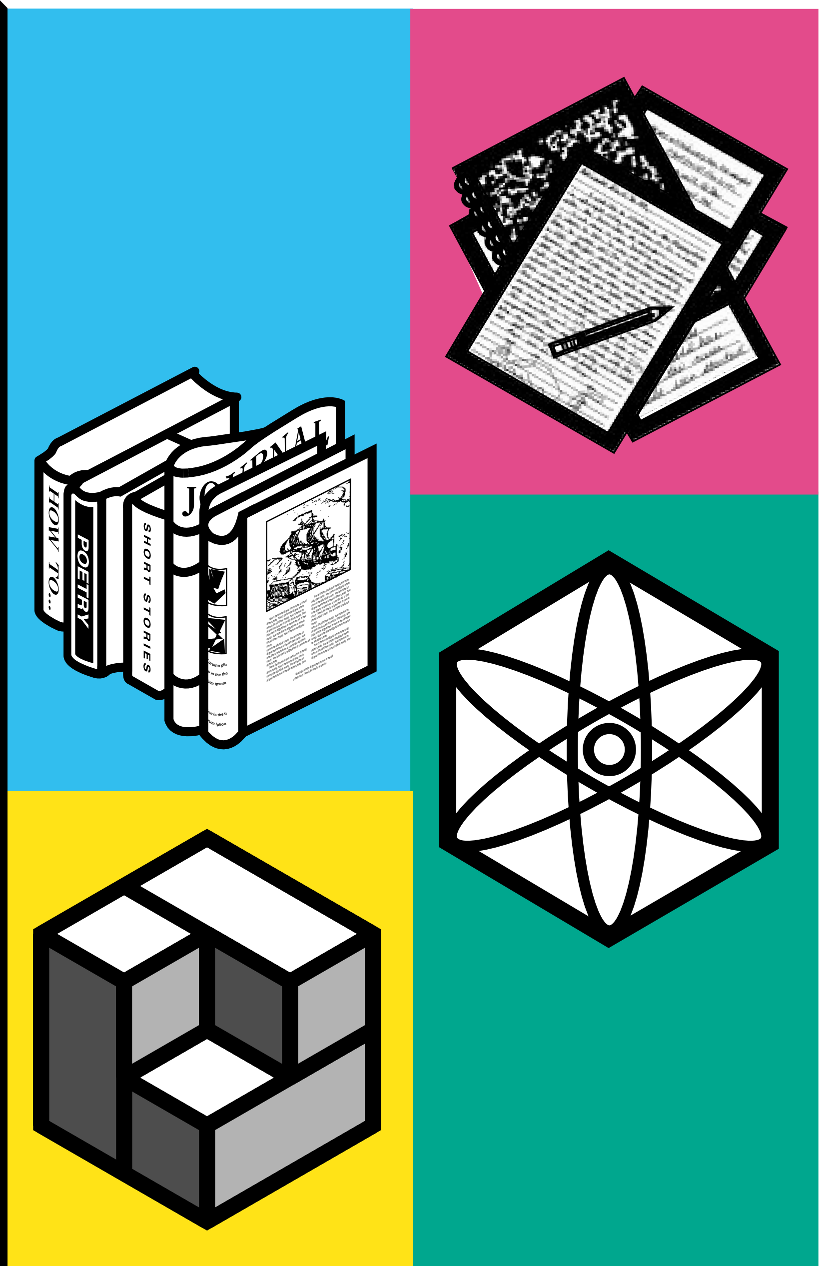
NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP 1994 Trends in Academic Progress

Achievement of U.S. Students in
Science, 1969 to 1994 • Mathematics, 1973 to 1994
Reading, 1971 to 1994 • Writing, 1984 to 1994



What is The Nation's Report Card?

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

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

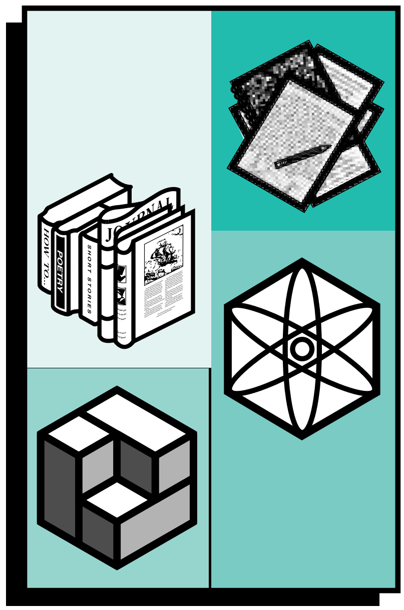
In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels;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

The National Assessment Governing Board

NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP 1994 Trends in Academic Progress

Achievement of U.S. Students in
Science, 1969 to 1994 • Mathematics, 1973 to 1994
Reading, 1971 to 1994 • Writing, 1984 to 1994



Jay R. Campbell
Clyde M. Reese
Christine O'Sullivan
John A. Dossey

in collaboration with

Patricia L. Donahue, Peggy Carr,
Brent Sandene, Claudia Gentile,
Karen Miller, and John Mazzeo

November 1996

THE NATION'S REPORT CARD

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

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

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr., Ph.D.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

FOR MORE INFORMATION:

For ordering information on this report, write:

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

or call 1-800-424-1616 (in the Washington, D.C. metropolitan area call 202-219-1651).

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

CONTENTS

PART I
Trends in Science Achievement From 1969-70 to 1994

CONTENTS

CONTENTS

CONTENTS

PART III
Trends in Reading Achievement From 1971 to 1994

CONTENTS

CONTENTS

PART IV
Trends in Writing Achievement From 1984 to 1994

CONTENTS

CONTENTS

PROCEDURAL APPENDIX

CONTENTS

DATA APPENDIX

ACKNOWLEDGMENTS

Executive Summary

Introduction

Educational reform continues to be a major concern of parents, educators, and policy makers, as well as the general public. Reorganizing schools, enhancing the curriculum, establishing performance standards, and rethinking traditional instructional methods are just some of the efforts being made across the country to increase student achievement. As a part of these efforts, in 1990 the President and governors adopted a set of six ambitious national education goals for the 21st century: ensuring that children start school ready to learn, raising high school graduation rates, increasing levels of educational achievement, promoting science and mathematics achievement as well as literacy and lifelong learning, and freeing schools of drugs and violence.[1] In the Spring of 1994, Congress broadened the goals to include improvements in teacher preparation and increased parental involvement in schools.[2]

Measuring students' progress toward higher achievement has been the purpose of the National Assessment of Educational Progress (NAEP) since its inception in 1969. Students in both public and nonpublic schools have been assessed in various subject areas on a regular basis. In addition, NAEP

---

[1] Executive Office of the President, *National Goals for Education.* (Washington, DC: Government Printing Office, 1990).

[2] Goals 2000: Educate America Act, Pub. L. No. 103-227 (1994).

collects information about relevant background variables that provide an important context for interpreting the assessment results and for documenting the extent to which educational reform has been implemented.

One important feature of NAEP is its ability to monitor trends in academic achievement in core curriculum areas over an extended period of time. By readministering materials and replicating procedures from assessment to assessment, NAEP provides valuable information about progress in academic achievement and about whether the United States can meet the challenge of its national education goals.

The NAEP long-term trend assessments are separate from the main assessments conducted by NAEP that involve more recently developed instruments. While the long-term trend assessments have used the same sets of questions and tasks so that trends across time can be measured, the main assessments in each subject area have been developed to reflect current educational priorities. The use of both long-term trend and main assessments allows NAEP to provide information about students' achievement over time, and assess their achievement of more contemporary educational objectives. As each of these assessments is based on different sets of questions or tasks, the results from each cannot be directly compared.

This report presents results of the NAEP 1994 trend assessments in science, mathematics, reading, and writing. To provide a numeric summary of students' performance on the assessment questions and tasks, NAEP uses a 0 to 500 scale for each subject area. Comparisons of average scale scores are provided across the years in which trend assessments have been administered and among subpopulations of students. National representative samples totaling approximately 31,000 students were involved in the NAEP 1994 trend assessments.

In the following sections of this report, trend assessment results are given in science, mathematics, reading and writing. These results chart trends going back to the first year in which each NAEP assessment was given: 1969/70 in science; 1973 in mathematics; 1971 in reading; and 1984 in writing. Trends in average performance over these time periods are discussed for students at ages 9, 13, and 17 for the science, mathematics, and reading assessments and for grades 4, 8, and 11 for the writing assessment. Trends in average performance differences between White students and Black students, White students and Hispanic students, and male students and female students are also discussed.

The descriptions of trend results are based on the results of statistical tests that consider both the estimates of average performance in each assessment year as well as the degree of uncertainty associated with these estimates. The purpose of basing descriptions on such tests is to restrict the discussion of observed trends and group differences to those that are statistically dependable. Hence, the patterns of results that are discussed are unlikely to be due to the chance factors associated with the inevitable sampling and measurement errors inherent in any large-scale survey effort like NAEP. Throughout this report, all descriptions of trend patterns, differences between assessment years, and differences between subgroups of students which are cited are statistically significant at the .05 level.

Two distinct sets of statistical tests have been applied to the trend results. First, each sequence of assessment results (whether it be overall performance or differences in performance for race/ethnicity and gender subgroups) was tested for linear and quadratic trends. Separate tests were carried out in each subject area at each age or grade level. The purpose of this first set of general tests was to determine whether the results of the series of assessments in a given subject could be generally characterized by a line or a simple curve. A linear relationship indicates that results have steadily increased (or decreased) at a relatively constant rate over the time period of interest. Simple curvilinear (i.e., quadratic) relationships capture more complex patterns. For example, one possible pattern is to have initial score declines over part of the time period followed by score increases in more recent assessments. Another possible pattern is to have a sequence of several assessments in which scores have increased followed by a period of relative stable assessment performance. These examples are two, but not all, of the simple curvilinear relationships that were tested.

Simple linear and curvilinear patterns do not always provide a satisfactory summary description of the pattern of trend results. Hence, tests of linear and quadratic trends were supplemented by a second set of statistical tests which compared results for selected pairs of assessment years within each trend sequence. Again, separate tests were carried out in each subject area at each age or grade level. Two families of pairwise tests were carried out. One family of tests consisted of comparing the results from each trend assessment year to the results for the first assessment year. The second family of tests consisted of comparing the results from each trend assessment year to the 1994 results. The statistical tests in both families were carried out at a significance level that adjusted for the multiple comparisons being carried out within each family. The characterizations of trend data that appear below are based on the combined results of both the general (i.e., linear and quadratic) and the two families of pairwise tests.

Overall Trends in Average Scale Scores

The overall trends in science, mathematics, reading, and writing achievement are presented in Figure 1. In general, the trends in science and mathematics show early declines or relative stability followed by improved performance, but reading and writing results show few indications of positive trends.[3]

Science. Students in all three age groups demonstrated declines in science performance in the 1970s, but have since improved. In 1994, the improvements resulted in an average score for 9-year-olds that was higher than that in 1970. However, the 1994 average score for 13-year-olds was not significantly different than that in 1970, and the 1994 average score for 17-year-olds was lower than the 1969 average.

Mathematics. The overall picture of mathematics achievement provided by the long-term trend results is one of early declines or relative stability followed by a pattern of increased performance. For 9-year-olds, average scale scores began to increase with the 1990 assessment and were higher in 1994 than in 1973. For 13-year-olds, average scores began to increase with the 1982 assessment, resulting in a 1994 average score higher than the average score in 1973. For 17-year-olds, after a period of decline from 1973 to 1982, the average score increased to a level in 1994 that did not differ significantly from that in 1973.

Reading. The overall picture of trends in reading achievement is one of only minimal changes across the assessment years. At age 9, higher levels of performance that were observed in the 1980s have not been maintained, and in 1994 the average score returned to a level not significantly different from that in 1971. The average score of 13-year-olds increased in 1992 to a level that was higher than in 1971. However, the 1994 average score did not differ significantly from either the 1971 or 1992 averages. Seventeen-year-olds had average scores from 1988 to 1992 that were higher than the 1971 average. However, the 1994 average score was not significantly different from that in either 1971 or 1992.

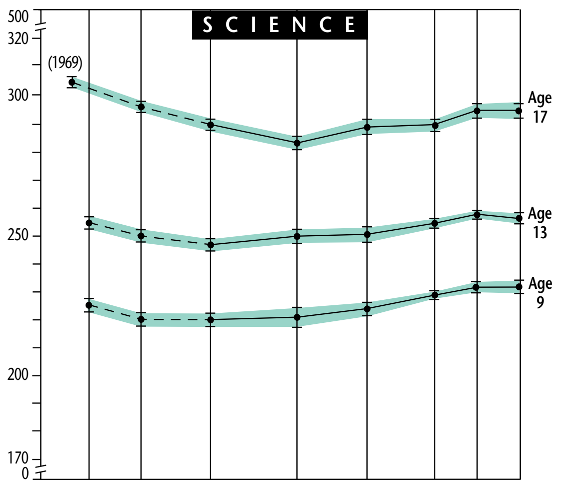
Writing. Across the three grades assessed in writing, mixed results were observed. Fourth graders had relatively stable performance across the assessment years. At the eighth grade, a decline observed in 1990 was reversed in 1992. However, the average score decreased from 1992 to 1994, resulting in an average score that was not significantly different from that in 1984. The average score of eleventh graders decreased from 1984 to 1994.

---

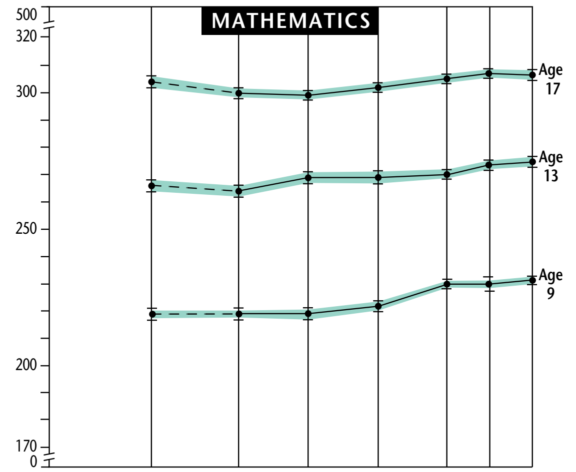
[3] The examination of overall trends includes discussions of statistically significant linear and quadratic trends. For science, linear (positive for ages 9 and 13, and negative for age 17) and quadratic (positive for all three ages) trends were noted. For mathematics, positive linear and quadratic trends were noted for all three ages. For reading, positive linear trends were noted for ages 13 and 17, and a negative quadratic trend was noted for age 9. For writing, a negative linear trend was noted for grade 11 and a positive quadratic trend was noted for grade 8.

