a sentence or two for the answer to be considered complete. Extended constructed-response questions, however, may require a response of a paragraph or more for the answer to receive full credit. Each constructed-response question has its own unique scoring guide that is used by trained scorers to rate students' responses. (See the "Data Collection and Scoring" section of this appendix.)

The grade 4 assessment consisted of eight 25-minute blocks: four blocks of "literary" texts and questions and four blocks of "informative" texts and questions. Each block contained at least one passage corresponding to one of the contexts for reading and 9–12 multiple-choice and constructed-response questions. In each block, one of the constructed-response questions required an extended response. As a whole, the 2002 fourth-grade assessment consisted of 49 multiple-choice questions, 45 short constructed-response questions, and 8 extended constructed-response questions.

The grade 8 assessment consisted of nine 25-minute blocks (three literary, three informative, and three task-related) and one 50-minute block (informative). Each block contained at least one passage corresponding to one of the contexts for reading and 8 to 13 multiple-choice and constructed-response questions. Each block contained at least one extended constructed-response question. As a whole, the eighth-grade assessment consisted of 58 multiple-choice questions, 68 short constructed-response questions, and 15 extended constructed-response questions.

The assessment design allowed maximum coverage of reading abilities at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of items in which representative samples of students took various portions of the entire pool of assessment questions. Individual students are required to take only a small portion, but the aggregate results across the entire assessment allow for broad reporting of reading abilities for the targeted population.

In addition to matrix sampling, the assessment design used a procedure for distributing blocks across booklets that controlled for position and context effects. Students received different blocks of passages and comprehension questions in their booklets according to a procedure called "partially balanced incomplete block (pBIB) spiraling." This procedure assigned blocks of questions in a manner that balanced the positioning of blocks across booklets and balanced the pairing of blocks within booklets according to context for reading. Blocks were balanced within each context for reading and were partially balanced across contexts for reading. The spiraling aspect of this procedure cycled the booklets for administration so that, typically, only a few students in any assessment session received the same booklet.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment—a teacher questionnaire, a school questionnaire, and a questionnaire regarding students with disabilities/limited English proficient students (SD/LEP). The teacher questionnaire was administered to teachers of fourth- and eighth-grade students participating in the assessment and included four sections. The first section focused on the teacher's background, the second section on instruction, the third section on professional development, and the fourth section on standards and assessment.

The school questionnaire was given to the principal or another administrator in each participating school and included questions related to school policies, programs, and the composition and background of the student body.

The SD/LEP 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 having an Individualized Education Program (IEP) or equivalent plan and/or being limited English proficient (LEP). An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire asked about the student and the special programs in which he or she participated.

NAEP Samples

National Sample

The national results, presented in *The Nation's Report Card: Reading 2002,* are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. At grades 4 and 8, the national sample in 2002 was a subset of the combined sample of students assessed in each participating state, plus an additional sample from the states that did not participate in the state assessment and a private school sample. In accordance with the NAEP legislation, the program uses a random selection process in order to obtain a representative sample of students for reporting national and state or jurisdiction results.

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 applied to the national and state samples account for disproportionate representation due to the oversampling of students who attend nonpublic schools and schools with high concentrations of Black and/or Hispanic students. Among other uses, sampling weights also account for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.[2] Appropriate sampling weights were applied to the trial urban district samples.

Testing accommodations (e.g., extended time, small group testing) were permitted for special-needs students selected to participate in the NAEP reading assessments. NAEP inclusion rules were applied, and accommodations were offered when a student had an Individualized Education Program (IEP) because of a disability, was protected under Section 504 of the Rehabilitation Act of 1973 because of disability, and/or was identified as being a limited English proficient student (LEP); all other students were asked to participate in the assessment under standard conditions.[3]

[2] Additional details regarding the design and structure of the national and state samples will be found in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard.*

[3] Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

Standards for Sample Participation and Reporting of Results

In carrying out the 2002 state and Trial Urban District Assessments, 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. 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. The NCES guidelines used to report results in the state assessments, and the guidelines for notation when there is some risk of nonresponse bias in the reported results, are presented in this section. These guidelines also applied to the Trial Urban District Assessments.

Guideline 1

The publication of NAEP results

The conditions that will result in the publication of a jurisdiction's results are presented below.

Guideline 1 - Publication of Public School Results

A jurisdiction will have its public school results published in the NAEP 2002 reading report card (or in other reports that include all state-level results) if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent. Similarly, a jurisdiction will receive a separate NAEP State Report if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent.

