NCES
National Center for Education Statistics

The Nation's Report Card NAEP

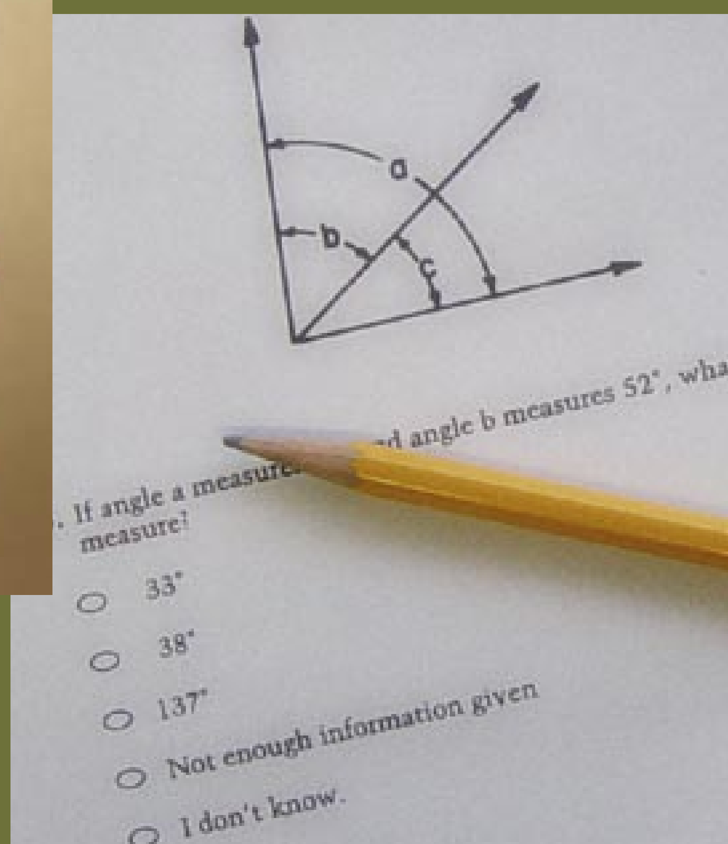
U.S. Department of Education
Institute of Education Sciences
NCES 2005–464

The Nation's Report Card™

NAEP 2004
Trends in Academic Progress

Three Decades of Student Performance in Reading and Mathematics

READING, 1971–2004



MATHEMATICS, 1973–2004

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 subjects.

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.

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 and relevant variables is collected under this program. The privacy of individual students and their families is protected, and the identities of participating schools are not released.

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 nonideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of NAEP reports.

NAEP 2004 Trends in Academic Progress

Three Decades of Student Performance in Reading and Mathematics

July 2005

U.S. Department of Education
Institute of Education Sciences
NCES 2005–464

Marianne Perie
Rebecca Moran
Anthony D. Lutkus
Educational Testing Service

William Tirre
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Grover J. Whitehurst
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in other countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

July 2005

The NCES World Wide Web Home Page address is http://nces.ed.gov
The NCES World Wide Web Electronic Catalog address is http://nces.ed.gov/pubsearch

SUGGESTED CITATION

Perie, M., Moran, R., and Lutkus, A.D. (2005). *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics* (NCES 2005–464). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

FOR ORDERING INFORMATION ON THIS REPORT, WRITE:

U.S. Department of Education, ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398, or call toll free 1-877-4ED-Pubs; or order online at http://www.edpubs.org

CONTENT CONTACT:

William Tirre
202-502-7361
William.Tirre@ed.gov

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

Since its inception in 1969, NAEP has tracked trends in student performance over time.

Executive Summary

The citizens and leaders of the United States have long valued education as a foundation for democracy, a resource for economic prosperity, and a means of realizing personal goals and individual potential. Throughout the nation's history, the commitment to educate children has grown stronger and more inclusive, and in recent decades, so has the expectation that our nation's schools and teachers be accountable (Ravitch 2002). In 2002, the reauthorization of the Elementary and Secondary Education Act—also known as the No Child Left Behind (NCLB) Act—further strengthened that commitment and expectation.

Since its inception in 1969, the National Assessment of Educational Progress (NAEP) has served the important function of measuring our nation's educational progress by regularly administering various subject-area assessments to nationally representative samples of students. One of the primary objectives of NAEP is to track trends in student performance over time. This report presents the results of NAEP long-term trend assessments in reading and mathematics, which were most recently administered in 2004 to students ages 9, 13, and 17. Because the assessments have been administered at different times in the 35-year history of NAEP, they make it possible to chart educational progress since 1971 in reading and 1973 in mathematics. Prior to 2004, the most recent long-term trend assessment was given in 1999, when results were reported for reading, mathematics, and science.

It should be noted that these long-term trend assessments are different from more recently developed assessments in the same subjects that make up the "main NAEP" assessment program. Because the instruments and methodologies of the two assessment programs are different, comparisons between the long-term trend results presented in this report and the main assessment results presented in other NAEP reports are not possible.

Approximately 38,000 students participated in the reading assessment, and 37,000 participated in the mathematics assessment. Appendix A provides technical information on this study, including sample sizes and a description of the significance tests done on each set of results. Only differences that have been determined to be statistically significant at the 0.05 level after controlling for multiple comparisons are included in this report.

National Results

National results, provided in chapter 2, are described in three ways: average score, score at selected percentiles, and percentage of students performing at or above each performance level. Student performance in each subject area is summarized as an average score on a 0–500 scale. The five long-term trend performance levels presented in this report were set at 50-point intervals on the two subject-area scales to provide a verbal description of student performance at different points on the scale. All national findings are reported from 1971–2004 for reading and 1973–2004 for mathematics. The primary findings include the following:

Average Scores

- Between 1999 and 2004, average reading scores increased at age 9 and average mathematics scores increased at ages 9 and 13. No measurable changes in average scores were found at age 17 in either subject between 1999 and 2004.
- In reading, 9-year-olds scored higher in 2004 than in any previous assessment year, with an increase of 7 points between 1999 and 2004. Average scores for age 13 showed no measurable differences between assessment years 1999 and 2004, but still were higher in 2004 than the scores in 1971 and 1975. For age 17, the average score in 2004 was not measurably different from the average score in the first assessment year, 1971.
- The average score in mathematics at age 9 was higher in 2004 than in any previous year—9 points higher than in 1999. The average score for 13-year-olds increased between 1999 and 2004 by 5 points. The average score at age 17 was not measurably different from 1973 or 1999.

Percentiles

- The reading score of 9-year-olds at the median (50th percentile) was higher in 2004 than the median score in every other year.
- Overall gains in reading scores for 13-year-olds were evident among higher performing students—those scoring at the 75th and 90th percentiles—between 1971 and 2004.
- Seventeen-year-olds showed no measurable improvements in reading scores at any of the selected percentiles between 1999 and 2004 or between 1971 and 2004.
- Mathematics scores for 9-year-olds at each of the selected percentiles showed gains between 1978 and 2004, increasing 26 points at the 10th percentile, 23 points at the 50th percentile, and 18 points at the 90th percentile.
- The mathematics score for 13-year-olds at each of the five percentile levels was higher in 2004 than in every previous assessment year, except at the 10th percentile.
- Mathematics scores for 17-year-olds in 2004 showed no measurable change since 1992 at any of the five percentiles.

Performance Levels

- The partially developed skills and understanding associated with reading at level 200 were demonstrated by 70 percent of 9-year-olds in 2004, more than in any other assessment year except 1980; by 94 percent of 13-year-olds; and by almost all 17-year-olds.
- The percentages of 13-year-olds and 17-year-olds who demonstrated the ability to interrelate ideas and make generalizations in reading (level 250) were 61 percent and 80 percent, respectively, in 2004, not measurably different from those in 1971 and 1999.
- Reading performance at or above level 300—understanding complicated information—was demonstrated by 38 percent of 17-year-olds in 2004, down from 41 percent a decade earlier in 1994.

- The beginning skills and understandings characteristic of level 200 in mathematics were demonstrated by 89 percent of 9-year-olds in 2004, more than in any other assessment year. Approximately 99 percent of 13-year-olds also demonstrated at least this level of performance in 2004.
- At age 13, the percentages of students at level 300 in mathematics increased from 17 percent in 1990 to 23 percent in 1999 and then to 29 percent in 2004. Students at this level could perform moderately complex procedures and use logical reasoning to solve problems. In 2004, 59 percent of 17-year-olds were at or above level 300 in mathematics, an increase of 7 percentage points from 1978.
- Across the assessment years in mathematics, between 5 and 8 percent of 17-year-olds performed at level 350, the highest performance level, in which students applied a range of reasoning skills to solve multistep problems.

Student Group Results

Chapter 3 describes the average scores for various groups of students, including male and female students; White, Black, and Hispanic students; and student-reported levels of parents' education, which included less than high school, graduated from high school, some education after high school and graduated from college. Some of the results were as follows:

Gender

- At all three ages in 2004, female students had higher average reading scores than their male counterparts.
- In 2004, there was no measurable difference between the average mathematics scores of male and female students at age 9, but at ages 13 and 17, male students scored higher on average than female students.
- The gender gap for 9-year-olds' reading scores in 2004 was smaller than the gaps in the first three assessment years and 1996. This gap did not change measurably between 2004 and any previous assessment year for 13-year-olds. This score gap in 2004 showed no measurable difference for 17-year-olds from the gap in 1999 or 1971.

Race/Ethnicity

- White students had higher average reading scores in 2004 than in 1971 at ages 9 and 13.
- For Black students at all three ages, average reading scores in 2004 were higher than in 1971.
- Although White students continue to outscore Black students, the White-Black score gap in reading narrowed from 1971 to 2004 at all three ages. The White-Black reading score gap for 9-year-olds decreased from 35 points in 1999 to 26 points in 2004.
- For Hispanic students, the average reading score at age 9 was higher in 2004 than in any other assessment year. Their average score at age 13 was higher in 2004 than in 1975, but not measurably different from that in 1999. No measurable difference was found between the average score for Hispanic students at age 17 in 2004 and that in 1999.
- Although White students continue to outscore Hispanic students, the White-Hispanic reading score gap for students at age 9 in 2004 was smaller than it was in 1994, 1984, 1980, and 1975. The White-Hispanic reading score gap for 13-year-olds showed no measurable difference between 2004 and 1999 or 1975. The score gap between White and Hispanic students at age 17 was measurably smaller in 2004 than in 1975.
- White students at all three ages scored higher, on average, in 2004 than in 1973 in mathematics.
- The average mathematics scores for Black students were higher in 2004 than in 1973 at all three ages. Average scores for Black students at ages 9 and 13 were higher in 2004 than in any previous assessment year.
- The differences in average scores for White and Black students at all ages decreased between the first (1973) and the most recent (2004) assessment in mathematics, although White students continued to outscore Black students in 2004. During this same period, the White-Black score gaps in mathematics narrowed by 12, 19, and 12 points for ages 9, 13, and 17, respectively.

- Hispanic students' performance in mathematics was higher at all three ages in 2004 than in any assessment year from 1973 through 1982. Average scores for Hispanic students at ages 9 and 13 were higher in 2004 than in any previous assessment year.
- White students scored higher on average than Hispanic students at all three age levels in 2004. For ages 13 and 17, the White-Hispanic score gap was smaller in 2004 than in 1973, but for age 9 there was no measurable difference in the size of the score gap between the first (1973) and most recent (2004) assessment year.

Parents' Education

- In 2004, the percentage of students reporting that at least one parent graduated from college has increased since 1980 for reading and 1978 for mathematics, while the percentage of students reporting that the highest level of education for their parents was a high school diploma or less has decreased.
- At age 13, there have been no measurable changes in average reading scores between 2004 and any previous assessment year regardless of the level of parents' education reported by the student.
- The average reading score for 17-year-olds who indicated that at least one parent had some education after high school was lower in 2004 than in any previous assessment year. For 17-year-olds who indicated that at least one parent graduated from college, the average score in 2004 (298) was lower than the average scores in 1990 (302) and 1984 (302).
- Students who reported that their parents had less than a high school education showed no measurable change in average mathematics score between 1999 and 2004 at either age 13 or 17, but their 2004 scores were higher than those in 1978.
- For students whose parents' highest education level was high school graduation or some education after high school, the average mathematics score at age 13 was higher in 2004 than in any other assessment year, while at age 17 there were no measurable changes between 1978 and 2004.
- For students with at least one parent who graduated from college, the average mathematics score in 2004 was higher than in any other assessment year at age 13; no measurable difference was seen at age 17 between 1978 and 2004.