Figure 1

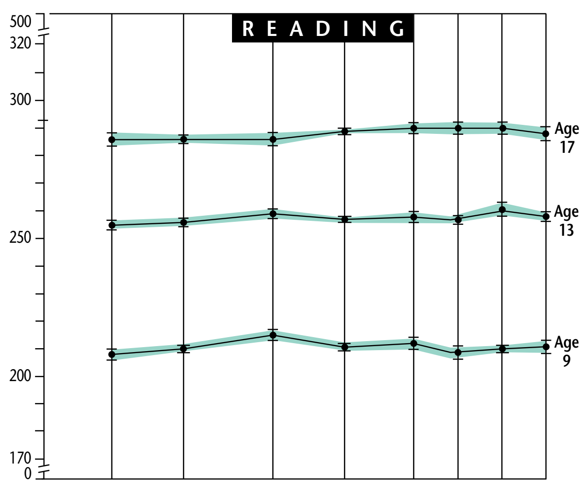
National Trends in Average Scale Scores in Science, Mathematics, Reading, and Writing



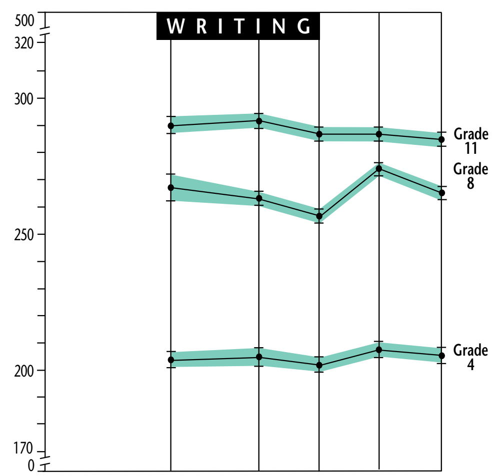
	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	305(1.0)*	296(1.0)†	290(1.0)†	283(1.2)*†	289(1.4)†	290(1.1)†	294(1.3)†	294(1.6)†
Age 13	255(1.1)	250(1.1)*†	247(1.1)*†	250(1.3)*†	251(1.4)*	255(0.9)	258(0.8)	257(1.0)
Age 9	225(1.2)*	220(1.2)*†	220(1.2)*†	221(1.8)*	224(1.2)*	229(0.8)	231(1.0)†	231(1.2)†



	1973	1978	1982	1986	1990	1992	1994
Age 17	304(1.1)	300(1.0)*	299(0.9)*†	302(0.9)*	305(0.9)	307(0.9)	306(1.0)
Age 13	266(1.1)*	264(1.1)*	269(1.1)*	269(1.2)*	270(0.9)*†	273(0.9)†	274(1.0)†
Age 9	219(0.8)*	219(0.8)*	219(1.1)*	222(1.0)*	230(0.8)†	230(0.8)†	231(0.8)†



	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	285(1.2)	286(0.8)	286(1.2)	289(0.6)	290(1.0)†	290(1.1)†	290(1.1)†	288(1.3)
Age 13	255(0.9)	256(0.8)	259(0.9)	257(0.5)	258(1.0)	257(0.8)	260(1.2)†	258(0.9)
Age 9	208(1.0)	210(0.7)	215(1.0)†	211(0.7)†	212(1.1)†	209(1.2)	211(0.9)	211(1.2)



	1984	1988	1990	1992	1994
Grade 11	290(1.6)*	291(1.3)*	287(1.0)	287(1.4)	285(1.2)†
Grade 8	267(2.0)	264(1.3)	257(1.2)*†	274(1.3)*†	265(1.3)
Grade 4	204(1.5)	206(1.6)	202(1.5)	207(1.5)	205(1.6)

95 percent confidence interval. [– – –] Extrapolated from previous NAEP analyses.

* Statistically significant difference from 1994 at 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 for science, 1973 for mathematics, 1971 for reading and 1984 for writing at 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Scale Scores by Race/Ethnicity and by Gender

Changes in average scale scores on the NAEP trend assessments in science, mathematics, reading, and writing are presented by race/ethnicity and gender in Table 1. Results are presented for 1994 and the first trend assessment in each subject area — 1969 or 1970 for science, 1973 for mathematics, 1971 for reading, and 1984 for writing.

Science. White 9- and 13-year-old students had higher average science scores in 1994 than in 1970. The average score of 17-year-old White students, however, was lower. Nine-year-old Black students and 13-year-old Hispanic students showed an increase in average scores between the first and most recent assessment.[4] At age 9, female students had a higher average score in 1994 than in 1970. At age 17, both male and female students, reflecting the national trend results, had lower scores in the most recent assessment.

Mathematics. In mathematics, White 9- and 13-year-olds had higher average scores in 1994 than in 1973. A higher average score in the most recent assessment was also observed for Black students in each age group, and for Hispanic students aged 13 and 17. At ages 9 and 13, both males and females had average mathematics scale scores in 1994 that were higher than those of their counterparts in 1973.

Reading. White 13-year-olds attained an average score in 1994 that was higher than that attained by their counterparts in 1971. At all three ages, Black students displayed a gain in average scale scores between 1971 and 1994. The average scores of male 9-year-olds and female 13-year-olds were higher in 1994 than the average scores earned by their counterparts in the first trend assessment.

Writing. Reflecting the overall performance of students on the trend writing assessments, no significant changes were observed between 1984 and 1994 in the average scores of fourth- and eighth-grade students by race/ethnicity and by gender. At the eleventh grade, the average score of White students in 1994 was lower than the average score of their counterparts in 1984.

[4] For Hispanic students, the science differences are calculated between 1977 and 1994.

Table 1

Trends in Average Scale Scores in Science, Mathematics, Reading, and Writing by Race/Ethnicity and Gender

	AGE 9		AGE 13		AGE 17	
	Average Scale Score					
SCIENCE	**1970†**	**1994**	**1970†**	**1994**	**1969†**	**1994**
Nation	225(1.2)*	231(1.2)	255(1.1)	257(1.0)	305(1.0)*	294(1.6)
White	236(0.9)*	240(1.3)	263(0.8)*	267(1.0)	312(0.8)*	306(1.5)
Black	179(1.9)*	201(1.7)	215(2.4)	224(4.2)	258(1.5)	257(3.1)
Hispanic	192(2.7)	201(2.7)	213(1.9)*	232(2.4)	262(2.2)	261(6.7)
Male	228(1.3)	232(1.3)	257(1.3)	259(1.2)	314(1.2)*	300(2.0)
Female	223(1.2)*	230(1.4)	253(1.2)	254(1.2)	297(1.1)*	289(1.7)
	Average Scale Score					
MATHEMATICS	**1973**	**1994**	**1973**	**1994**	**1973**	**1994**
Nation	219(0.8)*	231(0.8)	266(1.1)*	274(1.0)	304(1.1)	306(1.0)
White	225(1.0)*	237(1.0)	274(0.9)*	281(0.9)	310(1.1)	312(1.1)
Black	190(1.8)*	212(1.6)	228(1.9)*	252(3.5)	270(1.3)*	286(1.8)
Hispanic	202(2.4)	210(2.3)	239(2.2)*	256(1.9)	277(2.2)*	291(3.7)
Male	218(0.7)*	232(1.0)	265(1.3)*	276(1.3)	309(1.2)	309(1.4)
Female	220(1.1)*	230(0.9)	267(1.1)*	273(1.0)	301(1.1)	304(1.1)
	Average Scale Score					
READING	**1971†**	**1994**	**1971†**	**1994**	**1971†**	**1994**
Nation	208(1.0)	211(1.2)	255(0.9)	258(0.9)	285(1.2)	288(1.3)
White	214(0.9)	218(1.3)	261(0.7)*	265(1.1)	291(1.0)	296(1.5)
Black	170(1.7)*	185(2.3)	222(1.2)*	234(2.4)	239(1.7)*	266(3.9)
Hispanic	183(2.2)	186(3.9)	233(3.0)	235(1.9)	252(3.5)	263(4.9)
Male	201(1.1)*	207(1.3)	250(1.0)	251(1.2)	279(1.2)	282(2.2)
Female	214(1.0)	215(1.4)	261(0.9)*	266(1.2)	291(1.3)	295(1.5)
	GRADE 4		GRADE 8		GRADE 11	
	Average Scale Score					
WRITING	**1984**	**1994**	**1984**	**1994**	**1984**	**1994**
Nation	204(1.5)	205(1.6)	267(2.0)	265(1.3)	290(1.6)*	285(1.2)
White	211(1.9)	214(1.5)	272(2.1)	272(1.4)	297(1.8)*	291(1.4)
Black	182(5.0)	173(3.2)	247(5.7)	245(3.4)	270(3.6)	267(2.2)
Hispanic	189(5.8)	189(3.1)	247(6.4)	252(3.3)	259(6.6)	271(4.0)
Male	201(2.8)	196(1.7)	258(2.3)	254(1.8)	281(1.4)	276(1.5)
Female	208(3.1)	214(2.2)	276(2.4)	278(1.4)	299(2.5)	293(1.5)

†NOTE: For Hispanic students, the science differences are calculated between 1977 and 1994, and the reading differences are calculated between 1975 and 1994.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The set of comparisons include all intervening assessment years. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Differences in Average Scale Scores Between Racial/Ethnic Groups of Students and Between Males and Females

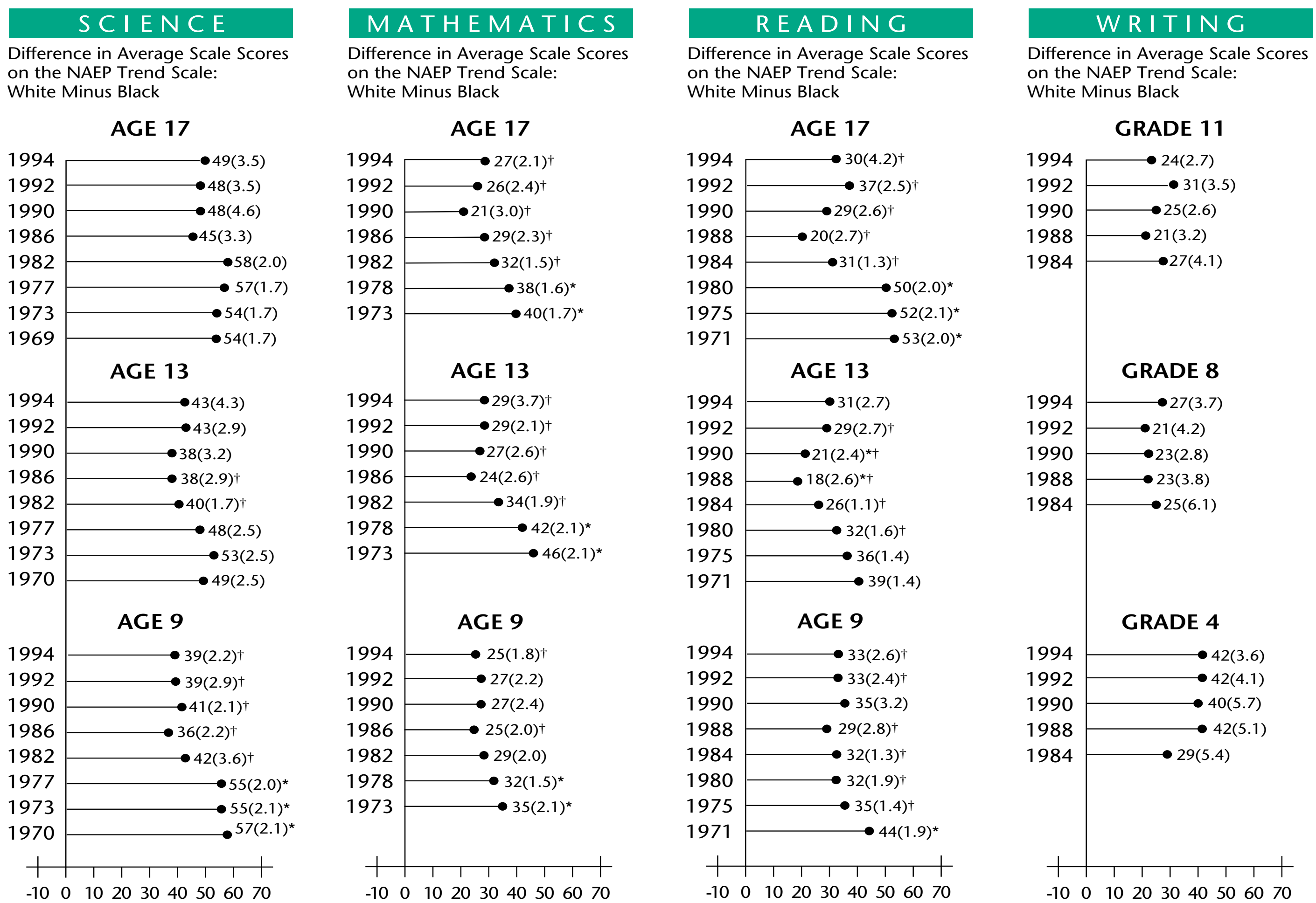
As noted earlier, one of the national education goals emphasizes increases in students' academic achievement.[5] A stated objective of this goal is that the performance distribution for minority students will more closely reflect that of the student population as a whole. In some subject areas, results indicated progress toward meeting this goal. Trends in the differences between average scores for subgroups of students are reported in this section.

Differences between White and Black Students. Trend analyses of the differences between White and Black students' average scale scores at each age revealed an overall pattern of narrowing gaps in mathematics and reading, resulting in a smaller gap between White and Black students in 1994 than in the first trend assessment. (The 1971 and 1994 gaps for 13-year-olds in reading were not significantly different.) However, there was some indication that the gap has increased since the late 1980s for 13-year-olds in mathematics and for all ages in reading (see Figure 2). The gap between White and Black students in science scores has decreased for 9- and 17-year-olds. However, for 17-year-olds the gap was not significantly different in 1994 than in 1969. At age 13, the gap between White and Black students in science scores decreased slightly until 1986. Since that time it returned to a level not significantly different from that in 1970. The gap between White and Black students' average writing scores has remained relatively stable since 1984 at each grade. Despite a narrowing of the gap between the average performance of White and Black students in three of the four subject areas assessed, in 1994 White students at all three ages or grades had average scores in each subject area that were higher than the average scores of Black students.

[5] Executive Office of the President, *National Goals for Education.* (Washington, DC: Government Printing Office, 1990).