Discussion: If a jurisdiction's public school participation rate for the initial sample of schools is below 70 percent, there is a substantial possibility that bias will be introduced into the assessment results. This possibility remains even after making statistical adjustments to compensate for school nonparticipation. There remains the likelihood that, in aggregate, the substitute schools are sufficiently dissimilar from the originals they are replacing and represent too great a proportion of the population to discount such a difference. Similarly, the assumptions underlying the use of statistical adjustments to compensate for nonparticipation are likely to be significantly violated if the initial response rate falls below the 70 percent level. Guideline 1 takes this into consideration. This guideline is congruent with current NAGB policy, which requires that data for jurisdictions that do not have a 70 percent before-substitution participation rate be reported "in a different format," and with the Education Information Advisory Committee (EIAC) resolution, which calls for data from such jurisdictions not to be published.

The following guidelines concerning school and student participation rates in the NAEP state assessment program were established to address four significant ways in which nonresponse bias could be introduced into the jurisdiction sample estimates. The four significant ways are overall school nonresponse, strata-specific school nonresponse, overall student nonresponse, and strata-specific student nonresponse. Presented on the following pages are the conditions that will result in a jurisdiction's receiving a notation in the 2002 reports. Note that in order for a jurisdiction's results to be published with no notations, that jurisdiction must satisfy all guidelines.

Guideline 2

Reporting school and student participation rates with possible bias due to school nonresponse

Guideline 2 - Notation for Overall Public School Participation Rate

A jurisdiction that meets Guideline 1 will receive a notation if its weighted participation rate for the initial sample of public schools was below 85 percent and the weighted public school participation rate after substitution was below 90 percent.

Discussion: For jurisdictions that did not use substitute schools, the participation rates are based on participating schools from the original sample. In these situations, the NCES standards specify weighted school participation rates of at least 85 percent to guard against potential bias due to school nonresponse. Thus the first part of these guidelines, referring to the weighted school participation rate for the initial sample of schools, is in direct accordance with NCES standards.

To help ensure adequate sample representation for each jurisdiction participating in the NAEP 2002 state assessments, NAEP provided substitutes for nonparticipating public schools. For jurisdictions that used substitute schools, the assessment results will be based on the student data from all schools participating from both the original sample and the list of substitutes (unless both an initial school and its substitute eventually participated, in which case only the data from the initial school will be used).

The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that decide not to participate in the assessment. However, considerable technical consideration was 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 bias due to the nonparticipation of initially selected schools. Thus, for the weighted school participation rates including substitute schools, the guidelines were set at 90 percent.

If a jurisdiction meets either standard (i.e., 85 percent or higher prior to substitution or 90 percent or higher after substitution), there will be no notation for the relevant overall school participation rate.

Guideline 3

Important segments of the jurisdiction's student population that must be adequately represented to avoid possible nonresponse bias

Guideline 3 - Notation for Strata-Specific Public School Participation Rates

A jurisdiction that is not already receiving a notation under Guideline 2 will receive a notation if the sample of public schools included a class of schools with similar characteristics that had a weighted participation rate (after substitution) of below 80 percent, and from which the nonparticipating schools together accounted for more than 5 percent of the jurisdiction's total weighted sample of public schools. The classes of schools from each of which a jurisdiction needed minimum school participation levels were determined by degree of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, if some important segment of the jurisdiction's population is not adequately represented, it is of concern, regardless of the overall participation rate.

If nonparticipating schools are concentrated within a particular class of schools, the potential for substantial bias remains, even if the overall level of school participation appears to be satisfactory. Nonresponse adjustment cells for public schools have been formed within each jurisdiction, and the schools within each cell are similar with respect to minority enrollment, degree of urbanization, and/or median household income, as appropriate for each jurisdiction.

If the weighted response rate, after substitution, for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the sampled schools are nonparticipants from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific school response rates.

Guideline 4

Possible student nonresponse bias

Guideline 4 - Notation for Overall Student Participation Rate in Public Schools

A jurisdiction that meets Guideline 1 will receive a notation if the weighted student response rate within participating public schools was below 85 percent.

Discussion: This guideline follows the NCES standard of 85 percent for overall student participation rates. The weighted student participation rate is based on all eligible students from initially selected or substitute schools who participated in the assessment in either an initial session or a make-up session. If the rate falls below 85 percent, the potential for bias due to students' nonresponse is too great.

Guideline 5

Possible nonresponse bias from inadequately represented strata

Guideline 5 - Notation for Strata-Specific Student Participation Rates in Public Schools

A jurisdiction that is not already receiving a notation under Guideline 4 will receive a notation if the sampled students within participating public 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 5 percent of the jurisdiction's weighted assessable public school student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the student, whether or not the student was classified as a student with a disability (SD) or limited English proficient (LEP), and the type of assessment session, as well as school level of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: This guideline addresses the fact that if nonparticipating students are concentrated within a particular class of students, the potential for substantial bias remains, even if the overall student participation level appears to be satisfactory. Student nonresponse adjustment cells have been formed using the school-level nonresponse adjustment cells, together with the student's age and the nature of the assessment session.