Contextual Variables

As described in chapter 4, examining student scores in the context of their learning and home environments provides useful information. Learning and home factors for which trends are reported include students' reports of how often they read for fun, completed homework, used computers, and watched television, and the advanced mathematics courses they had taken. Some of the findings include the following:

Homework. Students who took the reading assessment were asked how many hours they had spent on homework the previous day.

- The percentage of students at age 9 indicating that no homework was assigned or that they did not do any homework decreased between 1984 and 2004. In 2004, a greater percentage of 9-year-olds indicated that they spent less than 1 hour on homework than in any other year in which the question was asked.
- In 2004, the average reading score of 9-year-olds who spent less than 1 hour on homework was higher than the average reading scores of students who did not do the homework that was assigned or who spent more than 2 hours on homework.
- At age 13, the percentage of students spending less than 1 hour on homework increased from 36 percent in 1984 to 40 percent in 2004. At the same time, the percentage of students spending 1 to 2 hours on homework decreased from 29 percent in 1984 to 26 percent in 2004.

- At age 13, students who spent 1 to 2 hours or 2 or more hours on homework had higher average reading scores than their peers who spent less than 1 hour on homework, did not do their homework, or did not have any homework to do.
- At age 17, the percentage of students reporting that they were not assigned homework increased from 22 to 26 percent. At the same time, the percentage of 17-year-olds indicating they had spent 1 to 2 hours on homework the previous day decreased from 27 to 22 percent between 1984 and 2004.
- At age 17, students who spent 2 or more hours on homework had higher average reading scores in 2004 than those who spent 1 to 2 hours, whose scores were higher than those who spent less than 1 hour, whose scores in turn were higher than those who did not do any homework.

Reading for Fun. Students who took the reading assessment were asked to estimate how often they read for fun.

- There were no measurable changes between 1984 and 2004 in the percentage of 9-year-olds indicating that they read for fun almost every day. At ages 13 and 17, the percentage saying they read for fun almost every day was lower in 2004 than in 1984. This trend was accompanied by an increase over the same 20-year time period in the percentage indicating that they never or hardly ever read for fun.
- At all three ages, students who indicated that they read for fun almost every day had higher average reading scores in 2004 than those who said that they never or hardly ever read for fun. Students at all three age levels who said that they read for fun once or twice a week had higher average scores than those who never or hardly ever read for fun.

Computer Access and Usage. Students at ages 13 and 17 who took the mathematics assessment were asked three questions about their access to computers and how they used them.

- The percentage of 13-year-olds with access to computers in schools increased from 12 percent in 1978 to 57 percent in 2004. The percentage of students receiving instruction in computers at age 13 also increased, from 14 percent in 1978 to 48 percent in 2004. In the 2004 assessment, 69 percent of 13-year-olds said that they had used a computer to solve a mathematical problem.
- Similar increases were also seen among 17-year-olds, where the percentage of students with access to a computer in school increased by 33 percentage points between 1978 and 2004. The percentage of 17-year-olds using a computer to solve mathematics problems increased from 46 percent in 1978 to 66 percent in 1999, then to 70 percent in 2004. In that year, 36 percent reported that they had studied mathematics using computers.
- There were no measurable differences in mathematics scores between 13-year-olds who responded positively and those who responded negatively to any of the computer access and usage questions in 2004. At age 17, students who indicated that they had access to a computer at school scored 5 points higher in 2004 than students who did not have such access.
- In 2004, students at age 17 who reported that they had used a computer to solve a mathematical problem scored 6 points higher on average than students who had not used a computer for that purpose. There was no measurable difference in average mathematics scores for 17-year-olds based on whether or not they had studied mathematics using computers.

Course-Taking Patterns in Mathematics. Students at age 17 who took the mathematics assessment were asked to check all the mathematics courses they had taken or were currently taking. The highest course checked was used for the analyses.

- A greater percentage of 17-year-olds indicated they were taking or had taken calculus in 2004 than in any previous assessment year. The percentage taking second-year algebra increased from 37 percent in 1978 to 53 percent in 2004, while the percentage of students who indicated that the highest level of mathematics they had taken by age 17 was pre-algebra or algebra was lower in 2004 than in 1978.
- The trend towards higher-level course-taking was seen across all three racial/ethnic groups shown. The percentage of White, Black, and Hispanic students who indicated that their highest course was second-year algebra was higher in 2004 than in 1978. In 2004, a higher percentage of White students took calculus (19 percent) compared to Black students at the same age (8 percent). At 14 percent, the percentage of Hispanic students taking calculus was not measurably different from the percentage of either White or Black students in 2004.

2004 Bridge Study

Several changes were made to the long-term trend assessment in 2004 to align it with current assessment practices and policies applicable to the NAEP main assessments. These changes, discussed in detail in chapter 5, included replacing items that had outdated material, eliminating blocks of items for subjects no longer reported, replacing background questions, and changing some administration procedures. In addition, the 2004 modified assessment provided for the inclusion of and accommodations for students with disabilities and English language learners.

A bridge study was conducted to ensure that the interpretation of the assessment results remains constant over time. A bridge study involves administering two assessments: one that replicates the assessment given in the previous assessment year (a bridge assessment), and one that represents the new design (a modified assessment). In 2003–2004, students were randomly assigned to take either the bridge assessment or the modified assessment. The bridge assessment replicated the instrument given in 1999 and used the same administration techniques. The modified assessment included the new items and features discussed above. This modified assessment will provide the basis of comparison for all future assessments, and the bridge study will link its results back to the results of the past 33 years. The results from the bridge study are presented in chapters 2 and 4, and comparisons between the two assessments are provided in chapter 5.

- Comparing the results of the modified and bridge assessments demonstrates that the link between the 2004 bridge and modified assessments was successful.

Acknowledgments

The authors wish to thank all of those who contributed to the design, writing, production, and review of this report for their thoughtful critique, careful fact checking, insightful suggestions, and creativity. In particular, the authors acknowledge the contributions of Steve Gorman, James Griffith, Lisa Hudson, Andrew Kolstad, Taslima Rahman, and Marilyn Seastrom of the National Center for Education Statistics (NCES); Mary Crovo, Larry Feinberg, Ray Fields, Susan Loomis, and Sharif Shakrani of the National Assessment Governing Board; and Tajuana Bates, Yang Chin, Kim Gettis, Dana Kelly, and Alan Vanneman of the Education Statistics Services Institute. Outside reviewers included Lisa Clarke and Nancy Horkay.

This report would not have been possible without the ETS team of researchers and data analysts who designed the administration and analyzed the results to produce the data included in this report. These ETS staff include Andreas Oranje, Brenda Tay-Lim, Amy Dresher, Lydia Gladkova, and Catherine McClellan from the research group; and Tatyana Petrovicheva, Yuxin Tang, David Freund, Laura Jerry, Matthew Kandathil, Edward Kulick, Youn-Hee Lim, Mei-Jang Lin, and Haiyang Liu from the data analysis group. Gloria Dion, David Garber, and Patricia Donahue prepared the items for release. Marilyn Binkley from NCES also worked on the assessment development and design of the long-term trend.

Many people at ETS were responsible for the careful edit and review of these chapters. In particular the authors wish to thank our advisors, Nancy Mead and Wendy Grigg; our editors, Arlene Weiner, Mary Daane, and Linda Myers; and our quality control staff, Ming Kuang, Carmen Payton, Janice Goodis, and Chrystal Murphy. Finally, we wish to thank the ETS production staff, Loretta Casalaina, Rick Hasney, and Susan Mills, for transforming our final chapters into a crisp and attractive report, and our special staff assistant, Karen Damiano, who helped ensure that the final details were completed.

Contents

List of Tables

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK.

The long-term trend assessment has been measuring student progress in reading for 33 years and in mathematics for 31 years.

Chapter 1
Introduction

The citizens and leaders of the United States have long valued education as a foundation for democracy, a resource for economic prosperity, and a means of realizing personal goals and individual potential. Throughout the nation's history, the commitment to educate children has grown stronger and more inclusive, and in recent decades, so has the expectation that our nation's schools and teachers be accountable (Ravitch 2002). In 2002 the reauthorization of the Elementary and Secondary Education Act—also known as the No Child Left Behind (NCLB) Act—further expanded that commitment and expectation.

As educators and policymakers turn their attention to student achievement as measured by assessments, examining trends—student performance now compared to in the past—can inform efforts to increase student performance in the future. The National Assessment of Educational Progress (NAEP) is one of the most important resources for monitoring the student achievement. Since its inception in 1969, NAEP has served the important function of measuring our nation's educational progress by regularly administering various subject-area assessments to nationally representative samples of students. One of the primary objectives of NAEP is to track trends in student performance over time. This report presents the results of NAEP long-term trend assessments in reading and mathematics, which were administered in the 2003–2004 school year (referred to hereafter as 2004) to students ages 9, 13, and 17. Because the same assessments have been administered at different times in the 35-year history of NAEP, they make it possible to chart educational progress since 1971 in reading and 1973 in mathematics.

The specific focus of this long-term trend report is to compare student performance in 2004 to past performance, measured by the most recent assessment in 1999 and previous assessments back to the early 1970s.

NAEP Assessments

NAEP is a project of the National Center for Education Statistics (NCES) within the Institute of Education Sciences of the U.S. Department of Education. The National Assessment Governing Board (NAGB), an independent group created by Congress in 1988, provides policy direction for NAEP. (Information about NAGB can be found on its website, http://www.nagb.org/.)

NAEP includes two components: the long-term trend assessments and the main assessments. The existence of the two national assessment programs—long-term trend and main—makes it possible for NAEP to meet two important objectives. The long-term trend program uses substantially the same assessments decade after decade, each time a subject is assessed, in order to measure student progress in that subject over time. In contrast, the main NAEP assessments are periodically adapted to reflect contemporary curriculum policies, content currently in use in the nation's schools, and improvements in techniques of educational measurement. In this way, main NAEP can provide valid data for those seeking evidence for contemporary questions, and long-term trend NAEP can provide data for evaluating change over long periods. For example, while the current main NAEP reading assessment, given in 2005, was first administered in 1992, the long-term trend reading assessment dates back to 1971.

This report presents the results from the long-term trend assessments only. Because the long-term trend assessments use different questions from those used in the main assessments, and because students are sampled by age for the long-term trend assessments, rather than by grade as in the main assessments, it is not possible to compare results from the two assessment programs.

Overview of the 2004 Long-Term Trend Assessments

The long-term trend assessment originally was given in four subjects: mathematics, science, reading, and writing. At the time of the last long-term trend report (1999), NAGB discontinued the assessment in writing for technical reasons. More recently, NAGB decided that changes were needed to the design of the science assessment and, given recent advances in the field of science, to its content. For instance, many science questions that were written in the late 1960s are no longer relevant, as they were first written before Neil Armstrong set foot on the moon, before computers could fit onto a desk, and without the knowledge of many medical and biotechnology breakthroughs of the late 20th century. NAGB decided that the long-term trend assessment in science required technical studies of the required changes, so that valid comparisons between the updated assessment and the original assessment could still be made. To allow time to update the assessment and study the changes, the decision was made not to assess science in 2004.

According to NAGB's new policy, reading and mathematics would continue to be assessed by the long-term trend and main NAEP instruments, but science and writing would be assessed only in main NAEP. As a result, changes were needed to separate out the sets of questions (blocks) for science and writing, which had been intermixed with the reading and mathematics blocks in the long-term assessment instruments. New booklets consist only of reading or only of mathematics blocks. The changes provided an opportunity to bring other aspects of the assessment up to date. Considerable progress in testing theory has been made since the late 1960s, and the 2004 administration provided a platform to bring these improvements to the long-term trend assessments, in areas such as scoring and scaling. In addition, main NAEP assessments had begun providing accommodations to allow students with disabilities and students who were not fluent in English to participate. In 2004, it was possible to implement the modifications to the long-term trend assessments resulting in the assessment of a greater proportion of students using accommodations.