Figure 2

Trends in Differences in Average Scale Scores of White and Black Students Across Subject Areas



* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

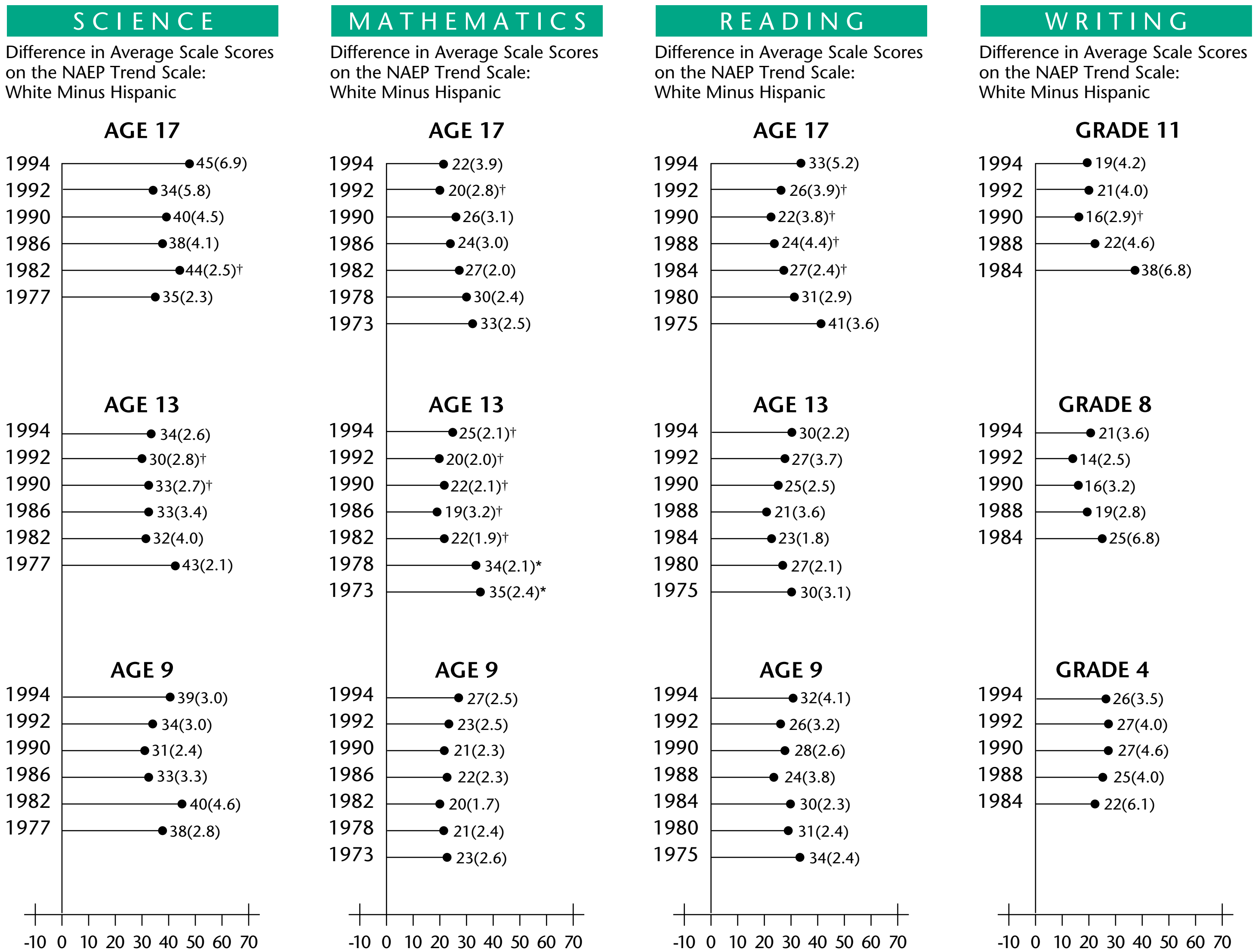
† Statistically significant difference from the initial assessment year in each subject. The standard errors of the estimated differences in scale scores appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Differences between White and Hispanic Students. Trend analyses of the differences between White and Hispanic students' average scale scores revealed an overall pattern of narrowing gaps for 17-year-olds in mathematics and reading, and for eleventh graders in writing. A similar pattern was observed for 13-year-olds in science and mathematics (see Figure 3). However, there was some indication that the trend of narrowing gaps reversed during recent assessments for 17-year-olds in reading and for 13-year-olds in science and mathematics. When differences between White and Hispanic students in the first assessment were compared to those in the most recent assessment, only one instance of a statistically significant change was revealed: the difference between White and Hispanic 13-year-olds' mathematics scores in 1994 was smaller than that observed in the first assessment. In 1994 the average scores of White students at all three ages or grades were higher than the average scores of Hispanic students in all subject areas.

Figure 3

Trends in Differences in Average Scale Scores of White and Hispanic Students Across Subject Areas



* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from the initial assessment year in each subject. The standard errors of the estimated differences in scale scores appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors for the sample. In comparing two estimates, one must use the standard error of the difference.

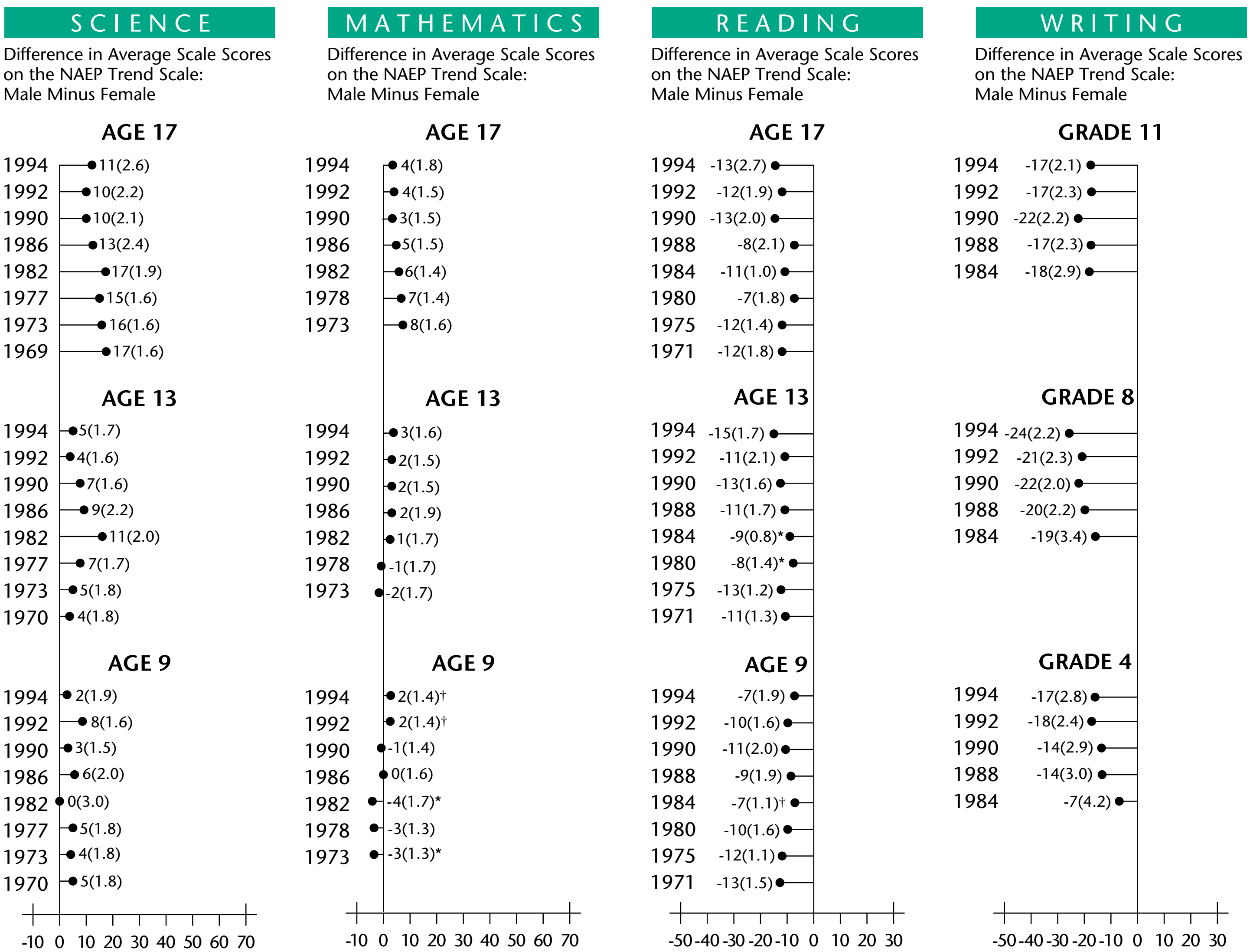
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Differences between Males and Females. Trend analyses of the differences between male and female students' average scale scores revealed an overall pattern of narrowing gaps for 17-year-olds in science and mathematics and for 9-year-olds in reading (see Figure 4). In science and mathematics, average scores were higher for males than for females, but in reading, females had higher average scores than males. The difference between fourth-grade male and female students' average writing scores increased across the assessment years, widening the gap in which females outperformed males. Trend analyses of the mathematics score gaps between male and female students aged 9 and 13 revealed a shift across time. At both ages, the trend was away from females outperforming males and toward either no difference or males outperforming females.

In 1994, the average science and mathematics scores of 13- and 17-year-old male students was higher than the average scores of their female counterparts. No significant difference between 9-year-old males and females in 1994 was observed in either science or mathematics. Across all three ages or grades in 1994, female students had higher average reading and writing scores than their male peers.

Figure 4

Trends in Differences in Average Scale Scores of Male and Female Students Across Subject Areas



* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from the initial assessment year in each subject. The standard errors of the estimated differences in scale scores appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Levels of Performance

To permit a closer look at the achievement of students in age groups or grades, five levels of performance have been identified and described along the NAEP scale for each subject area — 150, 200, 250, 300, and 350.[6] The procedure for describing the five performance levels was the same in science, mathematics, and reading. Sets of questions more likely to be answered correctly by students at one level, than at the next lower level, were identified. Educators and curriculum experts representing each of the subject areas then carefully studied the sets of questions to develop descriptions for the five levels. These descriptions outline the concepts, procedures, or processes associated with correct responses to the questions at each level.

The procedure for describing the writing performance levels was somewhat different. Because the NAEP writing assessment is a direct measure of students' writing abilities, it does not contain questions or tasks that can be scored as correct or incorrect. Instead, students' responses to the writing tasks are rated according to the extent of task accomplishment. The description of the five writing performance levels were developed by examining the ratings received by students whose overall performance was at one level in comparison to the ratings received by students at the next lower level.

Information about trends in students' performance at the scale levels is available back to 1977 in science, 1978 in mathematics, 1971 in reading, and 1984 in writing.

[6] In theory, performance levels above 350 or below 150 could have been defined; however, so few students in the assessment performed at the extreme ends of the subject-area scales that it was not practical to do so.

Tables 2 through 5 present the percentages of students performing at or above each of the five levels. In addition, the summary descriptions that characterized students' performance at each level are provided.

- Compared to 1977 for science and 1978 for mathematics, higher percentages of 9-year-olds in 1994 demonstrated understanding of the fundamentals in both subject areas (Levels 150, 200, and 250). However, no significant change occurred in the percentages of students at age 9 or grade 4 reaching any level of performance on the reading or writing scales.

- At age 13, virtually all students reached Level 150 in science and mathematics, and gains were observed in the percentages of students at or above Levels 200 and 250. Few changes were observed in levels of reading and writing performance. In reading, a larger percentage of 13-year-olds reached Level 300 in 1994 than in 1971. In writing, there were no statistically significant changes at grade 8.

- At age 17, the only change observed in science performance levels was an increase in the percentage of students reaching at least Level 300. In mathematics, 17-year-olds made gains at Levels 250 and 300. No significant change at any performance level was observed in reading. The percentage of eleventh graders at or above Level 250 in writing declined between 1984 and 1994.

Table 2

Percentages of Students Performing At or Above Science Performance Levels, Ages 9, 13, and 17, 1977 and 1994

	AGE 9		AGE 13		AGE 17	
Level	**Percent in 1977**	**Percent in 1994**	**Percent in 1977**	**Percent in 1994**	**Percent in 1977**	**Percent in 1994**
350 Can infer relationship and draw conclusions using detailed scientific knowledge	0(0.0)	0(0.1)	1(0.1)*	0(0.1)	9(0.4)	10(0.8)
300 Has some detailed scientific knowledge and can evaluate the appropriateness of scientific procedures	3(0.3)	4(0.4)	11(0.5)	12(0.9)	42(0.9)*	48(1.3)
250 Understands and applies general information from the life and physical sciences	26(0.7)*	34(1.2)	49(1.1)*	60(1.1)	82(0.7)	83(1.2)
200 Understands some simple principles and has some knowledge, for example, about plants and animals	68(1.1)*	77(1.1)	86(0.7)*	92(0.6)	97(0.2)	97(0.7)
150 Knows everyday science facts	94(0.6)*	97(0.4)	99(0.2)*	100(0.1)	100(0.0)	100(0.1)

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The set of comparisons include all NAEP science trend assessments between 1977 and 1994. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 3

Percentages of Students Performing At or Above Mathematics Performance Levels, Ages 9, 13, and 17, 1978 and 1994

Level	AGE 9		AGE 13		AGE 17	
	Percent in 1978	Percent in 1994	Percent in 1978	Percent in 1994	Percent in 1978	Percent in 1994
350 Can solve multi-step problems and use beginning algebra	0(0.0)	0(0.0)	1(0.2)	1(0.2)	7(0.4)	7(0.8)
300 Can compute with decimals, fractions, and percents; recognize geometric figures; solve simple equations; and use moderately complex reasoning	1(0.1)	1(0.4)	18(0.7)	21(1.4)	52(1.1)*	59(1.4)
250 Can add, subtract, multiply, and divide using whole numbers, and solve one-step problems	20(0.7)*	30(1.1)	65(1.2)*	78(1.1)	92(0.5)*	97(0.5)
200 Can add and subtract two-digit numbers and recognize relationships among coins	70(0.9)*	82(0.7)	95(0.5)*	99(0.3)	100(0.1)	100(0.0)
150 Knows some addition and subtraction facts	97(0.3)*	99(0.2)	100(0.1)	100(0.0)	100(0.0)	100(0.0)

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The set of comparisons include all NAEP mathematics trend assessments between 1978 and 1994. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 4

Percentages of Students Performing At or Above Reading Performance Levels, Ages 9, 13, and 17, 1971 and 1994

Level	AGE 9		AGE 13		AGE 17	
	Percent in 1971	Percent in 1994	Percent in 1971	Percent in 1994	Percent in 1971	Percent in 1994
350 Can synthesize and learn from specialized reading materials	0(0.0)	0(0.0)	0(0.0)	1(0.1)	7(0.4)	7(0.7)
300 Can find, understand, summarize, and explain relatively complicated information	1(0.1)	1(0.3)	10(0.5)*	14(0.8)	39(1.0)	41(1.2)
250 Can search for specific information, interrelate ideas, and make generalizations	16(0.6)	17(1.2)	58(1.1)	60(1.2)	79(0.9)	81(1.0)
200 Can comprehend specific or sequentially related information	59(1.0)	63(1.4)	93(0.5)	92(0.6)	96(0.3)	97(0.5)
150 Can carry out simple, discrete reading tasks	91(0.5)	92(0.7)	100(0.0)	99(0.2)	100(0.1)	100(0.1)

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The set of comparisons include all NAEP reading trend assessments between 1971 and 1994. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 5

Percentages of Students Performing At or Above Writing Performance Levels, Grades 4, 8, and 11, 1984 and 1994

	GRADE 4		GRADE 8		GRADE 11	
Level	Percent in 1984	Percent in 1994	Percent in 1984	Percent in 1994	Percent in 1984	Percent in 1994
350 Can write effective responses containing supportive details and discussion	0(0.0)	0(0.0)	0(0.1)	1(0.2)	2(0.7)	3(0.3)
300 Can write complete responses containing sufficient information	1(0.4)	0(0.2)	13(1.8)	17(1.2)	39(2.4)	33(1.5)
250 Can begin to write focused and clear responses to tasks	10(1.0)	12(0.8)	72(2.6)	67(1.3)	89(1.0)*	85(1.2)
200 Can write partial or vague responses to tasks	54(2.0)	56(2.0)	98(0.9)	96(0.6)	100(0.3)	99(0.2)
150 Can respond to tasks in abbreviated, disjointed, or unclear ways	93(1.3)	92(0.9)	100(0.0)	100(0.1)	100(0.0)	100(0.1)

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The set of comparisons include all NAEP writing trend assessments between 1984 and 1994. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in School and Home Contexts for Learning

Students' responses to background questions about school and home contexts for learning provide an important context for understanding students' educational progress. Although mixed results regarding the supportiveness of students' learning environment were observed, some positive trends were indicated in students' reports. For each school and home factor, results from the 1994 trend assessment are compared with results from the first assessment in which information on that contextual factor was collected.