If the weighted response rate for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the invited students who do not participate in the assessment are from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific student response rates.

In the 2002 Trial Urban District Assessment, New York City did not meet the initial public-school participation rate of 70 percent at eighth grade. Consequently the performance results for this grade were not reported.

At grade 4, New York City also did not meet the second guideline (i.e., the weighted participation rate for the initial sample of schools was below 85 percent and the weighted school participation rate after substitution was below 90 percent). Results for New York City at grade 4 are shown with a notation indicating possible bias related to nonresponse.

Trial Urban District Samples

Sampling for the Trial Urban District Assessment was modeled on NAEP's state sampling procedures. However, school substitution was not an option in the Trial Urban District Assessment. Tables C.3 and C.4 provide a summary of the 2002 urban district school and student participation rates for the reading assessment sample. The first rate in each table is the weighted percentage of schools participating in the assessment. 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 selected school that participated in the assessment. The denominator is the sum of the number of students represented by each of the selected schools that had eligible students enrolled.

Also presented in tables C.3 and C.4 are weighted student participation rates. The numerator of this rate is the estimated number of students that each student represents across all students *assessed* (in either an initial session or a makeup session). The denominator of this rate is the estimated number of students represented in the sample, across all *eligible* sampled students in participating schools. The number of students that each student represents is mainly determined by the probability that a student is included in the sample, with necessary adjustments made for other factors. The overall participation rates take into account the weighted percentage of school participation and the weighted percentage of student participation after makeup sessions.

For the grade 8 sample, New York City's school and student response rates did not meet NCES standards. Consequently the performance results for this grade were not reported.

Table C.3 Weighted school and student participation rates, grade 4: By urban district, 2002

Grade 4	Number of schools	Number of students	School rate	Student rate	Overall rate
Atlanta	49	1,509	98	93	92
Chicago	76	2,100	95	92	87
District of Columbia	117	2,810	100	90	90
Houston	49	1,326	98	95	93
Los Angeles	76	2,017	100	93	93
New York City ‡	38	947	76	89	67

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table C.4 Weighted school and student participation rates, grade 8: By urban district, 2002

Grade 8	Number of schools	Number of students	School rate	Student rate	Overall rate
Atlanta	15	1,281	100	91	91
Chicago	69	1,562	94	92	86
District of Columbia	36	1,795	100	85	85
Houston	34	1,110	96	89	86
Los Angeles	66	1,778	97	90	87
New York City ‡	31	692	63	81	51

‡ Indicates that the district did not meet the guideline for 70 percent school participation in 2002.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Results from the 2002 reading assessments are reported (on a trial basis) here on district-level samples of fourth- and eighth-grade students in the large urban school districts that participated in the Trial Urban District Assessment and in the District of Columbia. The samples of students in the urban school districts represent augmentations of students who would normally be selected as part of state samples. These samples allow reliable subgroup reporting in these districts. All students at "lower" geographical levels are assumed to be part of "higher-level" samples. For example, Houston is one of the urban districts included in the Trial Urban District Assessment. Data from students tested in the Houston sample are used to report results for Houston, and also contribute to the Texas and the national estimates.

Tables C.5 and C.6 display the target student and school sample sizes planned for the Trial Urban District Assessment. The first column contains the planned number of schools for each district. The second column contains the number of schools that would have been sampled by NAEP in each district had there been no Trial Urban District Assessment. The last column shows the planned sample size.

Table C.5 Number of schools and students planned for the Trial Urban District Assessment, grade 4: By urban district, 2002

Grade 4	Number of schools for trial assessment	Number of schools normally sampled in NAEP	Number of students
Atlanta	51	6	1250
Chicago	81	24	1900
Houston	51	5	1250
Los Angeles	76	12	1900
New York City ‡	52	41	1250

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table C.6 Number of schools and students planned for the Trial Urban District Assessment, grade 8: By urban district, 2002

Grade 8	Number of schools for trial assessment	Number of schools normally sampled in NAEP	Number of students
Atlanta	17	8	1250
Chicago	79	26	1900
Houston	39	6	1250
Los Angeles	68	13	1900
New York City ‡	52	37	1250

‡ Indicates that the district did not meet the guideline for 70 percent school participation in 2002.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Note that the sample sizes vary among districts. The study was designed to allow the examination of the quality of data that resulted from samples of different sizes. Therefore, larger samples were selected in two of the districts (Chicago and Los Angeles), and smaller samples in the remaining three.