Any time changes are made in a long-term trend assessment, studies are required to ensure that the results can continue to be reported on the same trend line—that is, that they are validly comparable to earlier results. So analyses were needed to ensure that the 2004 results under the new design were comparable to the results from 1971 through 1999, under the design that existed earlier. Therefore, two assessments were conducted in 2004. The modified assessment used the new design, and the "bridge" assessment replicated the former design. Comparisons of the results can then detect any shifts in results due to changes in test design. This bridge assessment links the old assessments to the new one.

2004 Bridge Study

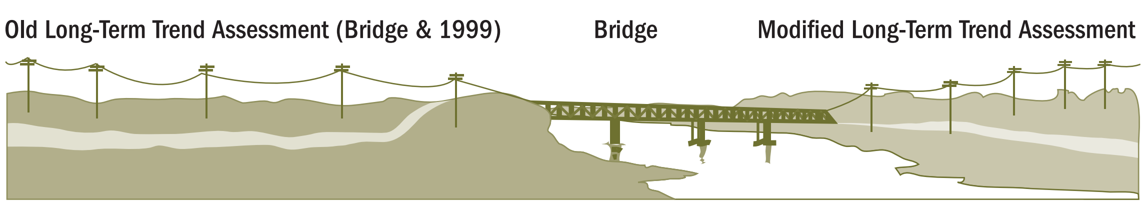
This section of the report presents a brief description of the 2004 bridge study, the modified assessment, and the long-term trend instruments. (More detailed information about the instruments and methodology is provided in appendix A.) The changes made for the modified 2004 assessment included replacing items containing outdated material, eliminating blocks of items for subjects no longer reported, replacing background questions, allowing accommodations for students who needed them, and changing some administrative procedures. For example, previous long-term trend assessments in mathematics included an audio portion that paced students, so they were always at the same place in the test booklet at the same time. The audiotape was eliminated in the modified design so that students could move at their own pace within each section. Another example is that students used to have the option of selecting "I don't know" as a response to a multiple-choice item. That response was eliminated in the modified assessment. Also, in prior assessments, the student's race/ethnicity was reported based on a test administrator's classification of the student's visual appearance. In 2004, both schools and students were asked to report each student's race/ethnicity as part of the school and student questionnaires. Finally, the 2004 modified assessment provided for the inclusion of and accommodations for students with disabilities and English language learners.

The changes were intended to improve the validity of the results while continuing to maintain the integrity of the long-term trend. Thus, studies were needed to ensure that the modifications did not affect the interpretation of the results. In other words, it was important to assess whether any changes in scores were due to actual changes in student performance rather than changes in the assessments themselves that may have made them easier or harder.

The bridge study was conducted to ensure that the interpretation of the assessment results remains constant over time. A bridge study involves developing two assessments: one that replicates the assessment given in the previous assessment year using the same questions and administration procedures (a bridge assessment), and one that represents the new design (a modified assessment). In 2004, students were randomly assigned to take either the bridge assessment or the modified assessment. The bridge assessment replicated the instrument given in 1999 and used the same administration techniques. The modified assessment included the new items and features discussed previously. This modified assessment will provide the basis of comparison for all future assessments, and the bridge will link its results back to the results of the past 30 years (see figure 1-1). Further detail on this study is provided in appendix A.

This report will be the final report of new results acquired under the old design using the bridge assessment. The greater part of the report uses the results from the bridge assessment to maintain the trend lines from 1971 (in reading) and 1973 (in mathematics). Differences between the old and modified formats are discussed only in one chapter, chapter 5. Beginning in 2008, only the modified design will be used, and the results will be linked back to the previous assessments through the 2004 bridge study.

Figure 1-1. Comparison of the old and new long-term trend assessment



- "I don't know" option
- Observed race/ethnicity
- Audio-paced portion
- No accommodations for SD/ELL students

- No "I don't know" option
- School-reported race/ethnicity
- Self-paced throughout each section
- Accommodations permitted

Content of the Assessments

The content of the NAEP long-term trend reading and mathematics assessments has not changed since its beginning. The reading assessment contains a range of reading materials, from simple narrative passages to complex articles on specialized topics. The selections include stories, poems, essays, reports, and passages from textbooks, as well as a sample train schedule, telephone bill, and advertisements. Students' comprehension of these materials is assessed with both multiple-choice questions, for which students choose a response from a list, and constructed-response questions, for which students are asked to write a response.

The long-term trend mathematics assessment measures students' knowledge of basic facts, their ability to carry out numerical algorithms using paper and pencil, their knowledge of basic measurement formulas as they are applied in geometric settings, and their ability to apply mathematics to daily-living skills (such as those related to time and money). The computational focus of the long-term trend assessment provides a unique opportunity to measure how students perform in traditional procedural skills.

The Long-Term Trend Background Questionnaires

In addition to assessing students' progress in reading and mathematics, the NAEP long-term trend assessments include questions about students' home and school experiences that may be related to educational achievement. For example, students are asked about the courses they have taken, activities in their classrooms, the amount of time they spend on homework, and educationally relevant uses of their time out of school. Their responses to these questions provide an informative context for interpreting the assessment results.

In the previous long-term trend assessments, these background questions were intermixed with the assessment questions. For example, students would answer questions about a reading passage to assess their understanding of that passage, and then they would respond to background questions about their reading habits. In the modified design, these background questions were reduced in number and assembled together in a separate section that students completed after finishing the assessment.

The Student Sample

The NAEP long-term trend assessments measure the performance of students at three ages—9, 13, and 17. The NAEP assessments measure the achievement of students nationally and are not intended to provide a measure of individual student performance. A nationally representative sample of students is selected, and their results are generalized to the nation as a whole. Small percentages of students with disabilities (SD) and of English language learners (ELL) are excluded in each assessment year based on their schools' judgment that they cannot be meaningfully assessed. Formerly, NAEP did not permit students so identified to receive accommodations (such as extended time, assessment in small groups, or use of bilingual dictionaries). In 2004, accommodations were permitted on the modified assessment, and therefore fewer students were excluded. Specifically, approximately 14 to 19 percent of students across the three ages and two subjects were identified as SD/ELL in 2004, resulting in an exclusion rate of 7 to 8 percent, depending on the age and subject assessed, in the nonaccommodated format. When accommodations were permitted, the exclusion rates dropped to approximately 3 percent for mathematics and 4 to 5 percent for reading. (See appendix A for information regarding exclusion criteria and exclusion rates.)

This report contains results representing the performance of all in-school 9-, 13-, and 17-year-olds in the nation who are capable of being meaningfully assessed without accommodations, except for the results from the modified assessment shown in chapter 5. In addition, it describes the performance of groups of students, such as males and females, in each age group. In 2004, more than 11,000 students at each of the three ages were assessed in each subject area, including both public and private school students. To ensure that the sample was nationally representative, a sampling plan was created to randomly select schools and students to participate. This sampling plan targeted certain schools and students for participation in NAEP. The degree to which the students who actually participated in the assessment matched the target is a measure of the reliability of the results. In 2004, approximately 80 to 81 percent of the students originally selected for the assessment at age 9 were actually assessed, 76 to 77 percent of the students at age 13, and 55 to 57 percent at age 17. (See appendix A for more information on sampling procedures and appendix B for the percentages of students in various reporting groups who were assessed.)

Reporting the Trend Results

Students' performance on the long-term trend assessments is summarized on a 0–500 scale for each subject area. For each year in which the assessments were administered, achievement in a particular subject area is described for a group of students by their average scale score and the score at the selected percentiles. Trends in student achievement are determined by examining the average scale scores attained by students in the current assessment year or the score at the selected percentiles and comparing them to the same scores in other assessment years. While the score ranges in both subjects are identical, the scale was derived independently for each subject. Therefore, average scale scores between subjects cannot be compared.

In addition to reporting average scores, student performance is described in terms of the percentages of students attaining specific levels of performance. These performance levels correspond to five points on the reading and mathematics scales: 150, 200, 250, 300, and 350. For each subject area, the performance levels from lowest to highest are associated with increasingly advanced skills and knowledge (Allen, McClellan, and Stoeckel 2005, pp. 21–22). Examining the percentages of students in each year that attained each performance level provides additional insight into student achievement.

Because the results presented in this report are based on a nationally representative sample of students, they are considered estimates of all students' average performance (excluding students who cannot be meaningfully assessed). As such, the results are subject to a degree of uncertainty, which is reflected in the standard errors of the estimates. The standard errors for all of the scale scores and percentages presented in this report can be viewed using the NAEP Data Explorer found at http://nces.ed.gov/nationsreportcard/naepdata/. Statistical tests that take into account these standard errors were conducted to determine whether apparent changes or differences in the results are measurably different in a statistical sense. When the term "significant" is used, it does not imply a judgment about the absolute magnitude or educational relevance of changes and differences in student performance. Rather, it is used to indicate that the observed changes are not likely to be due to chance factors associated with sampling and measurement error. The differences described in this report have been determined to be statistically significant at the 0.05 level with appropriate adjustments for multiple comparisons. In the tables and charts in this report, the symbol (*) is used to indicate that a score or percentage is measurably different from another. (See appendix A for additional information on analysis procedures.)

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

About This Report

This report describes trends in 9-, 13-, and 17-year-olds' achievement in reading and mathematics during the last three decades. Chapter 2 presents trends in terms of overall scale scores, percentiles, and percentages at selected performance levels for the nation. Chapter 3 examines trends in average scale scores for groups of students defined by gender, race/ethnicity, and the education level of the student's parents. Chapter 4 reports results from the NAEP long-term trend background questionnaires. In this chapter, students' school and home experiences, as shown in their responses to the background questions, are examined in relation to students' assessment scores. Chapter 5 explores the differences between the bridge assessment administered under the procedures used for earlier assessments and the modified assessment with the new design elements. The last chapter in this report provides sample items from the NAEP long-term trend assessments. For the first time, NCES is releasing items from the assessment, along with summary data that indicate how well students performed on these items. This report also contains three appendixes. Appendix A discusses technical procedures involved in collecting, analyzing, and reporting the assessment data, and appendix B is a data appendix showing the percentages of participating students in the bridge and modified samples by student groups. Appendix C provides a glossary of terms used in this report.

Additional information about the 2004 long-term trend assessments not included in this report, and other NAEP assessment reports and data, are available on the Internet at http://nces.ed.gov/nationsreportcard/. This site contains the data associated with all the figures in this report and further information on the technical features of the study. Additional data, such as the standard errors for each percentage, can also be found on this website.

Cautions in Interpreting the Long-Term Trend Results

The reader is cautioned against using the long-term trend results in this report to make simple causal inferences related to student performance, to the relative effectiveness of public and nonpublic schools, or to other educational variables discussed in this report. Simple cross-tabulations of a variable with measures of educational achievement, like the ones presented here, cannot constitute proof that differences in the variable cause differences in educational achievement. There are many possible reasons why the performance of one group of students will differ from that of another that are not discussed in this report. For example, group differences may be understood better by considering such factors as exposure to a rigorous curriculum, variations in course-taking patterns, and parental involvement.

A caution is also warranted for some small population group estimates. Smaller population groups may show increases or decreases across years in average scores; however, it is necessary to interpret such score changes with extreme caution. The effects of exclusion-rate changes for groups of students may be more marked for small groups than they are for the whole population. Another reason for caution is that the standard errors are often quite large around the score estimates for small groups, which in turn means the standard error around the gain is also large.

In addition, although in some figures trend lines for ages 9, 13, and 17 will appear in the same graphic, the reader is cautioned against making cohort comparisons. One cannot interpret the amount of growth between ages 9 and 13 from these figures by examining a 4-year time difference. Not all assessment years are four years apart, and the assessments were administered at different times of the year for the different ages. The relative merits of different types of comparisons are discussed in appendix A. Comparisons should be made within ages only.

National results are displayed using three reporting metrics: average scale scores, percentiles, and performance levels. Generally, all three metrics show improvements at age 9 in reading and mathematics.

Chapter 2
National Trends in Academic Achievement

For the past 35 years, NAEP's long-term trend assessments have documented trends in the academic achievement of America's students. Before the 2004 assessment, the last long-term trend assessment was conducted in 1999. This report examines the changes in students' performance in reading and mathematics over the past five years by comparing 2004 results to 1999 results and then provides a wider view of the overall trends in performance from the early 1970s through 2004.