Science and Mathematics Course Work. An increase in science and mathematics course work was observed across the trend assessments, even though the percentage of students taking advanced courses remained low. Between 1986 and 1994, increases were observed in the percentage of 17-year-old students (primarily twelfth graders) taking biology, chemistry, and physics. Trends in mathematics course taking at age 13 (primarily eighth graders) revealed increased percentages of students taking prealgebra and algebra in 1994 compared to 1986.[7]

As shown in Table 6, an increase in more advanced mathematics course work was also reported by 17-year-olds. Between 1978 and 1994, the percentage of 17-year-old students who had studied only prealgebra or general mathematics decreased. Conversely, there was an increase during the same time period in the percentage of 17-year-olds pursuing mathematics course work through algebra II, or precalculus or calculus.

[7] A complete discussion of science and mathematics course-taking patterns is presented in Chapers 3 and 6.

Table 6

Highest Level of Mathematics Course Work, Age 17, 1978 and 1994

	PERCENTAGE OF STUDENTS				
	General Mathematics or Prealgebra	Algebra I	Geometry	Algebra II	Precalculus or Calculus
1994	9(1.1)	15(0.9)	15(0.8)	47(1.6)	13(1.2)
1978	20(1.0)*	17(0.6)	16(0.6)	37(1.2)*	6(0.4)*

* Statistically significant difference between 1978 and 1994 with a 5 percent significance level. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference. Percentages do not total 100 percent because some students reported "other" (4 percent in 1978 and 1 percent in 1994).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Technology in the Classroom. Students' reports indicated an increased use of technology in the classroom. Between 1977 and 1994, there was an increase in the percentage of 9-year-olds who reported using a calculator, thermometer, or microscope in their classrooms.[8] As shown in Table 7, students in 1994 were much more likely to have used a computer in school than were students in the late 1970s and early 1980s. At ages 13 and 17, students reported considerably more access to and use of computers in mathematics classes in 1994 than in 1978. Also, between 1984 and 1994, there was a sharp increase in the percentage of students at grades 8 and 11 who reported using computers to write stories or papers.

[8] A complete discussion of technology use in classrooms is found in Chapters 3 and 6.

Table 7

Computer Usage in Mathematics and Writing Instruction, Ages 13 and 17, 1978/1984 and 1994

		PERCENTAGE OF STUDENTS ANSWERING "YES"	
		AGE 13	**AGE 17**
Studied Mathematics Through Computer Instruction	1994	50(1.8)	34(1.7)
	1978	14(0.9)*	12(1.1)*
		GRADE 8	**GRADE 11**
Used a Computer To Write Stories or Papers	1994	82(1.7)	87(2.0)
	1984	15(3.5)*	19(2.2)*

* Statistically significant difference between 1978 or 1984 and 1994 with a 5 percent significance level. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Homework. According to students' reports, there was little change between 1984 and 1994 in the amount of time overall that students spent each day working on homework for all subjects. The only change observed was at age 9, where more students in 1994 than in 1984 reported doing less than 1 hour of homework each day and fewer students reported doing more than 2 hours. In 1994, the percentage of students who reported doing at least 1 hour of homework daily was 16 percent at age 9, 37 percent at age 13, and 39 percent at age 17.[9]

Among 9- and 13-year-olds, there was evidence of slightly more reading for school and for homework in 1994 than in 1984. As shown in Table 8, the percentages of 9- and 13-year-old students who reported reading more

[9] A complete discussion of time spent on homework is presented in Chapter 9.

than 20 pages each day increased over the 10-year period. At age 9, there was a corresponding drop in the percentage of students who reported reading 5 or fewer pages each day. At age 13, there was also an increase in the percentage of students who reported reading 16 to 20 pages as well as a decline in the percentage of students who reported reading only 6 to 10 pages. The reports of 17-year-old students on the number of pages read each day did not change significantly between 1984 and 1994.

Table 8

Pages Read in School and for Homework Per Day, Ages 9, 13, and 17, 1984 and 1994

		PERCENTAGES OF STUDENTS		
		AGE 9	AGE 13	AGE 17
More than 20	1994	17(1.0)	14(0.8)	23(1.5)
	1984	13(0.4)*	11(0.4)*	20(1.0)
16 to 20 pages	1994	14(0.9)	13(0.5)	13(0.6)
	1984	13(0.5)	11(0.2)*	14(0.4)
11 to 15 pages	1994	14(0.5)	17(0.6)	18(0.6)
	1984	14(0.5)	18(0.4)	18(0.3)
6 to 10 pages	1994	26(0.6)	31(0.9)	25(0.9)
	1984	25(0.5)	35(0.5)*	26(0.6)
5 or fewer	1994	28(1.4)	26(0.9)	21(1.2)
	1984	35(1.0)*	27(0.6)	21(0.8)

* Statistically significant difference between 1984 and 1994 with a 5 percent significance level. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Value for Learning. Some positive trends were observed in students' perceptions about the value of learning different subjects. For example, there was an increase between 1977 and 1994 in the percentage of 17-year-olds who strongly agreed that learning science can be useful in the future and that science should be required in school. At age 13, a smaller percentage of students in 1994 than in 1978 indicated that they were taking mathematics only because it was required. Among fourth graders, a decrease between 1984 and 1994 was observed in the percentage of students who indicated that they would not write anything if it was not required in school.[10]

Factors in the Home. Based on the relatively small number of questions asked, factors in the home that previously have shown a relationship to achievement appeared to change little from assessment to assessment.[11] At age 9, there was an increase in the percentage of students who reported reading for fun daily (see Table 9). However, there was an increase in the percentage of 13-year-olds who reported reading for fun on a yearly basis. No significant change was observed in reading for fun among 17-year-olds. As shown in Table 10, television viewing habits appeared to have changed only slightly since the late 1970s and early 1980s. Nine- and 13-year-olds were less likely in 1994 than in 1982 to watch six or more hours of television each day. However, there was a small but significant increase between 1978 and 1994 in the percentage of 17-year-olds who reported watching six or more hours of television each day.

---

[10] A complete discussion of students' attitudes toward learning is presented in Chapters 3, 6, and 13.

[11] Past NAEP assessments have shown a relationship between achievement and both television watching and reading for fun. (See the *NAEP 1994 Reading Report Card for the Nation and the States.*)

Table 9

Trends in Reading for Fun, Ages 9, 13, and 17, 1984 and 1994

		PERCENTAGES OF STUDENTS		
		AGE 9	AGE 13	AGE 17
Daily	1994	58(1.6)	32(1.8)	30(2.6)
	1984	53(1.0)*	35(1.0)	31(0.8)
Weekly	1994	25(1.5)	32(2.1)	31(1.9)
	1984	28(0.8)	35(1.2)	34(1.1)
Monthly	1994	5(0.6)	14(1.7)	15(1.5)
	1984	7(0.6)*	14(0.8)	17(0.5)
Yearly	1994	3(0.6)	10(1.2)	12(1.5)
	1984	3(0.3)	7(0.5)*	10(0.5)
Never	1994	9(0.8)	12(1.7)	12(1.4)
	1984	9(0.5)	9(0.6)	9(0.6)

* Statistically significant difference between 1984 and 1994 with a 5 percent significance level. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 10

Trends in Television Watching, Ages 9, 13, and 17, 1978/1982 and 1994

	PERCENTAGE OF STUDENTS		
	NUMBER OF HOURS WATCHED PER DAY		
	0-2 Hours	3-5 Hours	6 or More Hours
Age 9			
1994	43(1.0)	38(0.9)	19(0.8)
1982	44(1.1)	29(0.6)*	26(1.0)*
Age 13			
1994	38(1.3)	49(1.1)	13(0.6)
1982	45(0.8)*	39(0.4)*	16(0.8)*
Age 17			
1994	53(1.7)	39(1.3)	8(0.7)
1978	69(0.7)*	26(0.6)*	5(0.2)*

* Statistically significant difference between 1978 or 1982 and 1994 with a 5 percent significance level. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

About This Report

A primary purpose of the National Assessment of Educational Progress is to measure trends in academic performance. The *NAEP 1994 Trends in Academic Progress* provides a broad examination of students' learning in four core academic areas — science, mathematics, reading, and writing. In addition to overall results, extensive subgroup (e.g., racial/ethnic subgroups, males and females) results are presented. Specific aspects of students' performance and of home and school factors related to achievement are reviewed at length.

This report is presented in six sections. The first four sections correspond to the four core academic areas assessed. In the science, mathematics, and reading sections, the first chapter discusses overall scale score results, the second discusses performance level results, and the third discusses home and school factors. In the writing section, the first chapter discusses overall scale score results, the second discusses performance level results, the third discusses aspects of writing (e.g., purposes for writing and writing mechanics), and the fourth discusses home and school factors. This report also contains a Procedural Appendix and a comprehensive Data Appendix.

Part I

Trends in Science Achievement from 1969-70 to 1994

Introduction

> *By the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography, and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our nation's modern economy.*[1]

The present impetus in science reform can be traced to a report issued in 1983 by the National Commission on Excellence in Education[2] that was critical of education in the United States. In the 1980s, publication of a

[1] Washington, DC: National Education Goals Panel, 1991.

[2] *A Nation at Risk: The Imperative for Education Reform* (Washington, DC: National Commission on Excellence in Education, 1983).

number of reports and analyses pointing out the deficiencies of the educational system and how to address them fueled interest in reform.[3] Recently, governmental, professional, and private organizations have taken on pivotal roles in providing support to state and local educational establishments as they reform their educational systems.[4] Areas looked at include the development of standards, revision of curricula, development of appropriate assessment techniques, and professional development. Several organizations have worked closely with the authors of the National Science Education Standards[5] and published documents to help teachers interpret these standards.[6]

To help policy makers and educators assess the outcomes of their quest for excellence in science learning, it is important to find out what American students know and can do in science. Over the past 25 years, NAEP has administered eight assessments to monitor progress in science performance by nationally representative samples of 9-, 13-, and 17-year-old students. These trend assessments were administered in 1969-70, 1972-73, 1976-77, 1981-82, 1985-86, 1989-90, 1991-92, and 1993-94. The subsequent text refers to each assessment by the last half of the school year in which it was administered: 1969 or 1970, 1973, 1977, 1982, 1986, 1990, 1992, and 1994. It should be noted that some of the analyses reported in this section do not go back to the first mathematics trend assessment because the data are not available.

In addition to the trend assessments, NAEP conducted a 1996 survey of science achievement among national samples of students in grades 4, 8, and 12 and state samples of students in grade 8. To keep abreast of current pedagogic research, this most recent main NAEP science assessment[7] included performance tasks such as hands-on investigations and

---

[3] *Educating Americans for the 21st Century: A Report to the American People and the National Science Board* (Washington, DC: National Science Board, Commission on Precollege Education in Mathematics, Science, and Technology, 1983).

[4] *Statewide Systemic Initiatives in Science, Mathematics, and Engineering* (Arlington, VA: The National Science Foundation, 1995-1996).

Scope, Sequence, and Coordination of High School Science (Washington, DC: National Science Teachers Association, 1995).

Benchmarks for Science Literacy (Washington, DC: Project 2061, American Association for the Advancement of Science, 1993).

New Standards Project (Washington, DC: National Center on Education and the Economy, 1993).

[5] *National Science Education Standards* (Washington, DC: National Research Council, 1995).

[6] *A High School Framework for National Science Education Standards* (Arlington, VA: National Science Teachers Association, 1995).

[7] *Science Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, 1995).

short-answer exercises, as well as multiple-choice questions. The first reports on this assessment will be available in 1997.

As distinct from the NAEP science assessment conducted in 1996 to collect national data for students in grades 4, 8, and 12, and state data for students in grade 8, the long term trend assessment conducted in 1994 sampled students from across the country at ages 9, 13, and 17. In addition to sampling students by age rather than grade, another important difference between the 1996 NAEP science assessment and the trend science assessment was the set of questions contained in each instrument. In order to allow for measuring trends in achievement since the first science trend assessment, the administration procedures and content have remained unchanged, and were once again replicated in 1994. While the new instrument developed for the 1996 national and state assessments placed particular emphasis on constructed-response questions and performance tasks, the trend science assessment contains only multiple-choice questions.

Because of the differences in sampling procedures and in content, results from the 1996 NAEP science assessment and the trend science assessment are not directly comparable. However, results from the trend assessment can provide valuable information about the attainment of long-held educational goals during a time of change and reform. For example, while school curricula shift toward increased emphasis on the application of science knowledge and the ability to communicate scientific concepts, trend results indicate whether students are maintaining their grasp of basic science knowledge and skills.

The trend science assessments contain a content dimension and a cognitive dimension.[8] The content dimension assesses life science, physical science, and earth and space science. The cognitive dimension assesses students’ ability to conduct inquiries, solve problems, and know science. NAEP also assesses students’ understanding of the nature of science within the context of both content area knowledge and cognition.

Estimates of average student performance in the science trend assessments were calculated using analysis techniques based on item response theory (IRT). The resultant scale, which spans 0 to 500, allows for comparisons of average scores across assessments, age groups, and demographic subpopulations. (The Procedural Appendix contains more detailed explanations of the analysis procedures and definitions of student subpopulations.) Five different levels of science performance have been

---

[8] *Science Objectives: 1985-1986 Assessment* (Princeton, NJ: National Assessment of Educational Progress, 1986). *Science Objectives: 1990 Assessment* (Princeton, NJ: National Assessment of Educational Progress, 1989).

defined on the NAEP trend scale: Level 150 — Knows Everyday Science Facts, Level 200 — Understands Simple Scientific Principles, Level 250 — Applies General Scientific Information, Level 300 — Analyzes Scientific Procedures and Data, and Level 350 — Integrates Specialized Scientific Information.