Data Collection and Scoring

The 2002 reading assessment was conducted from January to March 2002. Data collection for the 2002 assessment at the national, state and district levels was conducted by trained field staff from Westat.

Materials from the 2002 assessment were shipped to Pearson, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by Educational Testing Service (ETS). Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four-level guides, and almost all of the short constructed-response questions were rated according to three-level guides that permitted partial credit. Other short constructed-response questions were scored as either acceptable or unacceptable.

For the 2002 reading assessment, 4,023,861 constructed responses were scored. This number includes rescoring to monitor interrater reliability. The within-year average percentage of exact agreement for the 2002 national reliability sample was 92 percent at fourth grade and 91 percent at eighth grade.

Data Analysis and IRT Scaling

Subsequent to the professional scoring, all information was transcribed into the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information was compiled in the database, the data were weighted according to the population structure. The weighting for the national, state and trial urban district samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.[4]

Analyses 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. In analysis, omitted responses to multiple-choice items were scored as fractionally correct.[5] For constructed-response items, omitted responses were placed into the lowest score category. 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 NAEP 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 ques-

[4] Weighting procedures are described more fully in the "Weighting and Variance Estimation" section later in this document. Additional information about the use of weighting procedures will be found in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard.*

[5] Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

tion. However, for reading blocks that ended with extended constructed-response questions, the standard practice could result in extremely large drops in the proportion of students attempting some of the final questions. Therefore, for blocks ending with an extended constructed-response 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 reading scale scores for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. NAEP used IRT analysis to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity.

The results for 1992, 1994, 1998, 2000 and 2002 are presented on the NAEP reading scales in the *Nation's Report Card: Reading 2002*. In 1992, a scale ranging from 0 to 500 was created to report performance for each reading purpose — literary and information at grade 4, and literary, information, and task at grades 8 and 12. The scales summarize student performance across all three types of questions in the assessment (multiple-choice, short constructed-response, and extended constructed-response). Results from subsequent reading assessments (1994, 1998, 2000, and 2002) are reported on these scales.

Each reading scale was initially based on the distribution of student performance across all three grades in the 1992 national assessment (grades 4, 8, and 12). In that year, the scales had an average of 250 and a standard deviation of 50. In addition, a composite scale was created as an overall measure of students' reading performance. This composite scale is a weighted average of the three separate scales for the three reading purposes at grade 8 and two reading purposes at grade 4. The weight for each reading purpose is proportional to the relative importance assigned to the reading purpose by the specifications developed through the consensus planning process and given in the framework.

In producing the reading scales, three distinct IRT models were used. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as acceptable or unacceptable were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level scoring guide, as well as extended constructed-response questions rated on a four-level scoring guide, were scaled using a Generalized Partial-Credit (GPC) model.[6] 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.[7]

[6] Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

[7] More detailed information regarding the IRT analyses used in NAEP will be found in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard*.

The reading scale is composed of three types of questions: multiple-choice, short constructed-response (scored either dichotomously or allowing for partial credit), and extended constructed-response (scored according to a partial-credit model). Unfortunately, the question of how much information different types of questions contribute to the reading scale has no simple answer. The information provided by a given question is determined by the IRT model used to scale the question. It is a function of the item parameters and varies by level of reading proficiency.[8] Thus, the answer to the query "How much information do the different types of questions provide?" will differ for each level of reading performance. When considering the composite reading scale, the answer is even more complicated. The reading data are scaled separately by the two purposes for reading (reading for information and reading for literary experience) for grade 4, and the three purposes for reading (reading for information, reading for literary experience, and reading to perform a task) for grade 8, resulting in two or three separate subscales at each grade. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when the item parameters are estimated together. Because the composite scale is based on two or three separate estimation runs, there is no direct way to compare the information provided by the questions on the composite scale.

Because of the assessment booklet spiraling design used by NAEP, students do not receive enough questions about a specific topic to provide reliable information about individual performance. (For more information on the assessment booklet, see "The Assessment Design" section presented earlier in this appendix.) Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale-score level. However, it is NAEP's goal to estimate these population characteristics. As discussed by Mislevy and Sheehan (1987), NAEP's objectives can be achieved with methodologies that produce estimates of the population-level parameters directly, without the intermediary computation of estimates of individuals.[9] This is accomplished using marginal estimation scaling model techniques for latent variables. Under the assumptions of the scaling models, these population estimates will be consistent in the sense that the estimates approach the model-based population values as the sample size increases. This would not be the case for population estimates obtained by aggregating optimal estimates of individual performance.[10]

Weighting and Variance Estimation

A complex sampling design was used to select the students who were assessed. The properties of a sample selected through such a design may be very different from

[8] 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.