This chapter presents the results by subject, first examining the trends in reading and then discussing mathematics results. There have been 11 administrations of the reading assessment since 1971 and 10 administrations of the mathematics assessment since 1973 for ages 9, 13, and 17. The next section describes the different ways of reporting results, and the remainder of this chapter describes the national trends in reading and mathematics.

How the Results Are Presented

Performance results in this chapter are reported in three ways: as average scale scores, as percentile scores, and as percentages of students reaching predetermined performance levels.

- ***Average scale scores.*** The average scale scores represent the performance of 9-, 13-, and 17-year-olds in reading or mathematics averaged across the nation. Student performance is summarized on a 0–500 scale for both reading and mathematics, where the different points on the scale represent what students know and can do at a given point in time. Although the results from both subjects are reported on the same scale, the results cannot be compared with one another, as they measure different content.

  Line graphs are provided to depict student performance on this scale across the years in both subject areas. The average scale score attained by students in each assessment year is indicated on the graph. The average scores for years prior to 2004 are highlighted with an asterisk (*) when the score is significantly higher or lower than the average score in 2004. (See appendix A for information on the statistical tests conducted.)

- ***Percentile scores.*** Going beyond average scores, useful information can be gained by examining trends of student scores falling at specified percentiles along the performance distribution. Percentiles indicate the percentage of students whose scores fell below a particular point on the NAEP scale. For example, 25 percent of assessed students' scores fell below the 25th percentile score; 75 percent fell below the 75th percentile score. This chapter provides such information by examining the scores of students at five distinct percentiles (10th, 25th, 50th, 75th, and 90th) of the score distribution in each year. Examining student performance at different percentiles on the 0–500 scale indicates whether or not the changes seen in the overall national average score results are reflected in the performance of lower-, middle-, and higher-performing students.

- ***Performance levels.*** More detailed information about what students know and can do in each subject area can be gained by examining their attainment of specific performance levels in each assessment year. For each of the subject area scales, performance levels were set at 50-point increments from 150 through 350. The five performance levels—150, 200, 250, 300, and 350—were then described in terms of the knowledge and skills likely to be demonstrated by students who reached each level. To develop these descriptions, assessment questions were identified that students at a particular performance level were more likely to answer successfully than students at lower levels. The descriptions of what students know and can do at each level are based on these sets of questions. This process of developing the performance-level descriptions is quite different from that used to develop achievement-level descriptions in the main NAEP reports as they are not set through a judgmental process. The levels for long-term trends were set arbitrarily and do not represent perfomance standards. Specific descriptions for each subject are presented later in this chapter along with the results. (The procedures for describing the performance levels are discussed in more detail in appendix A.)

National Trends in Reading Performance

National trends are shown through the average score, the percentile scores, and the percentage of students at or above each performance level. Although at first glance it may appear that this report provides the same results in three formats, these different reporting metrics actually provide different perspectives. The average score summarizes student performance in one measure. The percentiles examine performance at five different points, demonstrating whether any changes in average score are more likely due to changes in the scores of lower-performing students or higher-performing students. These percentiles are based on a normative measure, while the performance levels are based on a criterion measure. That is, the performance levels show trends in student performance at five benchmarks. These benchmarks are valid within all three age groups, permitting comparisons of the attainment of absolute performance levels over time. Cross-age comparisons can be supported, but readers are encouraged to focus more appropriately on within-grade comparisons.

Overall, the national trend in reading shows improvement across most reporting metrics at age 9 between 1999 and 2004 as well as between 1971 and 2004. Students at age 13 show no significant improvement in recent years, although most reporting metrics indicate that performance in 2004 was higher than in 1971. At age 17, no measurable differences in performance were found between 1971 and 2004 for any reporting metric.

Average Scores

This measure provides a summary account of student performance. Figure 2-1 displays the trend lines for each age, and further details are given below.

Nine-year-olds. The average reading score at age 9 was higher in 2004 than in any previous assessment year.

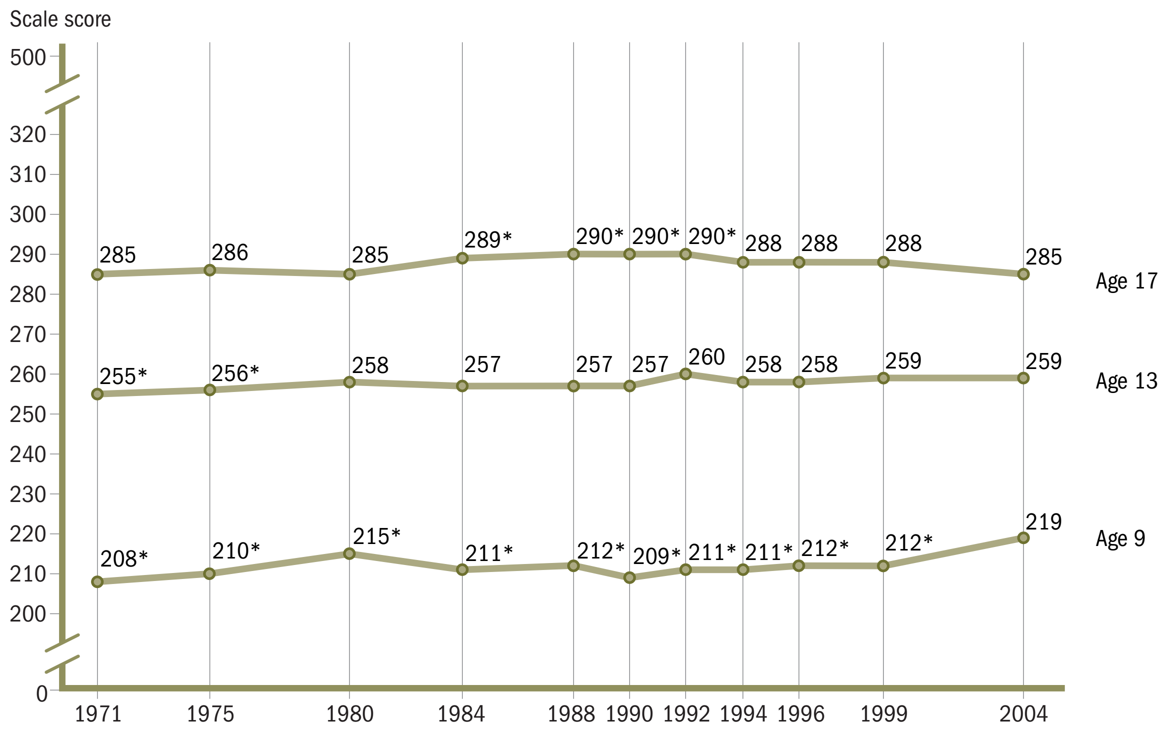
Thirteen-year-olds. The average score at age 13 was higher in 2004 than in 1971, but not measurably different from the average score in 1999.

Seventeen-year-olds. Between 1999 and 2004, average reading scores at age 17 showed no measurable changes. The average score in 2004 was similar to that in 1971.

How to interpret this graphic . . .

Graphics like these show the average scale score at each age for each year the assessment was given. Each score is plotted, and lines are drawn to connect the scores between the different years, creating trend lines. Examining the trend lines helps to determine whether scores appear to be increasing over time, or if there are any peaks or valleys in the 33-year trend. Statistically significant differences in scores between 2004 and previous years are marked with an asterisk. For example, figure 2-1 shows the trend lines of the average scores in reading for all three ages. The graphic shows that the average score at age 17 was about the same in 1971 as in 2004.

Figure 2-1. Trends in average reading scale scores for students ages 9, 13, and 17: 1971–2004



*Significantly different from 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading Assessments.

Percentile Scores

Examining the national trends at five percentiles shows whether changes seen in the national averages were sustained at every level of performance or were more likely to occur for students of specific ability levels. Figure 2-2 displays trends in reading scores for 9-, 13-, and 17-year-old students in the five percentile ranges. The results are discussed below for each age level.

Nine-year-olds. As seen in figure 2-2, only one significant increase was seen at the 90th percentile as compared to 2004. However, the score at the 50th percentile—the median—was higher in 2004 than in any other assessment year. The scores at the 10th, 25th, and 75th percentiles showed increases in performance between 1999 and 2004 and between 1971 and 2004.

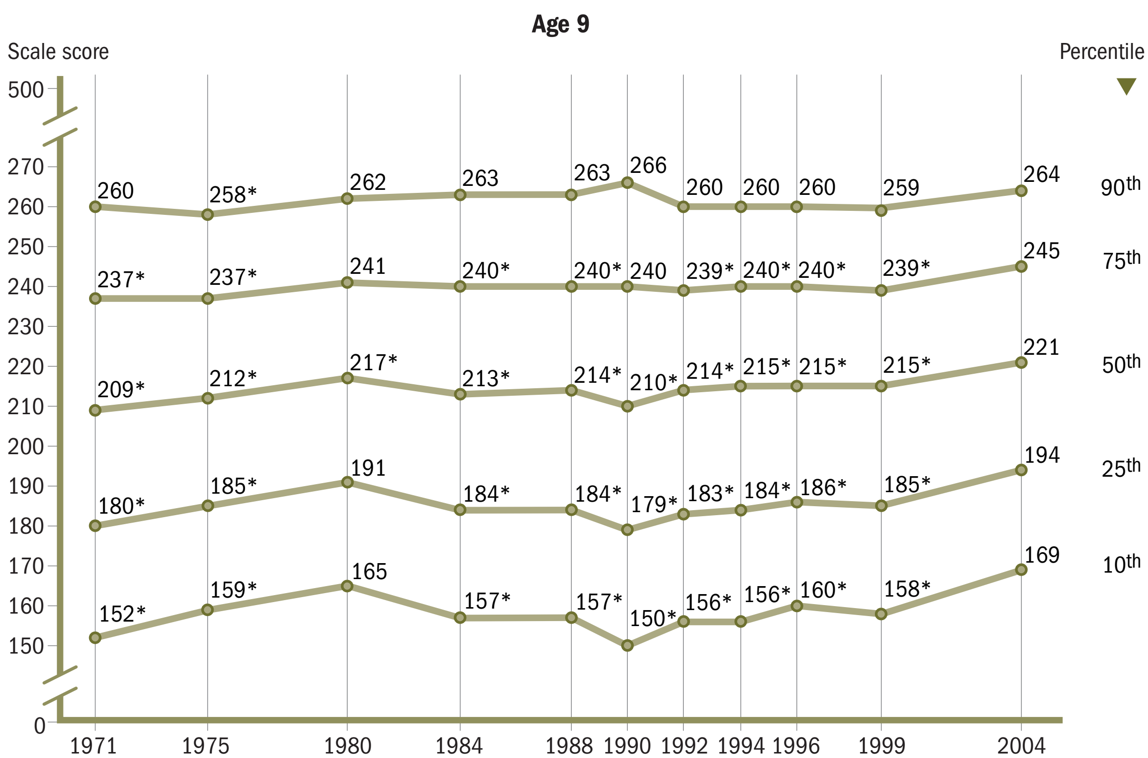
Thirteen-year-olds. The trends differ between upper and lower percentiles. The scores at the 10th, 25th, and 50th percentiles showed no measurable differences between 2004 and any previous assessment year. At the 75th and 90th percentiles, scores in 2004 were higher than in 1971, although no measurable differences were detected between the score in 2004 and that in 1999.

Seventeen-year-olds. Examining the scores at the five selected percentiles shows no measurable difference in the scores in 2004 compared to either 1971 or 1999.

How to interpret this graphic . . .