Two measures of performance are used in this report: the average scores of groups of students on the NAEP science scale and the percentages of students within each group attaining each of the five performance levels on the scale. Because the average scale scores and the percentages are based on samples of students and are subject to sampling and measurement error, standard errors are included with the results presented here. Statistically significant differences between previous assessments and the one conducted in 1994 are denoted with an asterisk, and statistically significant differences between 1969-70 and subsequent assessments are denoted with a dagger.

In addition to point-by-point multiple comparisons, a second test of significance was conducted to detect statistically significant linear and quadratic trends across the assessments. (See the Procedural Appendix for a discussion of the procedure.) This type of analysis makes it possible to discuss statistically significant patterns that may be missed by point-to-point comparisons. For example, from assessment to assessment, students' average scale scores may consistently increase by a small amount. Although these small increases between years may not be statistically significant under pairwise multiple comparisons, the overall increasing trend in average scores may be statistically significant and noteworthy. All of the differences and trend patterns discussed in this report are statistically significant at the .05 level.

The three chapters in Part I concentrate on different aspects of student performance. Trends in average science scale scores for the nation and demographic subpopulations are reported in Chapter 1. Chapter 2 defines levels of science performance and presents information on the percentage of students attaining successive levels in each assessment. Chapter 3 summarizes trends in students' responses to questions relating to classroom experiences, course taking, attitudes toward science, and perceived applications of science.

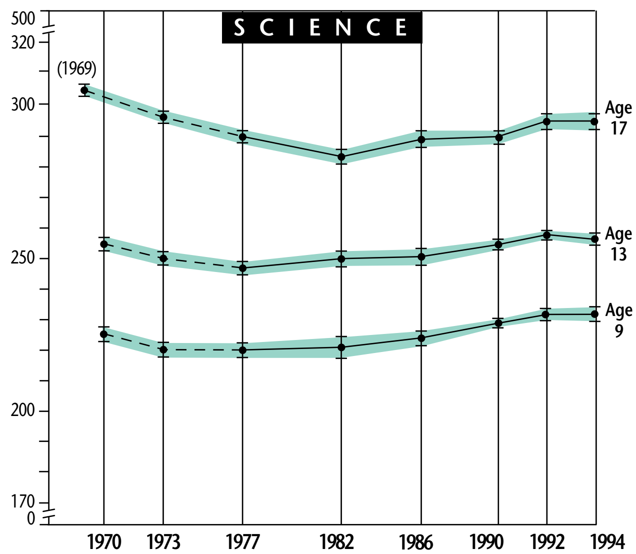
1

National Trends in Science Scale Scores from 1969-70 to 1994

Figure 1.1 depicts trends in average science scale scores for 9-, 13-, and 17-year-old students. The results for 1969 (17-year-olds only), 1970 (9- and 13-year-olds), and 1973 (all age groups) are extrapolated from previous analyses of NAEP data. Results for the 1977, 1982, 1986, 1990, 1992, and 1994 assessments are based on more recent analyses. (Please refer to the Procedural Appendix for details of scaling methodology and information about drawing inferences from trend analyses.)

Figure 1.1

Trends in Average Science Scale Scores for the Nation, 1969-70 to 1994



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	305(1.0)*	296(1.0)†	290(1.0)†	283(1.2)*†	289(1.4)†	290(1.1)†	294(1.3)†	294(1.6)†
Age 13	255(1.1)	250(1.1)*†	247(1.1)*†	250(1.3)*†	251(1.4)*	255(0.9)	258(0.8)	257(1.0)
Age 9	225(1.2)*	220(1.2)*†	220(1.2)*†	221(1.8)*	224(1.2)*	229(0.8)	231(1.0)†	231(1.2)†

95-percent confidence interval. [– – –] Extrapolated from previous NAEP analyses.

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Nine-Year-Olds. The performance of 9-year-olds declined in the 1970s but has since shown a pattern of improvement. In 1994, the average score for this age group was higher than the average score in 1970, but did not differ significantly from the 1992 average.

Thirteen-Year-Olds. The average score of 13-year-olds declined during the 1970s, but has since increased. Although the 1994 average score of these students was not significantly different than the 1970 average score, the 1994 average was higher than averages from 1973 to 1986.

Seventeen-Year-Olds. The performance of 17-year-olds declined from 1969 through 1982. Their average score has increased since that time, but the average for 1994 continued to be lower than the average score of 17-year-olds in 1969.

Trends in Science Scale Scores from 1977 to 1994 by Quartiles

Table 1.1 depicts the average science scale scores of 9-, 13-, and 17-year-old students who were in the upper quartile (upper 25 percent), middle two quartiles (middle 50 percent), and the lower quartile (lower 25 percent) of student performance in each assessment. As would be expected, standard errors are generally smaller for these more homogeneous groups than for the total group. (Please note that these trends are not extrapolated back to 1969 or 1970).

These data reveal the changes that have occurred in the last 25 years for students at different points along the distribution, demonstrating whether overall gains or losses were evident across the full range of performance in science. This information is particularly relevant in light of Goal 3 of *The National Education Goals Report,* which states that "the academic performance of elementary and secondary students will increase significantly in every quartile . . ."[9] The report emphasizes that students of all abilities should be granted access to educational opportunities and should demonstrate gains in educational achievement.

A similar pattern of results was evident for 9- and 13-year-old students across the performance distribution. The average scores for both age groups in each performance range increased since 1977. However, the last two assessments have shown no significant changes. For 17-year-old students in the upper and middle two quartiles, average scores also increased between 1977 and 1994. The average score of 17-year-olds in the lower quartile declined after the 1977 assessment and remained relatively stable until 1992 when some improvement was observed. In 1994, however, the average score of the students in this quartile continued to be lower than in 1977.

[9] *The National Education Goals Report,* National Education Goals Panel (Washington, DC: U.S. Government Printing Office, 1992).

Table 1.1

Trends in Average Science Scale Scores by Quartiles, 1977 to 1994

Quartile	Year	AVERAGE SCALE SCORE Age 9	Age 13	Age 17
Upper Quartile	1994	273(1.0)†	298(1.2)†	347(0.9)†
	1992	273(1.2)†	298(1.0)†	346(0.7)†
	1990	271(0.8)†	297(0.7)†	344(0.7)†
	1986	269(1.2)*	292(1.1)*	340(1.1)*†
	1982	268(1.8)	291(0.9)*	329(1.0)*†
	1977	266(0.9)*	291(0.5)*	334(0.9)*
Middle Two Quartiles	1994	233(1.1)†	259(0.7)†	296(0.7)†
	1992	233(0.7)†	260(0.5)†	295(1.0)†
	1990	231(0.5)†	256(0.6)*†	292(0.7)*
	1986	226(0.6)*†	252(0.7)*†	290(0.7)*
	1982	222(1.1)*	251(0.6)*	286(0.7)*†
	1977	222(0.5)*	249(0.6)*	291(0.5)*
Lower Quartile	1994	183(1.2)†	214(1.3)†	236(1.7)†
	1992	184(1.2)†	214(0.8)†	240(1.9)
	1990	182(0.9)†	211(1.2)†	234(1.2)†
	1986	177(1.0)*†	209(0.9)*†	235(1.3)†
	1982	171(2.0)*	208(0.8)*†	232(1.3)†
	1977	170(1.1)*	201(0.8)*	242(0.8)*

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1977, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Scale Scores from 1969-70 to 1994 by Race/Ethnicity and by Gender

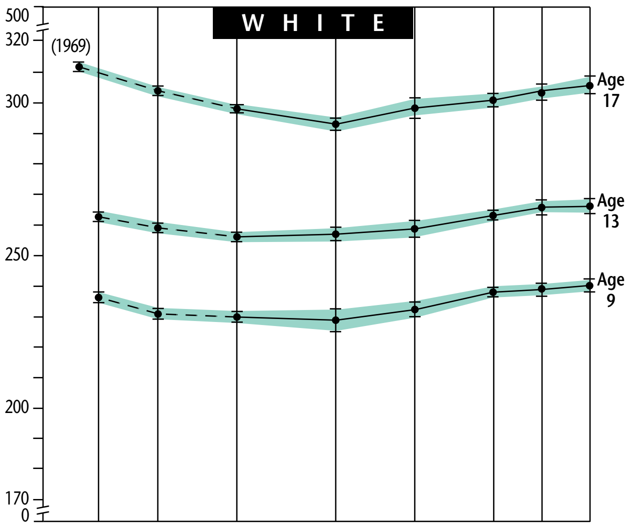
Race/Ethnicity. Figure 1.2 shows trends in average science scale scores for White, Black, and Hispanic students. After a period of decline during the 1970s and early 1980s for White 9-year-olds and during the 1970s for White 13-year-olds, performance improved. These gains resulted in 1994 average scores for both age groups that were higher than those in 1970 but did not differ significantly from 1992 levels. The average score of White 17-year-olds declined from 1969 to 1982. Despite gains that have been made since that time, the 1994 average score for these students continued to be lower than the average score of their counterparts in 1969.

The average score of Black 9-year-olds increased between 1970 and 1994. An overall pattern of increased performance was also observed for Black 13-year-olds. The 1994 average score was higher than the average score observed in 1973 and 1977, but did not differ significantly from the 1970 and 1992 average scores. Following a decline from 1969 to 1982 in the average score of 17-year-old Black students, an overall pattern of increased performance was evident. In 1994, however, the average score of these students remained at a level not significantly different from that in 1969, and average scores in the four most recent assessments did not differ significantly.

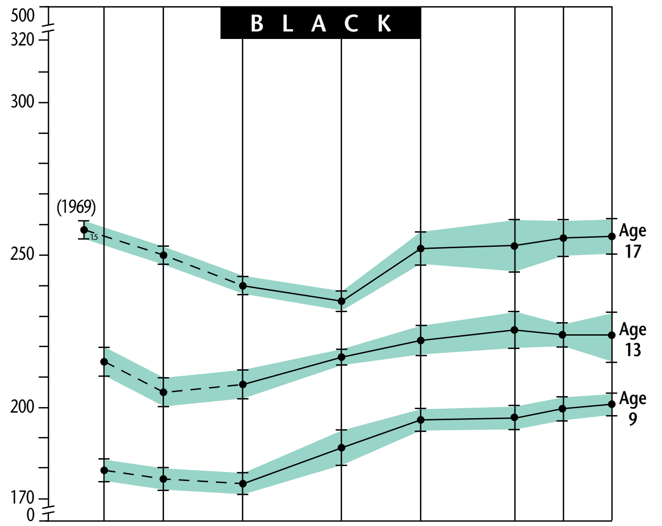
Among Hispanic students, 9-year-olds displayed an overall pattern of improved performance. However, the 1994 average score for these students was not significantly different from the 1977 average score. For 13-year-old Hispanic students, an overall gain observed across the assessment years resulted in a higher average score in 1994 than in 1977. Despite some fluctuations, no significant change between 1977 and 1994 was observed in the average scores of Hispanic 17-year-olds.

Figure 1.2

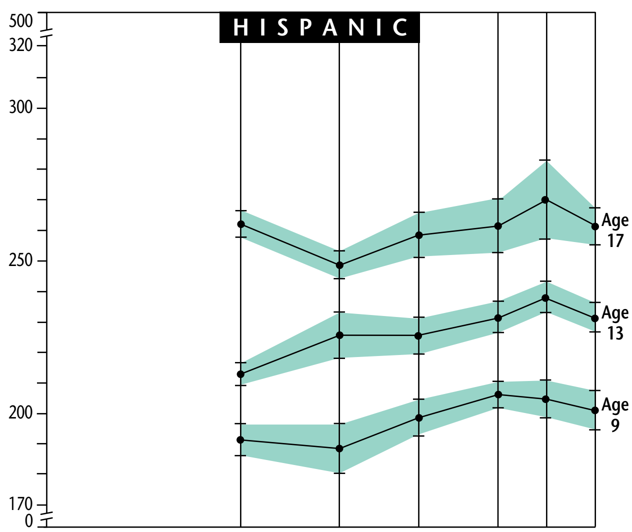
Trends in Average Science Scale Scores by Race/Ethnicity, 1969-70 to 1994



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	312(0.8)*	304(0.8)†	298(0.7)*†	293(1.0)*†	298(1.7)*†	301(1.1)*†	304(1.3)†	306(1.5)†
	—	—	83(1.3)*	81(2.0)*	78(0.5)*†	73(0.5)†	75(0.5)*†	72(0.6)†
Age 13	263(0.8)*	259(0.8)*†	256(0.8)*†	257(1.1)*†	259(1.4)*	264(0.9)	267(1.0)†	267(1.0)†
	—	—	80(1.6)*	79(2.1)*	77(1.0)*	73(0.7)†	74(0.5)†	72(0.8)†
Age 9	236(0.9)*	231(0.9)*†	230(0.9)*†	229(1.9)*†	232(1.2)*†	238(0.8)	239(1.0)	240(1.3)†
	—	—	80(1.6)*	79(2.6)	77(1.1)	75(1.1)†	75(0.9)	74(0.9)†



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	258(1.5)	250(1.5)†	240(1.5)*†	235(1.7)*†	253(2.9)	253(4.5)	256(3.2)	257(3.1)
	—	—	12(1.1)*	13(1.4)	14(0.3)	16(0.3)†	15(0.3)†	15(0.3)†
Age 13	215(2.4)	205(2.4)*†	208(2.4)*	217(1.3)	222(2.5)	226(3.1)†	224(2.7)	224(4.2)
	—	—	13(1.2)	14(1.9)	14(0.9)	16(0.3)	16(0.3)	15(0.3)
Age 9	179(1.9)*	177(1.9)*	175(1.8)*	187(3.0)*	196(1.9)†	196(2.0)†	200(2.7)†	201(1.7)†
	—	—	14(1.4)	14(2.1)	15(0.5)	16(0.7)	16(0.4)	15(0.5)



	1977	1982	1986	1990	1992	1994
Age 17	262(2.2)	249(2.3)†	259(3.8)	262(4.4)	270(5.6)	261(6.7)
	4(0.9)*	5(1.1)*	6(0.3)*	7(0.4)*†	7(0.5)†	9(0.3)†
Age 13	213(1.9)*	226(3.9)†	226(3.1)†	232(2.6)†	238(2.6)†	232(2.4)†
	5(1.1)*	5(1.0)*	7(1.1)	7(0.5)	7(0.5)	8(0.4)†
Age 9	192(2.7)	189(4.2)	199(3.1)	206(2.2)†	205(2.8)†	201(2.7)
	5(0.9)	5(1.3)	6(1.1)	6(0.6)	6(0.8)	6(0.4)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval. [– – –] Extrapolated from previous NAEP analyses.