[9] Mislevy, R. J., and Sheehan, K. M. (1987). Marginal Estimation Procedures. In A. E. Beaton (Ed.) *Implementing the New Design: The NAEP 1983-1984 Technical Report* (Rep. No. 15-TR-20), pp. 293–260. Princeton, NJ: Educational Testing Service.

[10] 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.

those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data were estimated using sampling weights. These weights included adjustments for school and student nonresponse.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. 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 portion of the cognitive domain of interest. 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 context of reading, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of groups and subgroups of students. The estimate of the variance of the students' posterior scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.[11]

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 estimation of standard errors may be quite large. Additional details concerning procedures for identifying such standard errors will be found in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard*.

[11] For further details, see Johnson, E. G., and Rust, K. F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics*, *17*(2), 175–190.

The reader is reminded that, as with 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 background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors 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.

Drawing Inferences from the Results

The reported statistics are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty related to the fact that they cannot ask all questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the estimated standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account, and observed similarities or differences should not be relied on solely. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of those statistics and the magnitude of the difference among the averages or percentages.

For the data presented in this report, all the estimates have corresponding estimated standard errors. For example, table C.7 shows the average scale score for the NAEP 2002 Trial Urban District Assessment and percentages of students at or above achievement levels by gender for grade 4. Also, table C.8 shows the scores and standard errors for the 25th, 50th, and 75th percentiles at grades 4 and 8. Estimated standard errors appear in parentheses next to each estimated scale score or percentage. For the estimated standard errors corresponding to other data in this report, the reader can consult the NCES web site at *http://nces.ed.gov/nationsreportcard/naepdata*.

Table C.7 Average reading scale scores, percentage of students at or above each achievement level, and estimated standard errors, by gender, grade 4 public schools: By urban district, 2002

Grade 4	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male						
Nation (Public)	51 (0.3)	214 (0.5)	41 (0.6)	59 (0.6)	26 (0.5)	5 (0.2)
Central city (Public) [1]	50 (0.4)	204 (0.7) **	53 (0.8) **	47 (0.8) **	19 (0.6) **	3 (0.3) **
Atlanta	47 (1.3)	191 (2.5) *,**	69 (2.6) *,**	31 (2.6) *,**	11 (2.3) *,**	2 (1.0)
Chicago	50 (1.3)	189 (2.0) *,**	70 (2.2) *,**	30 (2.2) *,**	9 (1.1) *,**	1 (0.4) *,**
District of Columbia	49 (1.2)	185 (1.2) *,**	74 (1.4) *,**	26 (1.4) *,**	8 (0.9) *,**	1 (0.4) *,**
Houston	51 (1.5)	204 (3.3) **	55 (3.9) **	45 (3.9) **	16 (3.4) **	3 (1.2)
Los Angeles	51 (1.2)	188 (2.2) *,**	70 (1.9) *,**	30 (1.9) *,**	10 (1.4) *,**	1 (0.4) *,**
New York City ‡	50 (1.5)	199 (2.7) **	61 (3.9) *,**	39 (3.9) *,**	14 (2.7) **	3 (0.9) **
Female						
Nation (Public)	49 (0.3)	220 (0.5)	35 (0.6)	65 (0.6)	33 (0.6)	8 (0.3)
Central city (Public) [1]	50 (0.4)	211 (0.7) **	45 (0.8) **	55 (0.8) **	24 (0.9) **	5 (0.4) **
Atlanta	53 (1.3)	200 (1.7) *,**	60 (2.4) *,**	40 (2.4) *,**	13 (1.4) *,**	4 (0.8) **
Chicago	50 (1.3)	198 (2.0) *,**	62 (2.3) *,**	38 (2.3) *,**	12 (1.4) *,**	2 (0.7) *,**
District of Columbia	51 (1.2)	196 (1.2) *,**	64 (1.5) *,**	36 (1.5) *,**	11 (1.0) *,**	2 (0.4) *,**
Houston	49 (1.5)	208 (2.7) **	50 (3.1) **	50 (3.1) **	19 (3.0) **	3 (1.5) **
Los Angeles	49 (1.2)	194 (2.2) *,**	64 (2.9) *,**	36 (2.9) *,**	12 (1.7) *,**	2 (0.5) *,**
New York City ‡	50 (1.5)	213 (3.0) **	45 (3.5) **	55 (3.5) **	23 (3.5) **	7 (2.5)

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
NOTE: Standard errors of the estimated percentages and scale scores appear in parentheses. Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Table C.8 Selected reading percentiles and estimated standard errors, grades 4 and 8 public schools: By urban district, 2002