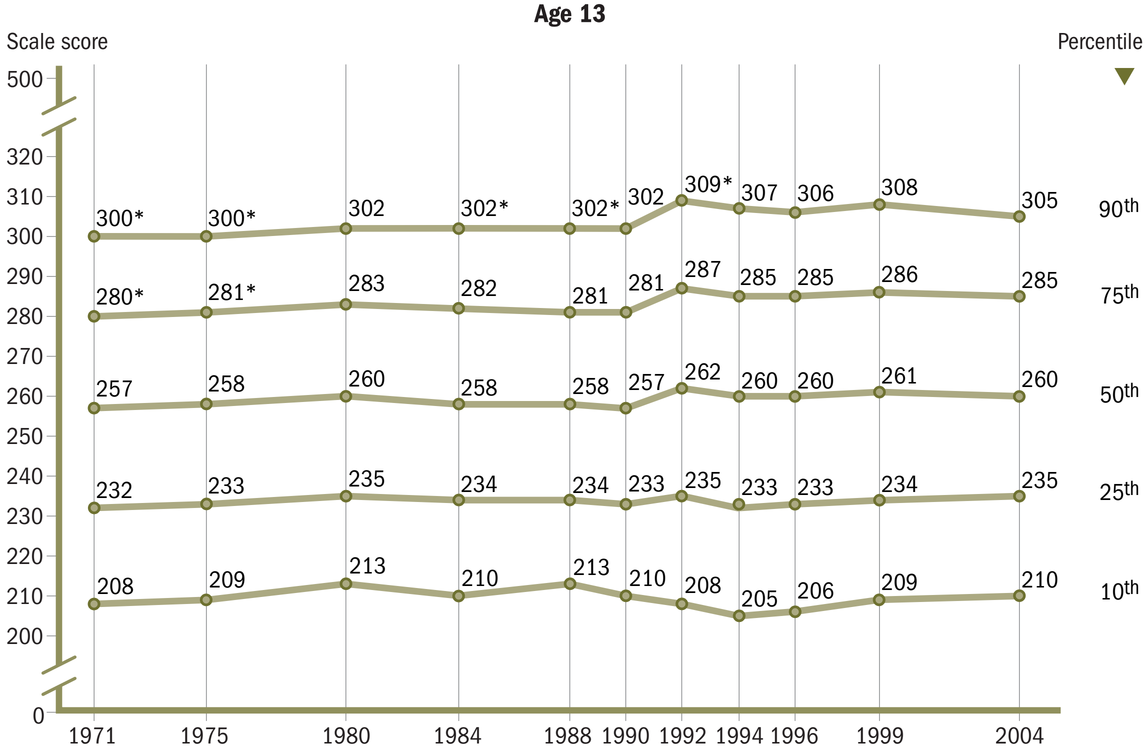
Graphics like figure 2-2 show the score at each percentile for five selected percentiles. For example, at age 9 in 2004, students at the 10th percentile scored 169 in reading, while students at the 90th percentile scored 264. Looking at the five trend lines together, it can be determined if more improvement took place at the upper end or at the lower end, or if the trend lines look the same at all five levels. For example, at age 9, the scores at the 10th, 25th, 50th, and 75th percentiles showed increases in performance between 1999 and 2004 and between 1971 and 2004.

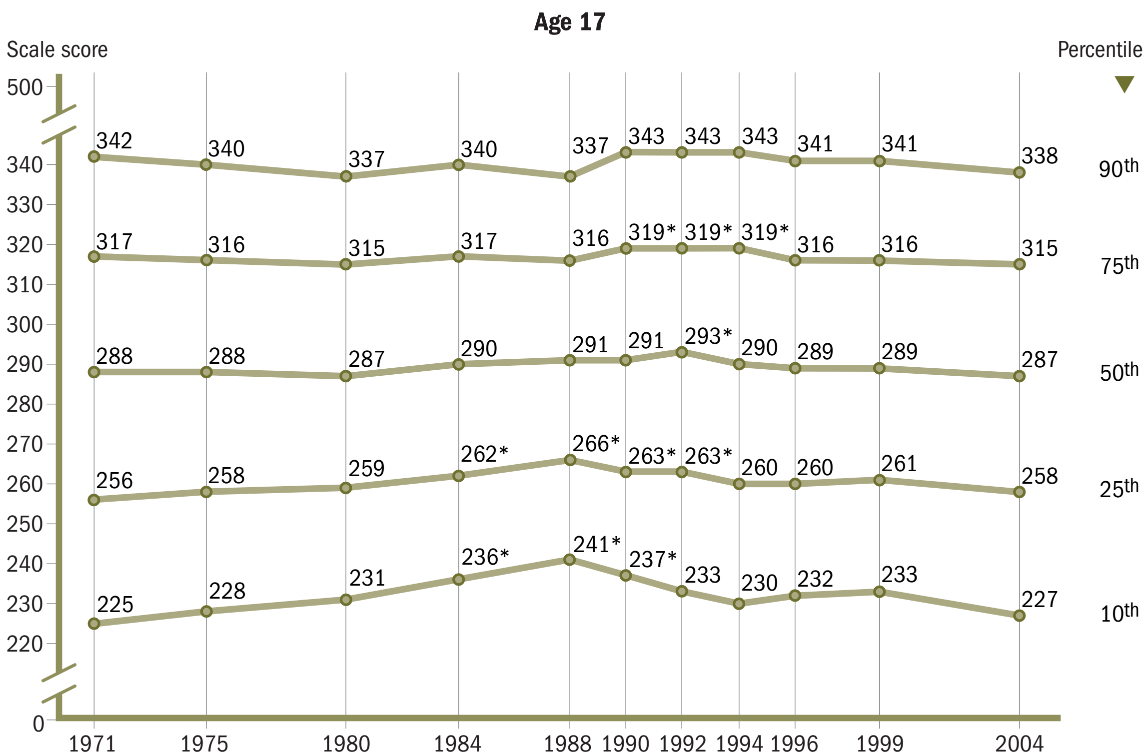
Figure 2-2. Trends in reading scale score at selected percentiles for students ages 9, 13, and 17: 1971–2004



See notes at end of figure.

Figure 2-2. Trends in reading scale score at selected percentiles for students ages 9, 13, and 17: 1971–2004—Continued





*Significantly different from 2004.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading Assessments.

Performance Levels

This section reports trend results using the performance-level reporting metric, examining the percentage of students demonstrating particular levels of performance over the past three decades. Although one would expect these trends to follow closely the trends in average scores, it is instructive to examine changes in what students now seem to know and be able to do.

The skills and abilities demonstrated by students at each reading performance level are described below. The five performance levels are applicable at all three age groups, although the likelihood of attaining higher performance levels is directly related to a student's age, because older students have completed more education in both subject areas. For this reason, only three performance levels are discussed for each age: levels 150, 200, and 250 for age 9; levels 200, 250, and 300 for age 13; and levels 250, 300, and 350 for age 17. One might expect younger students to reach only the first performance levels, as they have not yet been taught the material in the higher performance levels, and it is expected that nearly 100 percent of older students will meet the lowest performance levels. Thus, the performance-level results displayed for each age are those that are most likely to show significant change across the assessment years. The levels not shown here are those that nearly all or almost no students attained at a particular age in each year.

Reading Performance-Level Descriptions

LEVEL 350: ***Learn from Specialized Reading Materials***
Readers at this level can extend and restructure the ideas presented in specialized and complex texts. Examples include scientific materials, literary essays, and historical documents. Readers are also able to understand the links between ideas, even when those links are not explicitly stated, and to make appropriate generalizations. Performance at this level suggests the ability to synthesize and learn from specialized reading materials.

LEVEL 300: ***Understand Complicated Information***
Readers at this level can understand complicated literary and informational passages, including material about topics they study at school. They can also analyze and integrate less familiar material about topics they study at school as well as provide reactions to and explanations of the text as a whole. Performance at this level suggests the ability to find, understand, summarize, and explain relatively complicated information.

LEVEL 250: ***Interrelate Ideas and Make Generalizations***
Readers at this level use intermediate skills and strategies to search for, locate, and organize the information they find in relatively lengthy passages and can recognize paraphrases of what they have read. They can also make inferences and reach generalizations about main ideas and author's purpose from passages dealing with literature, science, and social studies. Performance at this level suggests the ability to search for specific information, interrelate ideas, and make generalizations.

LEVEL 200: ***Demonstrate Partially Developed Skills and Understanding***
Readers at this level can locate and identify facts from simple informational paragraphs, stories, and news articles. In addition, they can combine ideas and make inferences based on short, uncomplicated passages. Performance at this level suggests the ability to understand specific or sequentially related information.

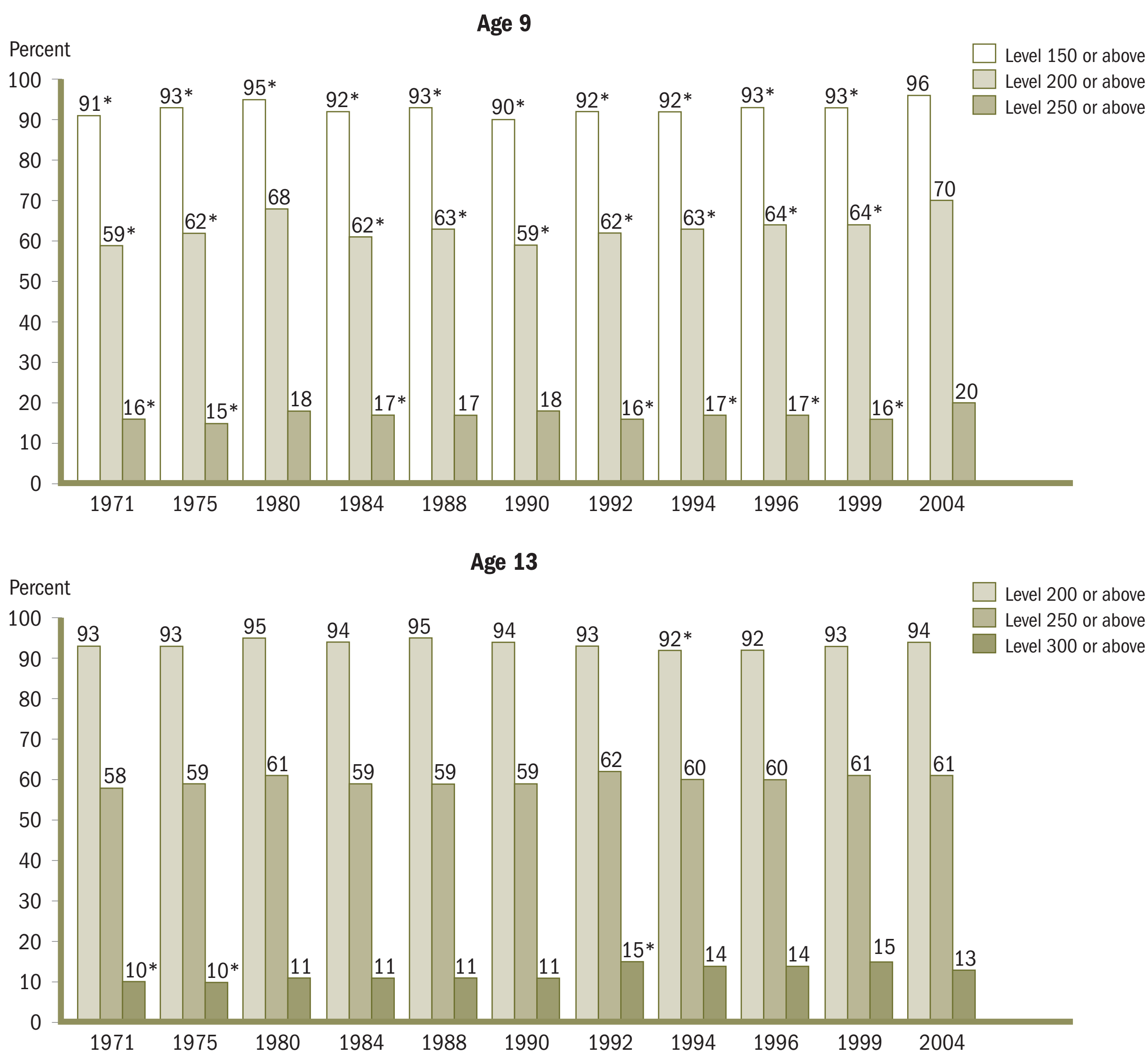
LEVEL 150: ***Carry Out Simple, Discrete Reading Tasks***
Readers at this level can follow brief written directions. They can also select words, phrases, or sentences to describe a simple picture and can interpret simple written clues to identify a common object. Performance at this level suggests the ability to carry out simple, discrete reading tasks.

Figure 2-3 shows the percentage of students reaching each performance level by age and assessment year. The following sections discuss the data for each age. It is important to keep in mind that the percentages reported for each level are cumulative. That is, the percentage shown for level 200 reflects the percentage of students who scored at 200 or above, so it also includes those who scored at 250, 300, or 350.