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores for White and Black students) or from 1977 (for scale scores for Hispanic students and for all other percentages), at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

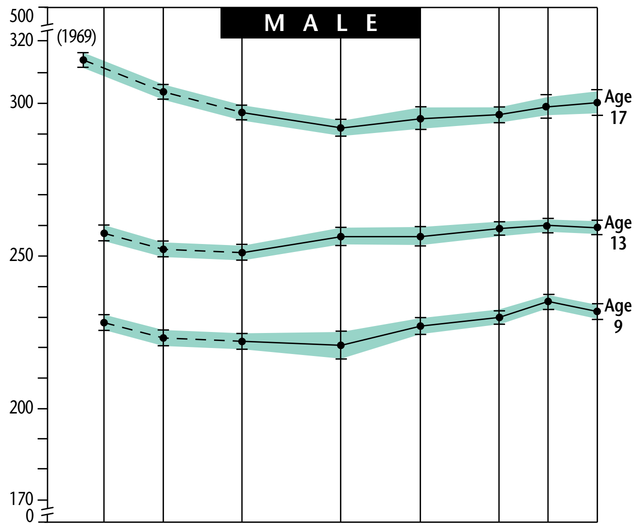
Gender. Trend analyses of data presented in Figure 1.3 indicated that after a period of decline (in the 1970s, and in some cases the early 1980s), the performance of male and female students at all three ages demonstrated a pattern of improvement.

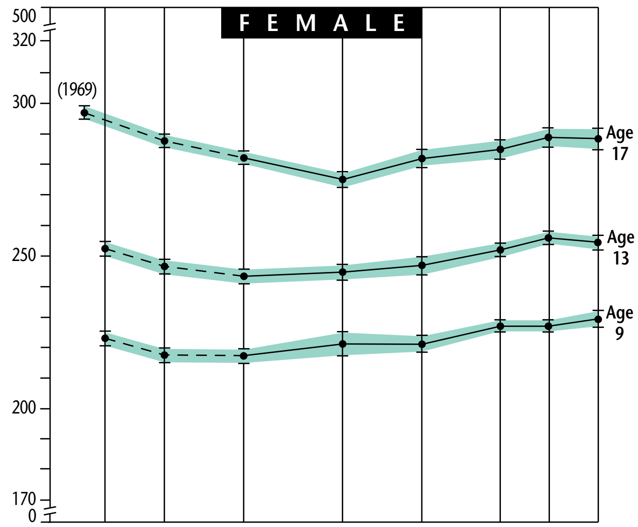
Among male students, the average scores of 9- and 13-year-olds remained at a level in 1994 that was not significantly different from that in 1970. Although 17-year-old males made gains since 1982, the 1994 average continued to be lower than the 1969 average.

Among female students, the gains made by 9-year-olds resulted in a 1994 average score that was higher than that in 1970. The average score of female 13-year-olds did not differ significantly from that of their counterparts in 1970. Despite the recent pattern of improvement among 17-year-old females, the 1994 average remained lower than the 1969 average.

Figure 1.3

Trends in Average Science Scale Scores by Gender, 1969-70 to 1994





Male

	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	314(1.2)*	304(1.2)†	297(1.2)†	292(1.4)*†	295(1.9)†	296(1.3)†	299(1.7)†	300(2.0)†
	—	—	50(0.7)	48(0.7)	49(1.2)	49(0.9)	51(1.2)	49(1.3)
Age 13	257(1.3)	252(1.3)*†	251(1.3)*†	256(1.5)	256(1.6)	259(1.1)	260(1.2)	259(1.2)
	—	—	50(0.3)	49(0.7)	50(0.9)	50(0.6)	50(0.8)	49(0.8)
Age 9	228(1.3)	223(1.3)*†	222(1.3)*†	221(2.3)*	227(1.4)	230(1.1)	235(1.2)†	232(1.3)
	—	—	51(0.4)	51(1.5)	50(0.6)	49(0.6)	49(0.6)†	49(0.8)

Female

	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	297(1.1)*	288(1.1)†	282(1.1)*†	275(1.3)*†	282(1.5)*†	285(1.6)†	289(1.5)†	289(1.7)†
	—	—	51(0.7)	52(0.7)	51(1.2)	51(0.9)	49(1.2)	50(1.3)
Age 13	253(1.2)	247(1.2)*†	244(1.2)*†	245(1.3)*†	247(1.5)*†	252(1.1)	256(1.0)	254(1.2)
	—	—	50(0.3)	52(0.7)	50(0.9)	50(0.6)	50(0.8)	51(0.8)
Age 9	223(1.2)*	218(1.2)*†	218(1.2)*†	221(2.0)*	221(1.4)*	227(1.0)	227(1.0)	230(1.4)†
	—	—	49(0.4)	50(1.5)	50(0.6)	51(0.6)	51(0.6)†	51(0.8)

Below each average scale score, the corresponding percentage of students is presented.

$\bar{\bullet}$ 95-percent confidence interval. [– – –] Extrapolated from previous NAEP analyses.

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores) or 1977 (for percentages), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Differences in Average Science Scale Scores

The previous section discussed the trends in science achievement for White, Black, and Hispanic students, and for male and female students. As with past NAEP assessments, significant differences were observed between racial/ethnic subgroups and between males and females.

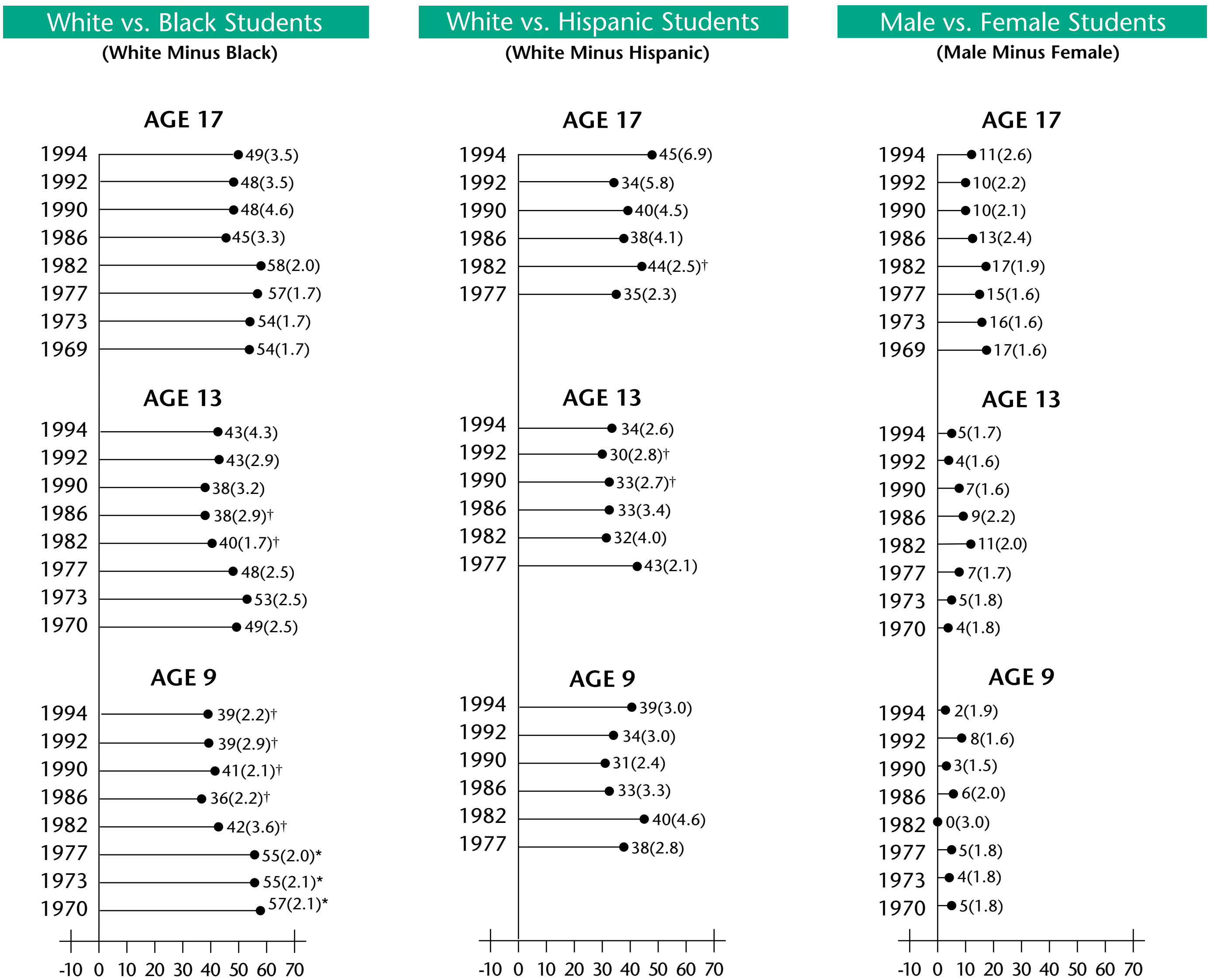
In 1994 at all three ages, White students outperformed Black and Hispanic students. The scale score gap between White and Black 9-year-olds declined between 1970 and 1986, and has remained relatively stable since that time. The gap in 1994, however, continued to be smaller than the gap in 1970. The reduction in the gap for 9-year-olds was the result of score increases for Black students between 1977 and 1986 while average scale scores for White students were relatively stable during the same time period. The average scale score gap between White and Black 13-year-olds also showed some evidence of decline through the mid-1980s. Again, this decline was the result of an increase in scores for Black students while scores for White students remained stable. However, despite some decline, the gap in the most recent assessment was not significantly different from that observed in 1970. For 17-year-olds, the overall trend analyses suggest that gaps observed in the 1986 to 1994 assessments are smaller than those in prior assessments. This reduction in the gap resulted from an 18-point gain by Black students between 1982 and 1986, compared to a 4-point gain for White students. However, as was the case for 13-year-olds, a direct comparison of the 1969 and 1994 scale score gaps showed no statistically significant change despite the apparent decline.

Trend analyses across the assessment years 1977 to 1994 showed no strong evidence of overall change in the average scale score gaps between White and Hispanic 9- and 17-year-old students. In contrast, the difference in average scale scores between White and Hispanic 13-year-olds decreased across the assessment years. The decrease in the gaps between average scale scores was the result of a 19-point increase for Hispanic students from 1977 to 1994, compared to a 10-point increase for White students during the same time period.

In 1994 male 13- and 17-year-olds continued to have higher average scale scores than their female peers. The gap between the average scale scores of 17-year-old male and female students were generally smaller in the last four assessments when compared to the gaps observed in assessments prior to 1986. The reduction in the gap resulted from gains for female students in 1986 (7 points) and 1990 (3 points) compared to male students (3 points in 1986 and 1 point in 1990). Among 9- and 13-year-olds, there was little evidence of change in the magnitude of the difference between male and female students' average scores.

Figure 1.4

Trends in Differences in Average Science Scale Scores



* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores for White vs. Black student and Male vs. Female student differences) or from 1977 (for White vs. Hispanic student differences), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale score differences appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Scale Scores from 1969-70 to 1994 by Region

Figure 1.5 depicts trends in average science scale scores by region. In the Northeast, 9-year-olds showed early declines followed by subsequent gains in performance. The overall pattern for these students was one of increased performance. Despite this overall pattern, the 1994 average score did not differ significantly from the 1970 average score. For 13-year-olds, there is some evidence of a decline in scores during the 1970s and early 1980s. Since that time, gains were observed, but these did not result in a 1994 average score that was significantly different from that in 1970. The average score of 17-year-olds declined from 1969 to 1982, but has since increased. No significant difference was observed between the 1969 and 1994 average scores for these students.

In the Southeast, an overall pattern of increased performance resulted in a 1994 average score for 9-year-olds that was higher than the 1970 average score of their counterparts. For 13-year-olds, scores declined during the 1970s, but have since increased. The average score in 1994 for these students was higher than that in the first assessment. The average score of 17-year-olds also declined during the 1970s, but subsequently increased. However, no significant difference was observed between the 1969 and 1994 average scores.

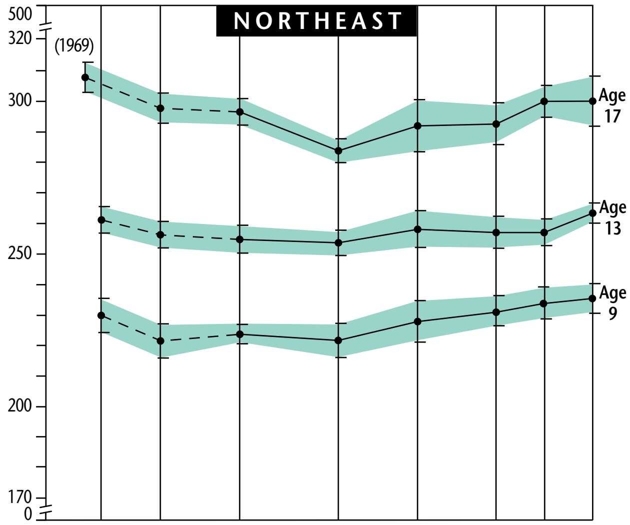
For 9-year-olds in the Central region, average scores declined during the 1970s. Higher average scores have been observed since that time, and the overall pattern has been one of increased performance. However, there was no difference between the 1970 and 1994 average scores for this group of students. Declining performance was observed among 13-year-olds from 1970 through 1986. Since that time, the average score has returned to a level not significantly different from that in 1970. Among 17-year-olds, the average score declined from 1969 to 1982, but has since increased. However, in 1994 the average score did not differ significantly from that in 1969.

In the West region, the overall pattern of performance for 9-year-olds was one of increased performance. However, 1970 and 1994 average scores did not differ significantly. After declines from 1970 to 1977, the average score of 13-year-olds increased. The pattern for these students shows overall improvement, although the 1970 and 1994 average scores do not differ significantly. Decreasing scores were observed for 17-year-olds from 1969 to 1982, followed by a pattern of increasing scores. However, the 1994 average score for these students continued to be lower than the average score of their counterparts in 1969.