	Scale score distribution		
	25th percentile	50th percentile	75th percentile
Grade 4			
Nation (Public)	194 (0.6)	219 (0.4)	242 (0.5)
Central city (Public) [1]	183 (0.8) **	209 (0.7) **	234 (0.7) **
Atlanta	171 (2.3) **,*	194 (1.4) **,*	219 (2.1) **,*
Chicago	170 (2.3) **,*	194 (2.1) **,*	217 (1.8) **,*
District of Columbia	167 (1.3) **,*	191 (0.9) **,*	215 (1.3) **,*
Houston	183 (2.9) **	206 (2.8) **	229 (3.9) **
Los Angeles	165 (2.5) **,*	190 (1.3) **,*	217 (2.2) **,*
New York City ‡	182 (2.3) **	206 (2.6) **	230 (3.6) **
Grade 8			
Nation (Public)	242 (0.5)	265 (0.6)	286 (0.5)
Central city (Public) [2]	232 (0.9) **	256 (0.7) **	278 (0.8) **
Atlanta	214 (2.3) **,*	236 (2.6) **,*	259 (1.4) **,*
Chicago	231 (1.8) **	251 (2.2) **,*	270 (2.6) **,*
District of Columbia	219 (3.3) **,*	241 (1.3) **,*	262 (1.7) **,*
Houston	226 (2.9) **	251 (2.2) **,*	273 (1.4) **,*
Los Angeles	213 (2.1) **,*	238 (1.4) **,*	261 (1.4) **,*

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 65 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price school lunch.
[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 47 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price school lunch.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Reading Assessment.

Using confidence intervals based on the standard errors provides a way to take into account the uncertainty associated with sample estimates and to make inferences about the population averages and percentages in a manner that reflects that uncertainty. An estimated sample average scale score plus or minus 1.96 standard errors approximates a 95-percent confidence interval for the corresponding population quantity. This statement means that one can conclude with an approximately 95 percent level of confidence that the average performance of the entire population of interest (e.g., all fourth-grade students in public schools) is within plus or minus 1.96 standard errors of the sample average.

For example, suppose that the average reading scale score of the students in a particular group was 256 with an estimated standard error of 1.2. An approximately 95 percent confidence interval for the population quantity would be as follows:

$$\text{Average} \pm 1.96 \text{ standard errors}$$

$$256 \pm 1.96 \times 1.2$$

$$256 \pm 2.4$$

$$(253.6, 258.4)$$

Thus, one can conclude with a 95 percent level of confidence that the average scale score for the entire population of students in that group is between 253.6 and 258.4. It should be noted that this example and the examples in the following sections are illustrative. More precise estimates carried out to one or more decimal places are used in the actual analyses.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. Extreme percentages should be interpreted with caution. Adding or subtracting the standard errors associated with extreme percentages could cause the confidence interval to exceed 100 percent or fall below 0 percent, resulting in numbers that are not meaningful. A more complete discussion of extreme percentages will appear in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard*.

Analyzing Group Differences in Averages and Percentages

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 or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. 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 an 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.

Standard Error of the Difference =

$$SE_{A\text{-}B} = \sqrt{(SE_A^2 + SE_B^2)}$$

The standard error of the difference, just like the standard error for an individual group average or percentage, 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 plus or minus 1.96 standard errors of the difference represents an approximately 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 at the 0.05 level.

The following example of comparing groups addresses the problem of determining whether the average reading scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are as follows:

Group	Average Scale 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 estimated standard error of this difference is

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

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference.

$$2 \pm 1.96 \times 1.4$$

$$2 \pm 2.7$$

$$(-0.7, 4.7)$$

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

Conducting Multiple Tests

The procedures in the previous 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 calculated. However, there are times when 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.[12] One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure was used to control the certainty level.[13]

Unlike the other multiple comparison procedures that control 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, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.[14]

Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures. A detailed description of the FDR procedure will appear in the technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard*.

To illustrate how the FDR procedure is used, consider the comparisons of current and previous years' average reading scale scores for the five groups presented in table C.9. Note that the difference in average scale scores and the estimated standard error of the difference are calculated in a way comparable with that of the example in the previous section. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference.

Table C.9 Example of FDR comparisons of average scale scores for different groups of students

	Previous year		Current year		Previous year and current year			
	Average scale score	Standard error	Average scale score	Standard error	Difference in averages	Standard error of difference	Test statistic	Percent confidence[1]
Group 1	224	1.3	226	1.0	2.08	1.62	1.29	20
Group 2	187	1.7	193	1.7	6.31	2.36	2.68	1
Group 3	191	2.6	197	1.7	6.63	3.08	2.15	4
Group 4	229	4.4	232	4.6	3.24	6.35	0.51	62
Group 5	201	3.4	196	4.7	-5.51	5.81	-0.95	35

[1] The percent confidence is 2(1-F(x)) where F(x) is the cumulative distribution of the t-distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.