How to interpret this graphic . . .

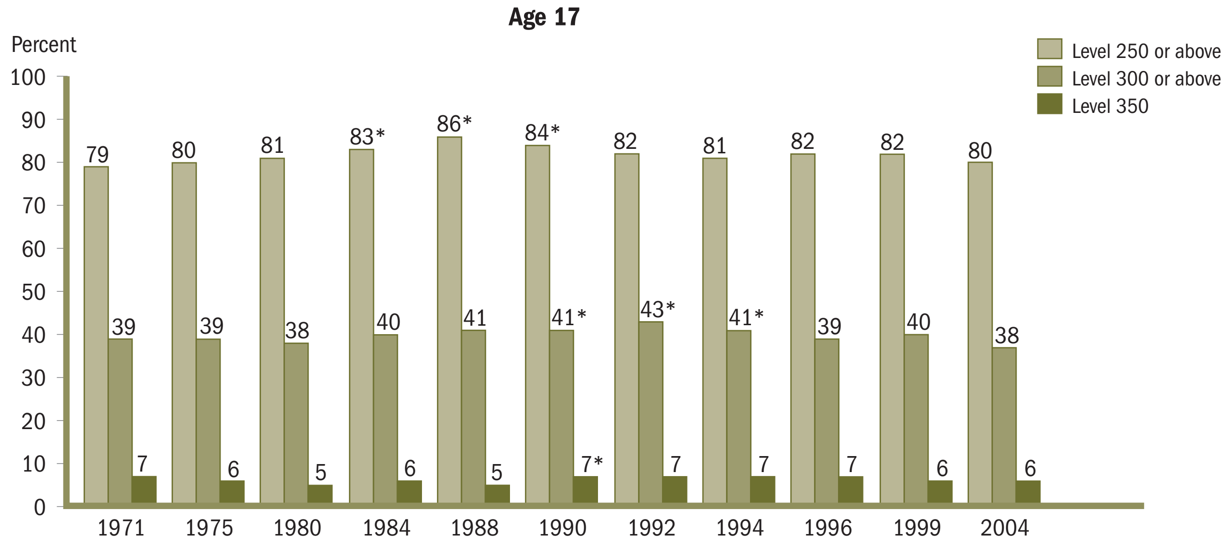
Bar charts are used to show the percentage of students who reach each performance level or above. For instance, figure 2-3 shows that 80 percent of 17-year-olds in 2004 reached level 250 or above, 38 percent reached level 300 or above, and 6 percent reached level 350. So, the 80 percent bar also includes those students in the 38 and 6 percent bars. Examining the height of the bars across years can help determine whether students are improving at the lower levels, higher levels, or both.

Figure 2-3. Trends in percentages at or above reading performance levels for students ages 9, 13, and 17: 1971–2004



See notes at end of figure.

Figure 2-3. Trends in percentages at or above reading performance levels for students ages 9, 13, and 17: 1971–2004—Continued



*Significantly different from 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading Assessments.

Nine-year-olds. Trends in the percentage of 9-year-olds scoring at or above reading performance levels 150, 200, and 250 are shown in the first panel of figure 2-3. In each assessment year, at least 90 percent of 9-year-olds performed the simple, discrete reading tasks described at level 150. In 2004, 96 percent of 9-year-olds reached level 150, a higher percentage than in any previous assessment year. The partially developed skills and understanding associated with level 200 were demonstrated by 70 percent of 9-year-olds in 2004. This number was higher than in every other assessment year with the exception of 1980, which showed no measurable difference from 2004. The ability to interrelate ideas and make generalizations (level 250) was demonstrated by 20 percent of 9-year-olds in 2004. This percentage was higher than both the more recent assessment year, 1999, and the first assessment year, 1971.

Thirteen-year-olds. The second panel of figure 2-3 displays trends in the percentage of 13-year-olds performing at or above reading performance levels 200, 250, and 300. In each assessment year, 92 percent or more of 13-year-old students performed at or above level 200, demonstrating at least partially developed skills and understanding. Ninety-four percent of students reached level 200 in 2004, which was not measurably different from the percentage in any other assessment year, except 1994, when the percentage fell to 92 percent. The ability to interrelate ideas and make generalizations (level 250) was demonstrated by 61 percent of 13-year-olds in 2004. Despite some apparent fluctuation, no measurable differences were found in the percentages of students at or above this level of performance across the assessment years. At level 300, students demonstrate the ability to understand complicated literary and informational passages. The percentage of students reaching level 300 in 2004 was higher than the percentage in 1971, mirroring the national trend for average score.

Seventeen-year-olds. Trends in the percentage of 17-year-olds scoring at or above reading performance levels 250 and 300 and at level 350 are shown in the last panel of figure 2-3. The ability to interrelate ideas and make generalizations (level 250) was demonstrated by 80 percent of 17-year-olds in 2004, which was not measurably different from 1999 or 1971. Performance at or above level 300—understanding complicated information—was demonstrated by 38 percent of 17-year-olds in 2004, which was not measurably different from the percentages in 1999 or 1971. Across all of the assessment years, only 5 to 7 percent of 17-year-olds demonstrated performance at level 350—the ability to learn from and synthesize specialized reading materials.

National Trends in Mathematics Performance

Overall, the national trend in mathematics shows improvement in performance at ages 9 and 13 in 2004 and few changes over the years at age 17. Note that the data from 1973 in figure 2-4 were extrapolated using a mean proportion correct, meaning that only average scores could be calculated. Results by percentile and performance levels are shown from 1978 through 2004. (See appendix A for further explanation of the extrapolated results.) The following sections examine the national results through the average score, the percentile scores, and the percentage of students at or above each performance level.

Average Scores

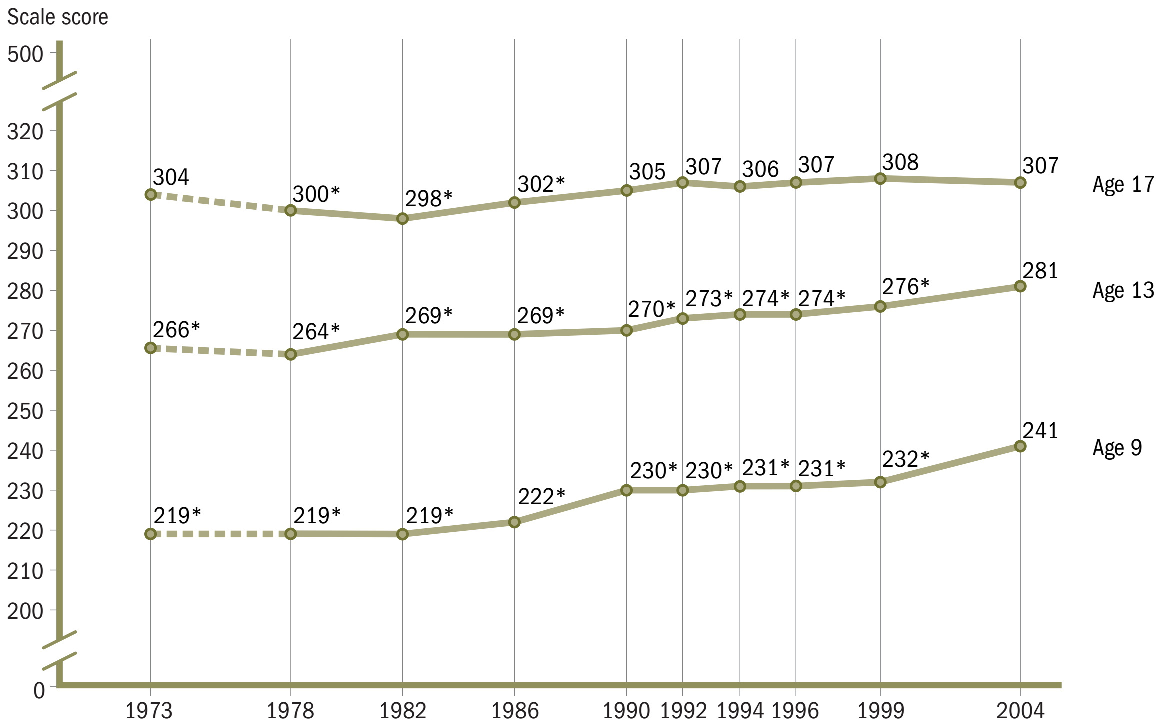
The first set of results shows trends in average scores in mathematics between 1973 and 2004. Figure 2-4 displays the trend lines for each age, and further details follow.

Nine-year-olds. At 241, the average score at age 9 was higher in 2004 than in any previous year—up 9 points from 1999 and 22 points from 1973.

Thirteen-year-olds. At age 13, the average score in 2004 was higher than in any other assessment year. The 5-point increase between 1999 and 2004 resulted in an average score in 2004 that was 15 points higher than the average score in 1973.

Seventeen-year-olds. The average score at age 17 was not measurably different from the average score in 1973 or 1999.

Figure 2-4. Trends in average mathematics scale scores for students ages 9, 13, and 17: 1973–2004



*Significantly different from 2004.
NOTE: Dashed lines represent extrapolated data.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1973–2004 Long-Term Trend Mathematics Assessments.

How to interpret this graphic . . .

Graphics like these show the average scale score at each age for each year the assessment was given. Each score is plotted, and lines are drawn to connect the scores between the subsequent assessment years, creating trend lines. Examining the trend lines helps to determine whether scores appear to be increasing over time, or if there are any peaks or valleys in the 31-year trend. Statistically significant differences in scores between 2004 and previous years are marked with an asterisk. For example, figure 2-4 shows that at age 17, the average score in 2004 was not measurably different from the scores shown in 1990 through 1999, but it was higher than the scores in 1978, 1982, and 1986.

Percentile Scores

This section examines the national trends at five percentiles to indicate whether changes seen in the national averages are sustained at every level of performance or occurred for students of specific ability levels. Figure 2-5 displays trends in mathematics scores for 9-, 13-, and 17-year-old students in the five percentile levels. Note that these trends are not available back to 1973 because only the overall average scores could be extrapolated for 1973.

Nine-year-olds. The trend lines shown in figure 2-5 appear very similar to one another at age 9. Nine-year-olds showed higher scores at each of the five selected percentiles in 2004 than in any other assessment year. Between the first year and the most recent assessment year—1978 and 2004—scores increased 26 points at the 10^{th} percentile, 26 points[1] at the 25^{th} percentile, 23 points at the 50^{th} percentile, 21 points[1] at the 75^{th} percentile, and 18 points at the 90^{th} percentile.

Thirteen-year-olds. At age 13, the score at each of the five percentile levels was higher in 2004 than in every previous assessment year, with the exception of the 10^{th} percentile. The score at the 10^{th} percentile in 2004 was higher than in 1978, but showed no measurable gain between 1999 and 2004.