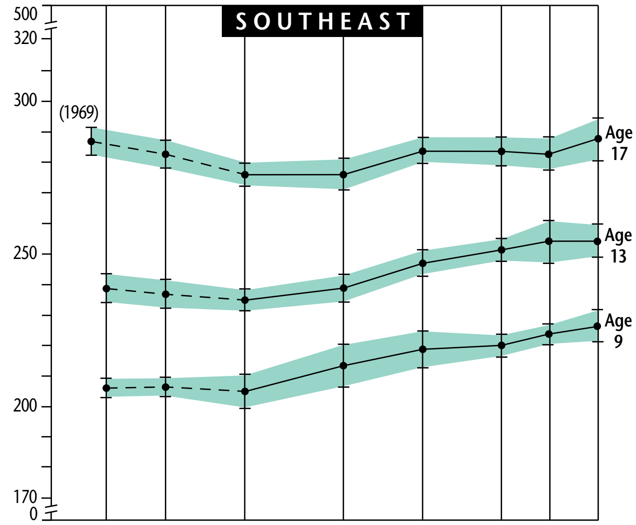
A comparison of the 1994 average scores of students from different regions revealed only two instances of differences. Among 13-year-olds, students in the Northeast had average scores that were higher than that of students in the Southeast or West regions.

Figure 1.5

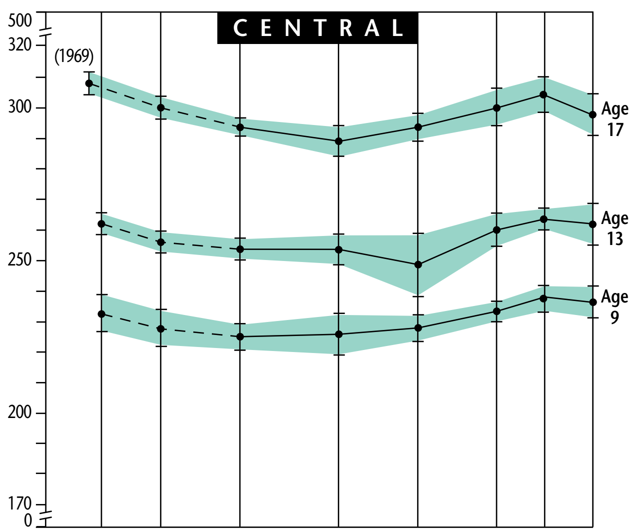
Trends in Average Science Scale Scores by Region, 1969-70 to 1994



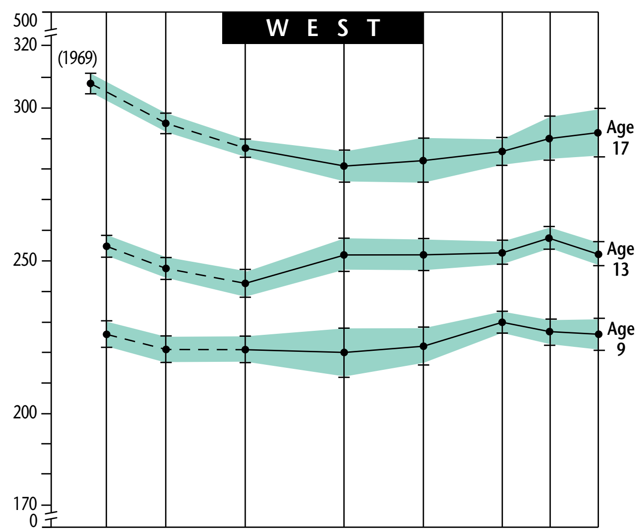
	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	308(2.5)	298(2.5)†	296(2.2)†	284(2.0)*†	292(4.3)†	293(3.2)†	300(2.4)	299(4.2)
	—	—	24(1.7)	25(3.0)	24(0.8)	22(1.0)	22(0.9)	23(1.0)
Age 13	261(2.2)	256(2.2)	255(2.3)	254(2.1)*	258(3.1)	257(2.7)	257(2.2)	263(1.7)
	—	—	26(1.5)*	24(2.9)	22(1.6)	24(1.0)*	21(0.8)†	20(0.9)†
Age 9	230(2.9)	222(2.9)*	224(1.6)*	222(2.9)*	228(3.5)	231(2.4)	234(2.8)	235(2.5)
	—	—	25(1.5)	19(2.3)	21(1.1)	23(0.9)	21(1.2)	22(0.9)



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	287(2.3)	283(2.3)	276(1.9)*†	276(2.7)*†	284(2.0)	284(2.4)	283(2.5)	288(2.8)
	—	—	18(1.0)*	21(2.4)	23(2.3)	24(0.9)†	25(1.4)†	25(1.5)†
Age 13	239(2.4)*	237(2.4)*	235(1.8)*	239(2.3)*	247(2.2)	251(1.9)†	254(2.8)†	253(2.6)†
	—	—	21(1.1)	22(2.3)	25(5.8)	23(0.8)	24(1.6)	22(1.7)
Age 9	206(1.6)*	207(1.6)*	205(2.9)*	214(3.6)*	219(3.1)†	220(1.9)†	223(1.7)†	227(2.2)†
	—	—	20(1.2)	23(2.7)	23(4.7)	24(0.9)†	24(1.3)	25(1.5)



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	308(1.9)	300(1.9)†	294(1.5)†	289(2.6)†	294(2.3)†	300(3.0)	304(2.7)	298(3.7)
	—	—	33(1.6)*	30(3.9)	28(2.2)	26(0.9)†	25(1.0)†	26(1.1)†
Age 13	262(1.8)	256(1.8)	254(1.8)†	254(2.6)	249(5.3)	260(2.8)	263(2.1)	261(3.5)
	—	—	27(1.9)	28(3.9)	25(5.0)	24(0.7)	27(1.0)	27(1.8)
Age 9	233(3.0)	228(3.0)	225(2.2)*	226(3.5)	228(2.2)	234(1.7)	238(1.8)	236(2.7)
	—	—	30(1.9)	30(4.3)	29(4.0)	24(0.6)†	28(0.8)	26(1.0)



	1970	1973	1977	1982	1986	1990	1992	1994
Age 17	308(1.7)*	295(1.7)†	287(1.5)†	281(2.7)†	283(3.8)†	286(2.3)†	290(3.8)†	292(4.1)†
	—	—	24(1.9)	25(2.6)	26(0.9)	28(0.9)	29(0.9)	27(1.0)
Age 13	255(1.8)	248(1.8)†	243(2.3)*†	252(2.8)	252(2.7)	253(2.1)	258(1.6)	252(2.1)
	—	—	26(2.1)	26(2.5)	28(1.5)	29(0.9)	29(0.9)	31(1.1)
Age 9	226(2.2)	221(2.2)	221(2.2)	220(4.1)	222(3.2)	230(1.8)	227(2.2)	226(2.7)
	—	—	25(2.0)	28(2.7)	28(1.6)	28(0.8)	27(0.8)	28(1.0)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval. [– – –] Extrapolated from previous NAEP analyses.

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores) or 1977 (for percentages), at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Scale Scores from 1977 to 1994 by Parents' Highest Level of Education

Table 1.2 presents trends in average science scores by parents' highest level of education. The percentage of students in each age group who reported that one or both parents had graduated from college increased between 1977 and 1994. Correspondingly, the percentage of students who reported high school graduation or less than high school graduation as their parents' highest level of education decreased during the same time period.

The average scores of 9-year-olds who reported that at least one parent had graduated from college increased between 1977 and 1994. This same pattern was observed for 9-year-olds who reported that neither parent had graduated from high school. The performance of 9-year-olds at the other levels of parental education remained relatively stable across the trend assessments.

Overall gains in average scale scores were observed for 13-year-olds at all levels of parental education. The overall gain for students who reported at least one parent graduated from college occurred in spite of a decline in the average score for this group in the 1980s. However, a higher average score in 1994 compared to 1977 was observed only among 13-year-olds who reported that neither parent had graduated from high school.

For 17-year-olds with at least one parent who graduated from college, a decline in the average score was observed between 1977 and 1982. Gains have been made since that time, and the overall pattern is one of increased performance. However, the 1994 average score did not differ significantly from the 1977 average score. The average score of 17-year-olds who reported high school graduation as their parents' highest level of education declined between 1977 and 1982, and has remained relatively stable since that time. In 1994, the average score of these students continued to be lower than the average score in 1977. The performance of 17-year-olds at the other levels of parental education showed little or no change across the assessment years.

Table 1.2

Trends in Average Science Scale Scores by Parents' Highest Level of Education, 1977 to 1994

Level of Education	Year	AGE 9		AGE 13		AGE 17	
		Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Graduated College	1994	45(0.8)†	239(1.4)†	46(1.3)†	269(1.3)	44(1.5)†	311(1.6)
	1992	42(1.2)†	239(1.2)†	44(1.3)†	269(1.0)	43(1.4)†	308(1.3)
	1990	40(1.1)*†	236(1.3)	41(1.2)*†	268(1.1)	39(1.4)*†	306(1.7)
	1986	38(1.1)*†	235(1.4)	38(2.0)*†	264(1.9)	37(1.2)*†	304(2.1)
	1982	42(2.3)†	231(2.3)*	37(1.5)*†	264(1.5)*	32(1.4)*	300(1.7)*†
	1977	23(0.7)*	232(1.4)*	27(1.0)*	266(1.0)	30(1.2)*	309(1.0)
Some Education After High School	1994	7(0.4)	239(2.8)	17(0.6)	260(2.0)	24(1.1)†	295(1.9)
	1992	8(0.4)	237(2.4)	18(0.7)†	266(1.1)†	25(0.9)†	296(1.7)
	1990	7(0.4)	238(2.1)	17(0.6)	263(1.2)	24(0.9)†	297(1.6)
	1986	7(0.6)	236(2.6)	16(0.6)	258(1.4)	24(1.0)†	295(2.5)
	1982	8(0.6)	229(3.2)	17(0.6)	259(1.5)	22(0.6)†	290(1.7)†
	1977	7(0.3)	237(1.5)	15(0.5)	260(1.3)	17(0.4)*	296(1.1)
Graduated High School	1994	14(0.6)†	225(1.4)	23(0.9)†	247(1.2)	22(0.8)†	279(1.7)†
	1992	14(0.7)†	222(1.9)	23(0.9)†	246(1.4)	21(0.9)†	280(2.4)
	1990	16(0.7)†	226(1.7)	27(0.8)*†	247(1.3)	26(1.1)*†	276(1.4)†
	1986	16(0.7)†	220(1.5)*	31(1.3)*	245(1.4)	28(1.1)*†	277(2.0)†
	1982	15(1.1)†	218(3.3)	26(1.1)†	243(1.3)	29(0.9)*†	275(1.6)†
	1977	27(0.5)*	223(1.4)	33(0.6)*	245(1.1)	33(0.6)*	284(0.8)*
Less Than High School	1994	4(0.4)†	211(3.4)†	6(0.4)†	234(2.5)†	7(0.5)†	256(4.2)
	1992	4(0.3)†	217(2.6)†	6(0.5)†	234(2.9)†	8(0.6)†	262(3.8)
	1990	5(0.4)†	210(2.7)†	8(0.5)†	233(2.1)†	8(0.6)†	261(2.8)
	1986	4(0.4)†	204(2.9)	8(1.1)†	229(2.7)	8(0.4)†	258(3.1)
	1982	7(0.9)*	198(6.0)	10(0.6)*†	225(1.9)*	13(0.7)*	259(2.4)
	1977	9(0.4)*	199(2.2)*	13(0.7)*	224(1.3)*	15(0.9)*	265(1.3)
I Don't Know	1994	30(0.8)†	223(1.9)†	8(0.5)†	230(2.5)†	3(0.3)†	247(6.7)
	1992	33(0.8)	224(1.4)†	8(0.4)†	232(2.0)†	3(0.3)†	258(7.4)
	1990	32(0.8)	222(1.2)†	8(0.5)†	224(2.1)	3(0.4)†	248(5.5)
	1986	35(1.0)*	215(1.5)*	8(0.4)†	227(2.7)	3(0.3)†	245(5.5)
	1982	29(1.8)†	211(2.8)*	11(1.2)	229(2.8)	5(0.8)	252(3.9)
	1977	34(0.7)*	211(1.4)*	13(1.1)*	222(1.8)*	4(0.4)*	253(3.2)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1977, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Scale Scores from 1977 to 1994 by Type of School

The percentage of students attending public and nonpublic schools remained approximately the same across the assessments for students in each age group (Table 1.3). At all three ages in 1994, the average scores of nonpublic school students were higher than the average scores of their public school peers.

The average scores of 9- and 13-year-old public school students increased from 1977 to 1994. For 17-year-old public school students, an early decline in performance observed between 1977 and 1982 has been reversed. In 1994 the average score for these students was higher than the average score in 1982, although it was not significantly different from the 1977 average.

Among nonpublic school students, the performance of 9-year-olds was relatively stable through the 1980s, but has since increased. However, the 1977 and 1994 averages for these students did not differ significantly. Little or no change has been observed across the assessments in the performance of 13-year-olds attending nonpublic schools. At age 17, a decline observed in average scores between 1977 and 1982 was subsequently reversed. However, the 1994 average for these nonpublic school students was not significantly different than the 1977 average.

Table 1.3

Trends in Average Science Scale Scores by Type of School, 1977 to 1994

		AGE 9		AGE 13		AGE 17	
Type of School	**Year**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**
Public	1994	88(1.8)	230(1.4)†	88(1.7)	255(1.1)†	88(2.3)	292(1.5)
	1992	87(1.6)	229(1.0)†	88(1.9)	257(1.0)†	90(2.4)	292(1.3)
	1990	89(2.1)	228(0.9)†	90(1.4)	254(1.1)†	93(1.8)	289(1.1)
	1986	84(2.7)	223(1.4)*	96(1.8)*	251(1.4)†	96(1.4)*	287(1.6)
	1982	90(2.3)	220(2.0)*	89(1.7)	249(1.4)*	90(2.0)	282(1.1)*†
	1977	89(1.2)	218(1.4)*	90(1.4)	245(1.2)*	94(1.8)	288(1.0)
Nonpublic	1994	12(1.8)	242(2.8)	12(1.7)	268(2.6)	12(2.3)	310(4.8)
	1992	13(1.6)	240(2.7)	12(1.9)	265(2.4)	9(2.1)	312(3.7)
	1990	11(2.1)	237(2.4)	10(1.4)	269(1.8)	7(1.8)	308(6.6)
	1986	16(2.7)	233(2.9)	4(1.8)*	263(6.4)	4(1.4)*	321(10.1)
	1982	10(2.3)	232(3.2)	11(1.7)	264(3.2)	10(2.0)	292(2.9)*†
	1977	11(1.2)	235(2.2)	10(1.4)	268(2.1)	6(1.8)	308(2.4)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1977, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Scale Scores from 1977 to 1994 by Modal Grade

Trends in average science scale scores by modal grade are presented in Table 1.4. The modal grade is the grade typically attended by students of a particular age. The modal grades are: grade 4 for age 9, grade 8 for age 13, and grade 11 for age 17. Students who are below modal grade are older than the expected age of students in their grade. Conversely, students who are above modal grade are younger than the expected age of students in their grade.

One interesting pattern is the change in the percentages of students below modal grade. Since the beginning of the trend comparisons by modal grade, the number of students below modal grade levels has increased in each age group.