12 Miller, R. G. (1966). *Simultaneous Statistical Inference*. New York: John Wiley & Sons, Inc.

13 Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society,* Series B, no. 1, 289–300.

14 Williams, V. S. L., Jones, L. V., and Tukey, J. W. (1999). Controlling Error in Multiple Comparisons with Examples From State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics*, *24*(1), 42–69.

The difference in average scale scores and the estimated standard error of that difference can be used to find an approximately 95 percent confidence interval, as in the example in the previous section, or they can be used to identify a confidence percentage. In the example in the previous section, because an approximately 95 percent confidence interval was desired, the number 1.96 was used to multiply the estimated standard error of the difference to create the approximate confidence interval. In the current example, the confidence interval for the test statistics is identified from statistical tables. Instead of checking to see if zero is within the 95 percent confidence interval about the mean, the significance level from the statistical tables can be directly compared to 100 - 95 = 5 percent.

If the comparison of average scale scores across two years were made for only one of the five groups, there would be a significant difference between the average scale scores for the two years if the significance level were less than 5 percent. However, because we are interested in the difference in average scale scores across the two years for all five of the groups, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons. However, comparisons of average scale scores for each pair of years were treated separately, so the steps described in this example would be replicated for the comparison of other current and previous year average scale scores.

Using the FDR procedure to take into account that all comparisons are of interest to us, the percents of confidence in the example are ordered from largest to smallest: 62, 35, 20, 4, and 1. In the FDR procedure, 62 percent confidence for the group 4 comparison would be compared to 5 percent, 35 percent for the group 5 comparison would be compared to $0.05 \times (5-1)/5 = 0.04 = 4$ percent,[15] 20 percent for the group 1 comparison would be compared to $0.05 \times (5-2)/5 = 0.03 = 3$ percent, 4 percent for the group 3 comparison would be compared to $0.05 \times (5-3)/5 = 0.02 =$ 2 percent, and 1 percent for the group 2 comparison (actually slightly smaller than 1 prior to rounding) would be compared to $0.05 \times (5-4)/5 = 0.01 = 1$ percent. The procedure stops with the first contrast found to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference in the current year and previous years' average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for current and previous year are not significantly different from one another. In practice, a very small number of counterintuitive results occur when the FDR procedures are used to examine between-year differences in subgroup results by jurisdiction. In those cases, results were not included in this report. NCES is continuing to evaluate the use of FDR and multiple-comparison procedures for future reporting.

[15] The level of confidence times the number of comparisons minus one divided by the number of comparisons is $0.05 \times (5-1)/5 = 0.04 = 4$ percent.

NAEP Reporting Groups

Results are provided for groups of students defined by shared characteristics—gender, race/ethnicity, school's type of location, and eligibility for free/reduced-price school lunch. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).[16] The first-stage sampling units in the selection of Trial Urban District Assessment samples are schools. However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations are presented below. Note that not all of the reporting groups used for the national report card are included in this report on the urban districts.

Gender

Results are reported separately for males and females.

Race/Ethnicity

In all NAEP assessments, data about student race/ethnicity is collected from two sources: school records and student self-reports. Previously, NAEP has used student self-reported race as the primary race/ethnicity reporting variable. Starting in 2002, school-recorded race has become the race/ethnicity variable presented in NAEP reports. The mutually exclusive racial/ethnic categories are White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. When a school reports a student's race as "Other," that category is used. If the school record for race is missing for the student, the student's response to the race/ethnicity question is then used. If student data are missing (i.e. the student did not respond or gave multiple responses), then the student is coded to the "Other/missing" category. The combination of these two sets of student categories is used for the "other" category. The race/ethnicity tables in this report omit the "Other" category because the percentages were found to be consistently under one percent. Information based on student self-reported race/ethnicity will continue to be available on the NAEP Data Tool (*http://nces.ed.gov/nationsreportcard/naepdata*).

Type of Location

In most NAEP assessments, results are reported for students attending schools in three mutually exclusive location types: central city, urban fringe/large town, and rural/small town. Results for the NAEP 2002 Trial Urban District Assessment are reported for students attending schools in one type of location— central city.

Following standard definitions established by the Federal Office of Management and Budget, the U.S. Bureau of the Census (see *http://www.census.gov/*) defines "central city" as the largest city of a Metropolitan Statistical Area (MSA) or a Consolidated Metropolitan Statistical Area (CMSA). An MSA is an area defined by the federal government for the purposes of presenting general-purpose statistics for metropolitan areas. Typically, an MSA contains a city with a population of at least 50,000 and includes its adjacent areas. An MSA becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has

[16] In 2002, the first-stage sampling units are public schools in the selection of the combined sample. Further details about the procedure for determining minimum sample size will appear in technical documentation section of the NAEP web site at *http://nces.ed.gov/nationsreportcard*.

a population of 1,000,000 or more, its component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation.