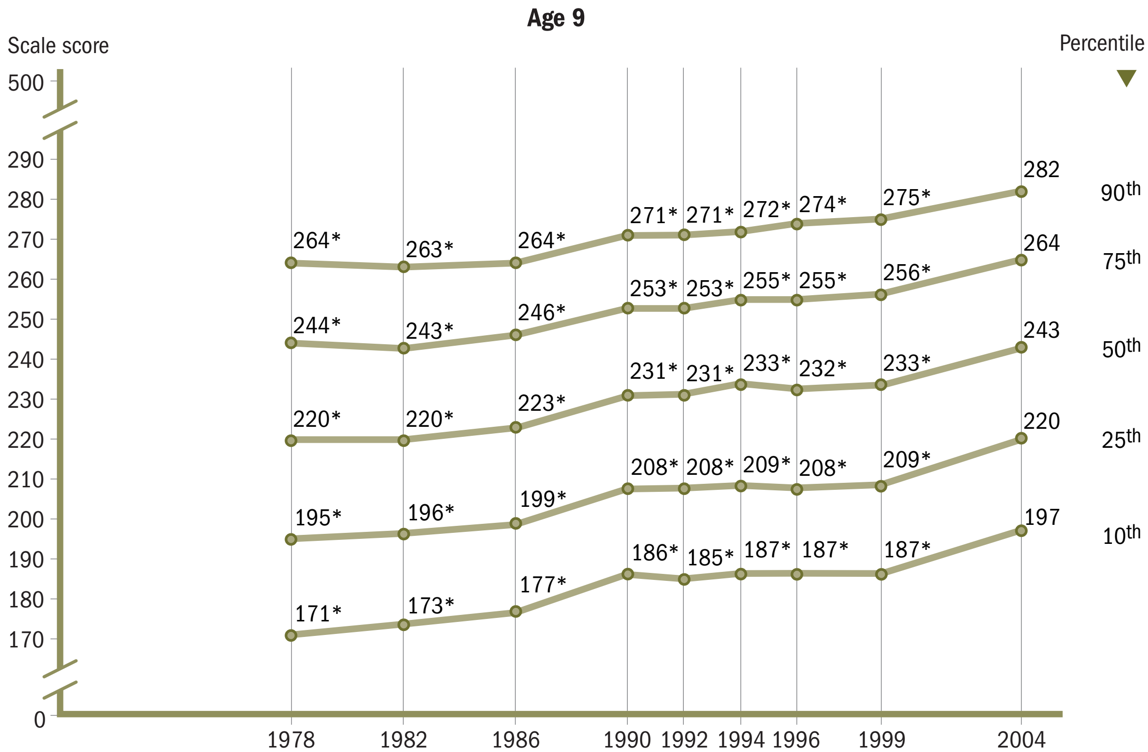
Seventeen-year-olds. Scores for 17-year-olds at the 10^{th}, 25^{th}, and 50^{th} percentiles were higher in 2004 than in 1978. The scores at the 75^{th} and 90^{th} percentiles were not measurably different in 2004 compared to 1999 or 1978.

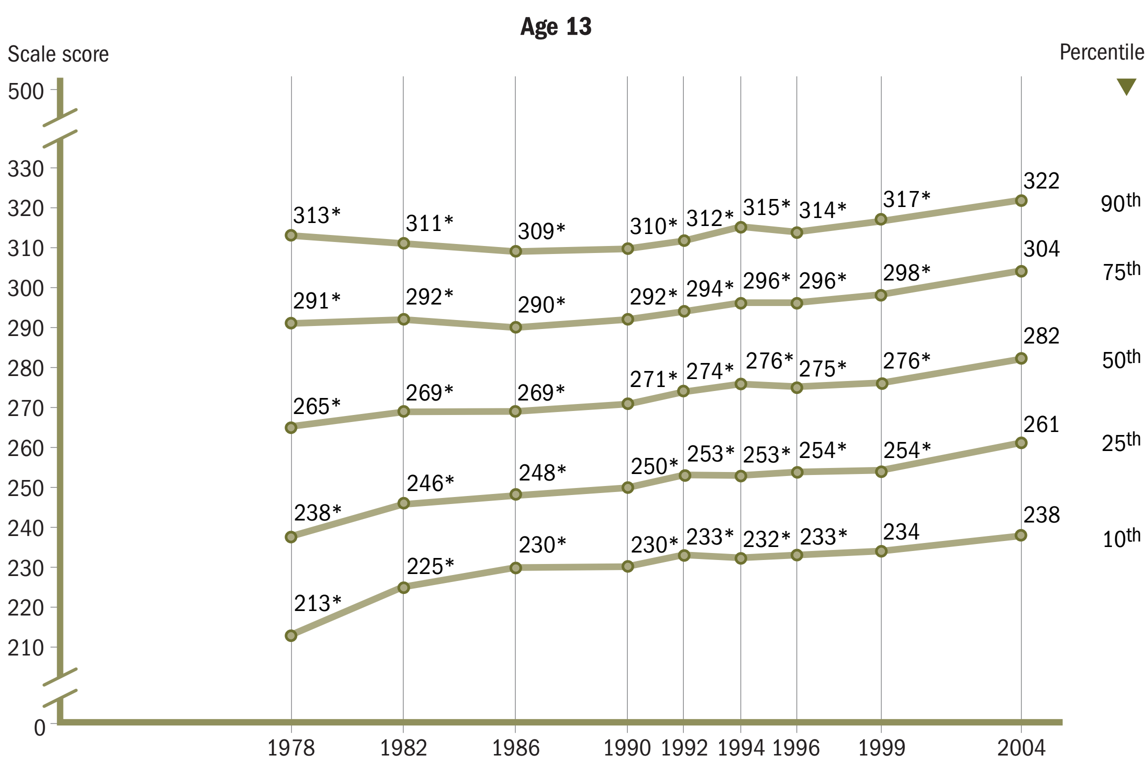
How to interpret this graphic . . .

Graphics like figure 2-5 show the score at each percentile for five selected percentiles. For example, at age 9 in 2004, students at the 10^{th} percentile scored 197 in mathematics, while students at the 90^{th} percentile scored 282. Both of these scores are higher than the scores in any previous assessment year. Looking at the five trend lines together, it can be determined if more improvement took place at the upper end or at the lower end, or if the trend lines look the same at all five levels.

[1] Detail may not sum to totals because of rounding. Differences between scores are calculated using unrounded values. In this instance, the result of the subtraction differs from what would be obtained by subtracting the rounded values shown in the accompanying figure.

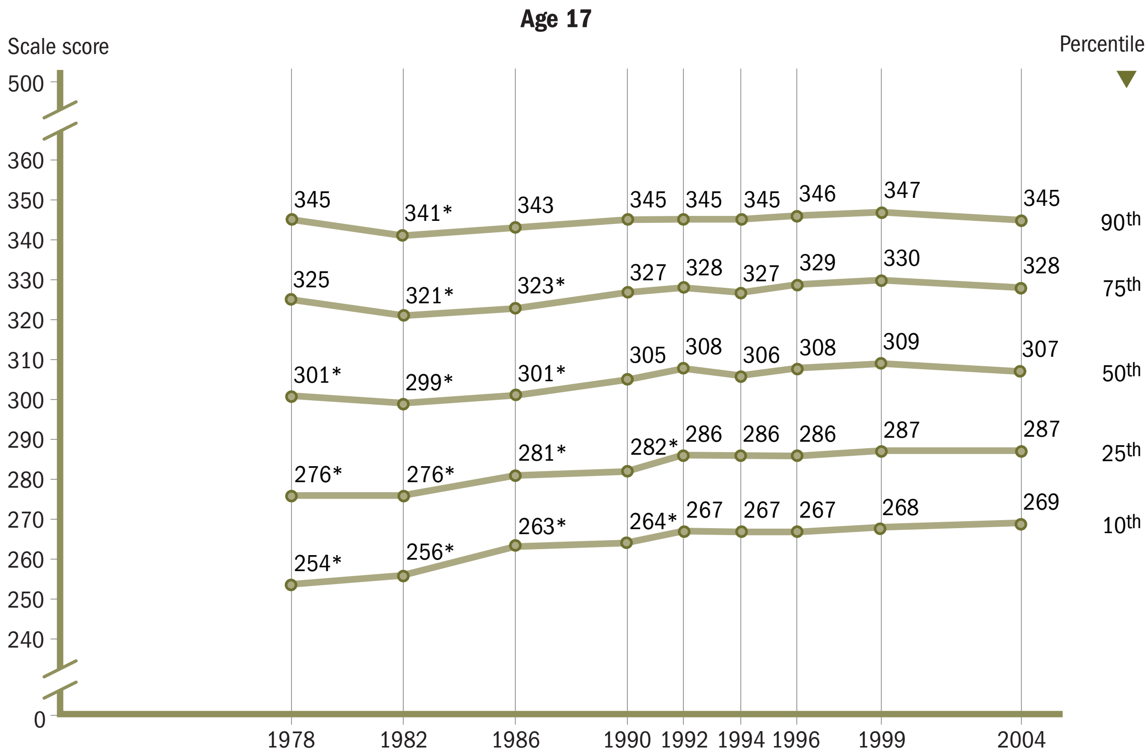
Figure 2-5. Trends in mathematics scale score at selected percentiles for students ages 9, 13, and 17: 1978–2004





See notes at end of figure.

Figure 2-5. Trends in mathematics scale score at selected percentiles for students ages 9, 13, and 17: 1978–2004—Continued



*Significantly different from 2004.

NOTE: Mathematics scores at selected percentiles are not available in 1973 because only the overall average scores were extrapolated for this year.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1978–2004 Long-Term Trend Mathematics Assessments.

Performance Levels

The skills and abilities demonstrated by students at each mathematics performance level are described below. As in reading, the five performance levels are applicable at all three ages, but only three performance levels are discussed for each age: levels 150, 200, and 250 for age 9; levels 200, 250, and 300 for age 13; and levels 250, 300, and 350 for age 17. These performance levels are the ones most likely to show significant change within an age across the assessment years and do not include the levels that nearly all or almost no students attained at a particular age in each year. Again, these trends are only available from 1978, because only the overall average scores could be extrapolated for 1973.

Mathematics Performance-Level Descriptions

LEVEL 350: *Multistep Problem Solving and Algebra*
Students at this level can apply a range of reasoning skills to solve multistep problems. They can solve routine problems involving fractions and percents, recognize properties of basic geometric figures, and work with exponents and square roots. They can solve a variety of two-step problems using variables, identify equivalent algebraic expressions, and solve linear equations and inequalities. They are developing an understanding of functions and coordinate systems.

LEVEL 300: *Moderately Complex Procedures and Reasoning*
Students at this level are developing an understanding of number systems. They can compute with decimals, simple fractions, and commonly encountered percents. They can identify geometric figures, measure lengths and angles, and calculate areas of rectangles. These students are also able to interpret simple inequalities, evaluate formulas, and solve simple linear equations. They can find averages, make decisions based on information drawn from graphs, and use logical reasoning to solve problems. They are developing the skills to operate with signed numbers, exponents, and square roots.

LEVEL 250: *Numerical Operations and Beginning Problem Solving*
Students at this level have an initial understanding of the four basic operations. They are able to apply whole number addition and subtraction skills to one-step word problems and money situations. In multiplication, they can find the product of a two-digit and a one-digit number. They can also compare information from graphs and charts and are developing an ability to analyze simple logical relations.

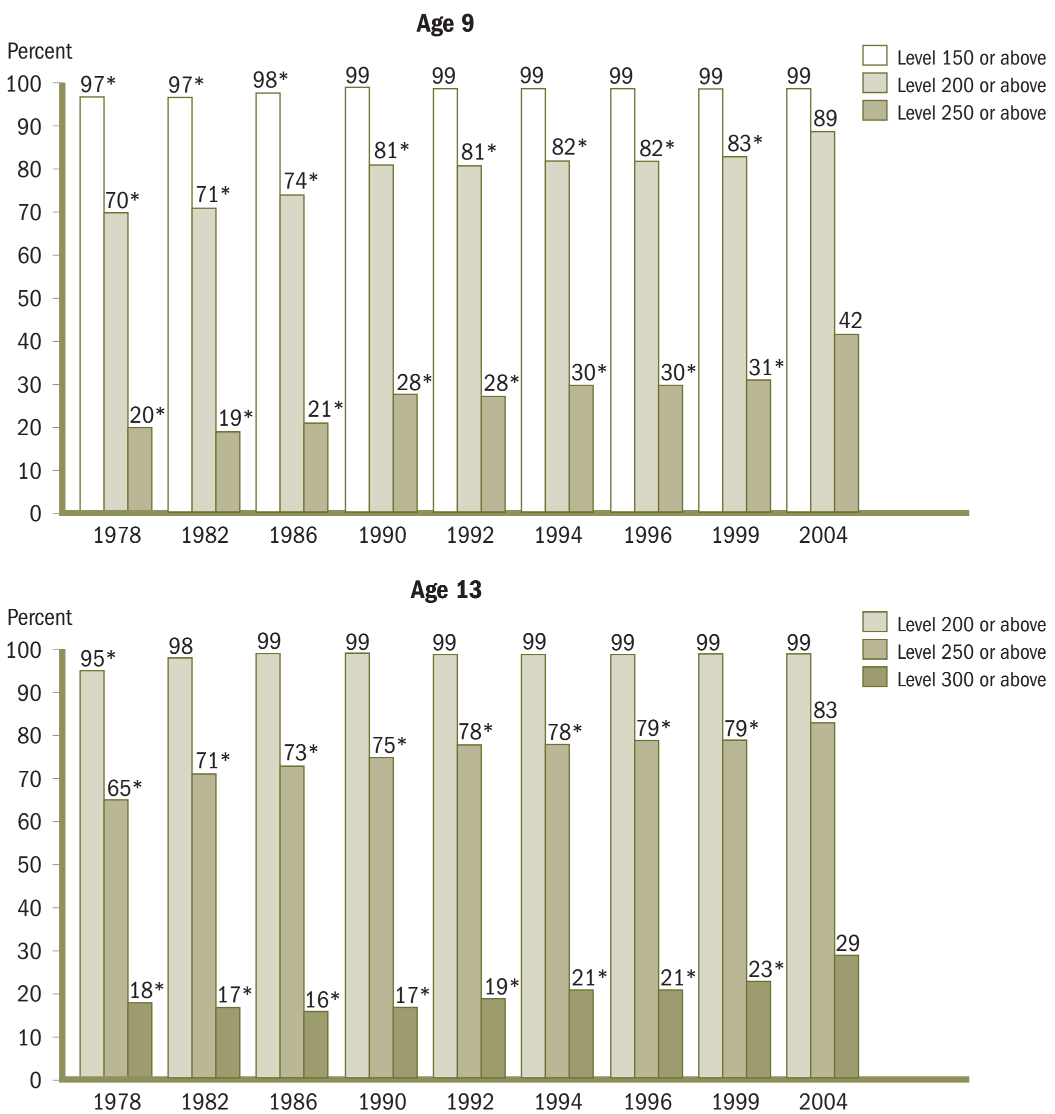
LEVEL 200: *Beginning Skills and Understandings*
Students at this level have considerable understanding of two-digit numbers. They can add two-digit numbers but are still developing an ability to regroup in subtraction. They know some basic multiplication and division facts, recognize relations among coins, can read information from charts and graphs, and use simple measurement instruments. They are developing some reasoning skills.

LEVEL 150: *Simple Arithmetic Facts*
Students at this level know some basic addition and subtraction facts, and most can add two-digit numbers without regrouping. They recognize simple situations in which addition and subtraction apply. They also are developing rudimentary classification skills.

Figure 2-6 shows the percentage of students reaching each performance level by age and assessment year. The following sections discuss the data for each age group.

Nine-year-olds. Trends in the percentage of 9-year-olds attaining mathematics performance levels 150, 200, and 250 are displayed in the upper panel of figure 2-6. In each assessment year, nearly all 9-year-olds (at least 97 percent) demonstrated understanding of simple arithmetic facts associated with level 150. In 2004, this percentage was 99, measurably higher by one percentage point than in 1986, and higher by three points[2] than in 1978, with no measurable change since 1990. The beginning skills and understandings characteristic of level 200 was demonstrated by 89 percent of 9-year-olds in 2004, higher than in any other assessment year. In the 2004 assessment, 42 percent of 9-year-olds performed the numerical operations and beginning problem solving associated with level 250, a higher percentage than in any other assessment year. There was an increase of 11 percentage points for 9-year-olds at this level between 1999 and 2004.

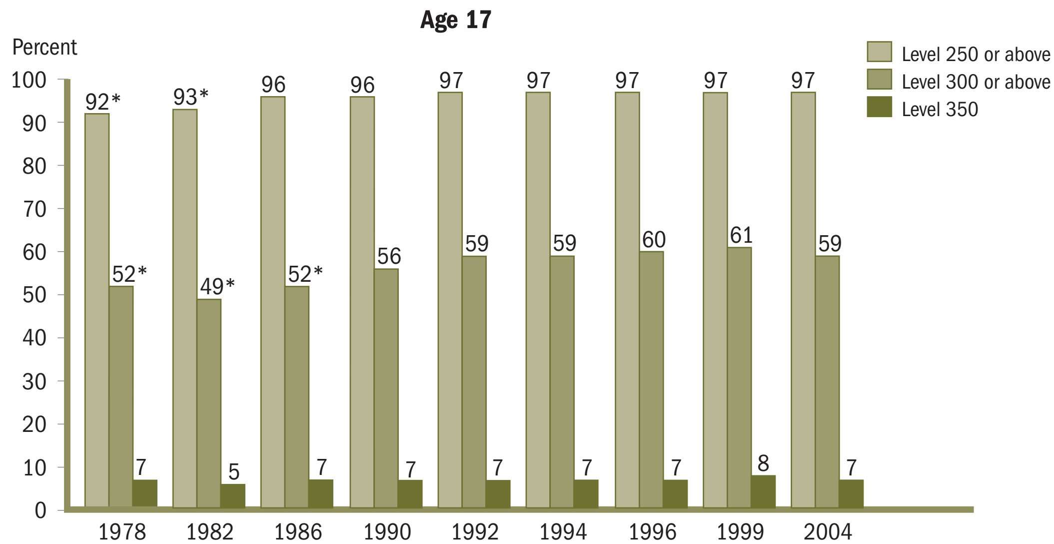
Figure 2-6. Trends in percentages at or above mathematics performance levels for students ages 9, 13, and 17: 1978–2004



See notes at end of figure.

2 Detail may not sum to totals because of rounding. Differences between percentages are calculated using unrounded values. In this instance, the result of the subtraction differs from what would be obtained by subtracting the rounded values shown in the accompanying figure.

Figure 2-6. Trends in percentages at or above mathematics performance levels for students ages 9, 13, and 17: 1978–2004—Continued



*Significantly different from 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1978–2004 Long-Term Trend Mathematics Assessments.

Thirteen-year-olds. The percentage of 13-year-old students scoring at or above mathematics performance levels 200, 250, and 300 across the assessment years are displayed in the middle panel of figure 2-6. Since 1986, 99 percent of 13-year-olds demonstrated the beginning skills and understandings associated with level 200. In 2004, 83 percent scored at or above level 250, demonstrating the ability to perform numerical operations and beginning problem solving. Overall gains are also evident at level 300, where students performed moderately complex procedures and reasoning. The percentage of students who scored at or above this level increased from 18 percent in 1978, to 23 percent in 1999, and to 29 percent in 2004.

Seventeen-year-olds. Trends in the percentage of 17-year-olds scoring at or above mathematics performance levels 250, 300, and 350 are displayed in the last panel of figure 2-6. Since 1986, at least 96 percent of 17-year-olds have performed at or above level 250, demonstrating the ability to perform numerical operations and beginning problem solving. The percentage of 17-year-olds who performed moderately complex procedures and reasoning (level 300) showed no measurable change from 1990 to 2004, but has increased by 7 percentage points from 1978. No measurable change between 2004 and all the previous assessment years can be detected at 350, the highest performance level, in which students applied a range of reasoning skills to solve multistep problems. Across the assessment years, between 5 and 8 percent of students performed at this level.

Summary

The results presented in this chapter give an overall view of national trends in reading and mathematics achievement. Average scores for the nation, scores for students in five different ranges of the performance distribution, and attainment of specific performance levels were discussed. Looking across the 33 years, upward trends are most noticeable at age 9 in both reading and mathematics. Also of interest is the increase in performance at age 13 in mathematics.

The following figures provide an overview of the major findings presented in this chapter by comparing students' performance in 2004 to that of their counterparts in the first year data were collected. In addition, 2004 and 1999 results are compared, providing a summary of trends over the last five years.

Arrows pointing upward (↑) indicate improvement, and horizontal arrows (→) indicate no measurable change in performance. For example, the first line of the display in figure 2-7 indicates that the national average reading score for 9-year-olds was higher in 2004 than it was in 1971 or 1999.

Figure 2-7. Summary of trends in reading and mathematics average scale scores for students ages 9, 13, and 17: 1971–2004

Reading

↑ 9-year-olds' average scale scores since 1971 (↑ since 1999)
↑ 13-year-olds' average scale scores since 1971 (→ since 1999)
→ 17-year-olds' average scale scores since 1971 (→ since 1999)

Mathematics

↑ 9-year-olds' average scale scores since 1973 (↑ since 1999)
↑ 13-year-olds' average scale scores since 1973 (↑ since 1999)
→ 17-year-olds' average scale scores since 1973 (→ since 1999)

↑ Significantly higher in 2004.
→ Indicates no significant difference between earlier year and 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading and Mathematics Assessments.

Figure 2-8. Summary of trends in reading and mathematics scale score percentiles for students ages 9, 13, and 17: 1971–2004

Reading

9-year-olds
↑ 10th percentile since 1971 (↑ since 1999)
↑ 25th percentile since 1971 (↑ since 1999)
↑ 50th percentile since 1971 (↑ since 1999)
↑ 75th percentile since 1971 (↑ since 1999)
→ 90th percentile since 1971 (→ since 1999)

13-year-olds
→ 10th percentile since 1971 (→ since 1999)
→ 25th percentile since 1971 (→ since 1999)
→ 50th percentile since 1971 (→ since 1999)
↑ 75th percentile since 1971 (→ since 1999)
↑ 90th percentile since 1971 (→ since 1999)

17-year-olds
→ 10th percentile since 1971 (→ since 1999)
→ 25th percentile since 1971 (→ since 1999)
→ 50th percentile since 1971 (→ since 1999)
→ 75th percentile since 1971 (→ since 1999)
→ 90th percentile since 1971 (→ since 1999)

Mathematics

9-year-olds
↑ 10th percentile since 1978 (↑ since 1999)
↑ 25th percentile since 1978 (↑ since 1999)
↑ 50th percentile since 1978 (↑ since 1999)
↑ 75th percentile since 1978 (↑ since 1999)
↑ 90th percentile since 1978 (↑ since 1999)

13-year-olds
↑ 10th percentile since 1978 (→ since 1999)
↑ 25th percentile since 1978 (↑ since 1999)
↑ 50th percentile since 1978 (↑ since 1999)
↑ 75th percentile since 1978 (↑ since 1999)
↑ 90th percentile since 1978 (↑ since 1999)

17-year-olds
↑ 10th percentile since 1978 (→ since 1999)
↑ 25th percentile since 1978 (→ since 1999)
↑ 50th percentile since 1978 (→ since 1999)
→ 75th percentile since 1978 (→ since 1999)
→ 90th percentile since 1978 (→ since 1999)

↑ Significantly higher in 2004.
→ Indicates no significant difference between earlier year and 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading and Mathematics Assessments.

Figure 2-9. Summary of trends in reading and mathematics percentages at or above performance levels for students ages 9, 13, and 17: 1971–2004

Reading	
9-year-olds	
↑ Level 150 (simple, discrete reading tasks) since 1971	(↑ since 1999)
↑ Level 200 (partially developed skills and understanding) since 1971	(↑ since 1999)
↑ Level 250 (interrelate ideas and make generalizations) since 1971	(↑ since 1999)
13-year-olds	
→ Level 200 (partially developed skills and understanding) since 1971	(→ since 1999)
→ Level 250 (interrelate ideas and make generalizations) since 1971	(→ since 1999)
↑ Level 300 (understand complicated information) since 1971	(→ since 1999)
17-year-olds	
→ Level 250 (interrelate ideas and make generalizations) since 1971	(→ since 1999)
→ Level 300 (understand complicated information) since 1971	(→ since 1999)
→ Level 350 (learn from specialized reading materials) since 1971	(→ since 1999)
Mathematics	
9-year-olds	
↑ Level 150 (simple arithmetic facts) since 1978	(→ since 1999)
↑ Level 200 (beginning skills and understandings) since 1978	(↑ since 1999)
↑ Level 250 (numerical operations and beginning problem solving) since 1978	(↑ since 1999)
13-year-olds	
↑ Level 200 (beginning skills and understandings) since 1978	(→ since 1999)
↑ Level 250 (numerical operations and beginning problem solving) since 1978	(↑ since 1999)
↑ Level 300 (moderately complex procedures and reasoning) since 1978	(↑ since 1999)
17-year-olds	
↑ Level 250 (numerical operations and beginning problem solving) since 1978	(→ since 1999)
↑ Level 300 (moderately complex procedures and reasoning) since 1978	(→ since 1999)
→ Level 350 (multistep problem solving and algebra) since 1978	(→ since 1999)

↑ Significantly higher in 2004.
→ Indicates no significant difference between earlier year and 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading and Mathematics Assessments.

THIS PAGE INTENTIONALLY LEFT BLANK.