In the NCES Common Core of Data (CCD) locale codes are assigned to schools. For the definition of central city used in this report, two locale codes of the survey are combined. The definition of each school's type of location is determined by the size of the place where the school is located and whether or not it is in an MSA or CMSA. School locale codes are assigned by the U.S. Bureau of the Census (see *http://www.census.gov/*). For the definition of central city NAEP reporting uses data from two CCD locale codes: large city (a central city of an MSA or CMSA with the city having a population greater than or equal to 250,000) and midsize city (a central city of a MSA or CMSA having a population less than 250,000). Central city is a geographical term and is not synonymous with "inner city."

The boundaries of an urban school district and a city may not always coincide. Los Angeles Unified, for example, extends beyond the city boundaries and includes urban fringe areas of the MSA, although the entire district is coded as central city. Most of the other districts included in this report have school districts that share the same boundaries as the city. The interested reader may view the School District Demographics website at *http://nces.ed.gov/surveys/sdds*, where the school district's boundaries can be shown on a map that also has county boundaries. In the Houston and Los Angeles districts, some students attended schools located in the category of urban fringe/large town. These included 6 percent and 19 percent of fourth-grade students in Houston and Los Angeles respectively, as well as 24 percent of grade 8 students in Los Angeles. Urban fringe/large town is a NAEP classification that combines three categories: urban fringe of large city, urban fringe of midsize city, and large town. An urban fringe includes all densely settled places within MSAs that are classified as urban by the U.S. Census Bureau. A large town is defined as a place outside MSAs with a population of less than 25,000 but greater than or equal to 2,500. Across the total sample for the NAEP reading assessment in 2002, 28 percent of students attended schools classified as central city, 42 percent attended schools classified as urban fringe/large town, and 30 percent attended schools classified as rural.

Eligibility for Free/Reduced-Price School Lunch

Based on available school records, students were classified as either currently eligible for the free/reduced-price school lunch component of the Department of Agriculture's National School Lunch Program or not eligible. Eligibility for the program is determined by students' family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level, and reduced-price lunch qualification is set at 170 percent of the poverty level. The classification applies only to the school year when the assessment was administered (i.e., the 2001–02 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."

Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 2002 reading assessment. The 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 Pearson Educational Measurement. Most importantly, NAEP is grateful to the students and school staff who made the assessment possible.

The NAEP 2002 reading assessment was funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. The Associate Commissioners of Education Statistics, Peggy Carr and Val Plisko, and the NCES staff—Arnold Goldstein, Steven Gorman, Carol Johnson, Andrew Kolstad, Taslima Rahman and Marilyn Seastrom—worked closely and collegially with the authors to produce this report.

The NAEP project at ETS is directed by Stephen Lazer and John Mazzeo, with assistance from John Barone. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities were conducted by Pearson Educational Measurement under the direction of Brad Thayer, Connie Smith, and William Buckles.

Test development activities took place at ETS under the direction of Patricia Donahue with assistance from Robert Finnegan.

The complex statistical and psychometric activities necessary to report results for the NAEP 2002 reading assessment were directed by Catherine Hombo and Jinming Zhang, with assistance from Andreas Oranje, Hui Deng, Kelvin Gregory, and Ying Jin.

The extensive data processing and computer programming activities underlying the statistical and psychometric analyses conducted at ETS are under the direction of David Freund, Edward Kulick, Bruce Kaplan, and Steven Isham. Data analyses presented in this report were managed by Steven Isham and Tatyana Petrovicheva with assistance from Laura Jerry, Youn-hee Lim, Norma Norris, Alfred Rogers, Fred Schaefer, Mike Weiss, and John Willey. The complex database work for this assessment was managed by Katherine Pashley with assistance from Gerry Kokolis.

The design and production of this report were overseen by Loretta Casalaina with assistance from Joseph Kolodey and Rick Hasney. Wendy Grigg coordinated the documentation and data checking procedures with assistance from Janice Goodis, Ming Kuan, Andrea Bergen, and Alice Kass. Arlene Weiner coordinated the editorial and proofreading procedures with assistance from Trish Hamill. The Web version of this report was coordinated by Rick Hasney with assistance from Loretta Casalaina. The consistency review process was coordinated by Pat O'Reilly.

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 the final version of this report: Peter Afflerbach, James Carlson, Christopher T. Cross, Kerry Gruber, Elvie Gemino Hausken, Sharon Lewis, and Aileen Waters.

United States
Department of Education
ED Pubs
8242-B Sandy Court
Jessup, MD 20794-1398

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17