The factors that have brought about changes in the modal grade level of 9-, 13-, and 17-year-olds have been the focus of numerous studies. Some have documented the increases in age requirements for entry into kindergarten classes in many school systems and states. In many cases, entrance age requirements have been increased or parents have chosen to enter their children later in order to maximize the potential for success in school for young children.[10] However, research in this area seems to indicate both positive and negative results of such practices.[11]

Since 1977, the performance of students in all age groups who were below modal grade has shown a pattern of overall improvement. For 9- and 13-year-olds below modal grade, the 1994 average scores were higher than the 1977 average scores. A significant difference was not observed for 17-year-olds. For students who were at modal grade, gains were observed between 1977 and 1994 for 9- and 13-year-olds. The performance of 17-year-olds at modal grade declined between 1977 and 1982, but has since improved. In 1994, the average score of these students was higher than that of their counterparts in 1977.

---

[10]Nurss, J. R. *Readiness for Kindergarten.* ERIC Digest, Office of Educational Research and Improvement. (Washington, DC: Government Printing Office, 1987).

[11]DeMeis, J. L., & Stearns, E. S., "Relationship of School Entrance Age to Academic and Social Performance," *Journal of Educational Research, 81,* 20-27, 1992.

Shepard, L. A., & Smith, M. L. "Escalating Academic Demand in Kindergarten: Counterproductive Policies," *Elementary School Journal, 89,* 135-145, 1988.

Table 1.4

Trends in Average Science Scale Scores by Modal Grade, 1977-1994

		Age 9		Age 13		Age 17	
Modal Grade	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Below Modal Grade	1994	33(1.3)†	215(1.7)†	38(1.3)†	244(1.5)†	21(1.6)†	262(3.4)
	1992	38(1.2)†	215(1.4)†	37(1.1)†	244(1.4)†	24(1.1)†	263(2.6)†
	1990	35(1.4)†	211(1.5)†	36(1.3)†	240(1.6)†	22(1.0)†	261(2.0)†
	1986	34(1.7)†	205(1.6)*†	33(2.1)	234(1.9)*†	17(0.9)	259(2.7)
	1982	30(1.9)†	198(2.9)*	28(1.3)*	229(1.6)*	16(1.0)*	251(2.2)*
	1977	24(1.0)*	198(1.6)*	27(0.9)*	223(1.6)*	14(0.6)*	253(1.4)
Modal Grade	1994	66(1.3)†	239(1.3)†	62(1.3)†	264(1.1)†	73(1.7)	302(1.3)†
	1992	62(1.2)†	240(1.0)†	62(1.0)†	266(1.0)†	70(1.0)†	304(1.2)†
	1990	65(1.4)†	238(1.0)†	63(1.4)†	264(1.0)†	70(1.0)†	299(1.0)†
	1986	66(1.7)†	234(1.2)*†	67(2.1)	260(1.3)	75(1.2)	294(1.6)*
	1982	70(1.9)	231(2.2)*	72(1.3)*	259(1.3)*	75(1.0)	289(1.1)*†
	1977	75(1.0)*	227(1.2)*	72(0.7)*	256(1.0)*	75(0.6)	295(0.9)*
Above Modal Grade	1994	0(0.1)	***(***)	***(***)	***(***)	6(0.6)†	303(4.2)
	1992	0(0.1)	***(***)	0(0.1)	***(***)	6(0.5)†	305(4.1)
	1990	0(0.1)	***(***)	1(0.2)	263(17.5)	8(0.6)†	298(2.5)
	1986	0(0.1)	***(***)	1(0.1)	266(6.3)	8(0.7)†	299(4.3)
	1982	1(0.2)	266(13.1)	0(0.1)	***(***)	9(0.7)*	293(2.6)†
	1977	1(0.1)	244(6.2)	1(0.4)	285(3.9)	11(0.5)*	301(1.5)

NOTE: The modal grades are: grade 4 at age 9, grade 8 at age 13, and grade 11 at age 17.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1977, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

Students in all three age groups demonstrated declines in science performance during the first few assessments, but have since improved. In 1994, these improvements resulted in an average score for 9-year-olds that was higher than that of 9-year-olds in 1970. Among 13-year-olds, recent gains have increased the average score to a level in 1994 that was not significantly different than the 1970 level. Despite the recent gains made by 17-year-olds, the 1994 average score for these students continued to be lower than the 1969 average.

The average science scores of 9- and 13-year-olds in the upper, middle two, and lower quartiles of the performance distribution increased between 1977 and 1994. This pattern was also observed among 17-year-olds in the upper and middle two quartiles. The 1994 average science score of 17-year-olds in the lower quartile was lower than that in 1977.

Earlier declines and more recent gains characterized the science performance of 9-, 13-, and 17-year-old White students across the assessment years. In 1994, the average scores of 9- and 13-year-olds were higher than the average scores of their counterparts in 1970. Despite the recent improvements made by 17-year-old White students, the 1994 average score was lower than the 1969 average score. Both 9- and 13-year-old Black and Hispanic students displayed overall gains across the trend assessments. However, a higher average score in 1994 compared to 1970 was observed only for Black 9-year-olds and Hispanic 13-year-olds. At age 17, Black students displayed a pattern of early declines followed by more recent gains. However, their average score in 1994 did not differ significantly from that in 1969. No significant changes across the assessment years were observed for 17-year-old Hispanic students.

In 1994, White students in all age groups continued to outperform their Black and Hispanic peers in science. Although there was some evidence that the gap between White and Black students has narrowed across the assessments, the magnitude of the difference in 1994 for 13- and 17-year-olds did not change significantly from that in the first assessment. For 9-year-olds, the gap between White and Black students' average science scores was smaller in 1994 than in 1970. The difference between White and Hispanic students aged 9 and 17 has shown no strong indication of change across the assessments. However, at age 13 the difference between White and Hispanic students has decreased.

Both male and female students at all ages demonstrated a pattern of declines during the 1970s (and in some cases the early 1980s), followed by a period of increased performance in science. For 9- and 13-year-old males, recent gains have not resulted in a 1994 average score that was significantly different from the 1970 average. For 17-year-old males, the average score in 1994 continued to be lower than that in 1969. A comparison of the 1994 average scores with those in the first assessment showed that 9-year-old females had a higher average in 1994, 13-year-old females demonstrated no significant difference, and 17-year-old females had lower scores in 1994. Despite some evidence of decreasing gender gaps across the assessments among 17-year- olds, male students outperformed female students at ages 13 and 17 in 1994.

In the Northeast and Central regions, students in all three age groups displayed a pattern of early declines followed by increased science performance. However, their 1994 averages did not differ significantly from those in the first trend assessment. In the Southeast, the average scores of 9- and 13-year-olds also declined and then increased, and in 1994 their average scores were higher than the average scores of their counterparts in 1970. Despite recent gains made by 17-year-olds in the Southeast, their 1994 average score was not significantly different from that observed in 1969. Although 9- and 13-year-olds in the West region have shown improved science performance since the early 1980s, their 1994 average scores did not differ significantly from 1970 averages. For 17-year- olds in the West region, a period of increased science performance was also observed since the early 1980s. However, their 1994 average score was lower than that in 1969.

Increases in average science scores between 1977 and 1994 were observed for 9-year-olds who reported that at least one parent graduated from college or who reported that neither parent had graduated from high school. Overall gains were observed for 13-year-olds at all levels of parental education. However, a higher average score in 1994 compared to 1977 was observed only for 13-year-olds who reported that neither parent had graduated from high school. An overall pattern of increased performance was observed for 17-year-olds with at least one parent who graduated from college, although there was no significant difference between 1977 and 1994 average scores. A decline in performance between 1977 and 1994 was observed for 17-year-olds whose parents' highest level of education was high school graduation.

The average science scores of 9- and 13-year-old public school students increased between 1977 and 1994. Although 17-year-old public school students have made gains since a decline in 1982, their 1994 average score was not significantly different than the 1977 average score. Despite some evidence of recent improvement among 9-year-old nonpublic school students, in 1994 their average score and the average score of 13-year-old nonpublic school students did not differ significantly from 1977 averages. The average score of 17-year-old nonpublic school students declined in 1982, but was not significantly different in 1994 than in 1977.

At ages 9 and 13, students who were below modal grade showed increases in average science scores since 1977. The pattern for 17-year-olds below modal grade was also one of overall improvement, although no significant difference was observed between average scores in 1977 and 1994. For students who were at modal grade, an increase in the average science score was observed for each age group.

2

National Trends in Levels of Science Performance from 1977 to 1994

To provide more information about students' knowledge and skills in science, five levels of performance were established on the science trend scale — 150, 200, 250, 300, and 350. Performance was anchored at the five levels by using empirical procedures that identified sets of assessment questions that students who performed at one level were more likely to answer correctly than students who performed at the next lower level. The types of knowledge and skills that these sets of questions assessed were then identified and used as a basis for constructing the descriptions of performance at the five scale levels (see Figure 2.1).

Figure 2.1 — Levels of Science Performance

Level 350: Integrates Specialized Scientific Information

Students at this level can infer relationships and draw conclusions using detailed scientific knowledge from the physical sciences, particularly chemistry. They also can apply basic principles of genetics and interpret the social implications of research in this field.

Level 300: Analyzes Scientific Procedures and Data

Students at this level can evaluate the appropriateness of the design of an experiment. They have more detailed scientific knowledge and the skill to apply their knowledge in interpreting information from text and graphs. These students also exhibit a growing understanding of principles from the physical sciences.

Level 250: Applies General Scientific Information

Students at this level can interpret data from simple tables and make inferences about the outcomes of experimental procedures. They exhibit knowledge and understanding of the life sciences, including a familiarity with some aspects of animal behavior and of ecological relationships. These students also demonstrate some knowledge of basic information from the physical sciences.

Level 200: Understands Simple Scientific Principles

Students at this level are developing some understanding of simple scientific principles, particularly in the life sciences. For example, they exhibit some rudimentary knowledge of the structure and function of plants and animals.

Level 150: Knows Everyday Science Facts

Students at this level know some general scientific facts of the type that could be learned from everyday experiences. They can read simple graphs, match the distinguishing characteristics of animals, and predict the operation of familiar apparatuses that work according to mechanical principles.

Table 2.1 presents the percentages of students performing at or above the five science performance levels in the six assessments conducted from 1977. (Performance level data are not available for assessment years with extrapolated results.)

Table 2.1

Trends in Percentages of Students At or Above Five Science Performance Levels, 1977 to 1994

		ASSESSMENT YEARS					
Performance Levels	**Age**	**1977**	**1982**	**1986**	**1990**	**1992**	**1994**
Level 350	9	0(0.0)	0(0.1)	0(0.1)	0(0.0)	0(0.1)	0(0.0)
Integrates Specialized	13	1(0.1)*	0(0.1)	0(0.1)†	0(0.1)	0(0.1)†	0(0.1)†
Scientific Information	17	9(0.4)	7(0.4)*	8(0.7)	9(0.5)	10(0.7)	10(0.8)
Level 300	9	3(0.3)	2(0.7)	3(0.5)	3(0.3)	3(0.3)	4(0.4)
Analyzes Scientific	13	11(0.5)	10(0.7)	9(0.9)	11(0.6)	12(0.8)	12(0.9)
Procedures and Data	17	42(0.9)*	37(0.9)*†	41(1.4)*	43(1.3)	47(1.5)†	48(1.3)†
Level 250	9	26(0.7)*	24(1.8)*	28(1.4)*	31(0.8)†	33(1.0)†	34(1.2)†
Applies General	13	49(1.1)*	51(1.6)*	53(1.6)*	57(1.0)†	61(1.1)†	60(1.1)†
Scientific Information	17	82(0.7)	77(1.0)*†	81(1.3)	81(0.9)	83(1.2)	83(1.2)
Level 200	9	68(1.1)*	71(1.9)*	72(1.1)*†	76(0.9)†	78(1.2)†	77(1.0)†
Understands Simple	13	86(0.7)*	90(0.8)*†	92(1.0)†	92(0.7)†	93(0.5)†	92(0.6)†
Scientific Principles	17	97(0.2)	96(0.5)	97(0.5)	97(0.3)	98(0.5)	97(0.7)
Level 150	9	94(0.6)*	95(0.7)	96(0.3)†	97(0.3)†	97(0.3)†	97(0.4)†
Knows Everyday	13	99(0.2)*	100(0.1)†	100(0.1)†	100(0.1)†	100(0.1)†	100(0.1)†
Science Facts	17	100(0.0)	100(0.1)	100(0.1)	100(0.2)	100(0.0)†	100(0.1)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1977, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Level 350: After a slight decline in the percentage of 17-year-olds attaining this highest level of performance in the early 1980s, characterized by the ability to integrate specialized scientific information, there was evidence of an increase. However, the 1994 percentage did not differ significantly from the 1977 percentage.

Level 300: For both 13- and 17-year-olds, there was evidence of early declines followed by increases in the percentage of students reaching at least Level 300. At this level, students' performance was characterized by the ability to analyze scientific procedures and data. The 1994 percentage of students at or above this level was higher than the 1977 percentage for 17-year-olds, but not for 13-year-olds.

Level 250: After a decline between 1977 and 1982, the percentage of 17-year-olds able to apply general scientific information (at or above Level 250) increased. However, the percentage in 1994 was not significantly different from the percentage in 1977. An increase between 1977 and 1994 was observed in the percentage of 9- and 13-year-olds at or above this performance level.

Level 200: Most 17-year-olds performed at or above this level in 1994 and in all previous assessments, demonstrating understanding of simple scientific procedures. The percentage of 13-year-olds reaching at least Level 200 has increased since 1977, although the percentage has been relatively stable since 1986. At age 9, there was an increase between 1977 and 1994 in the percentage of students at or above this level.

Level 150: In 1994, as in all previous trend assessments, most students at all three ages demonstrated knowledge of general scientific facts and an ability to perform tasks as described in Level 150. At age 9, an increase between 1977 and 1994 was observed in the percentage of students attaining at least this level of performance.

Trends in Levels of Science Performance between 1977 and 1994 by Race/Ethnicity

Table 2.2 depicts the percentages of White, Black, and Hispanic students at or above each of the performance levels for the assessment years 1977 and 1994.

Table 2.2

Trends in Percentages of Students At or Above Five Science Performance Levels by Race/Ethnicity, 1977 and 1994

		ASSESSMENT YEARS					
		1977			1994		
Performance Levels	**Age**	**White**	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**
Level 350	9	0(0.0)	0(0.0)	0(0.0)	0(0.1)	0(0.0)	0(0.0)
Integrates Specialized	13	1(0.1)*	0(0.0)	0(0.1)	0(0.1)	0(0.0)	0(0.0)
Scientific Information	17	10(0.4)*	0(0.2)	2(0.6)	13(1.1)	1(0.3)	2(0.7)
Level 300	9	4(0.3)	0(0.1)	0(0.4)	5(0.6)	0(0.4)	1(0.5)
Analyzes Scientific	13	13(0.5)	1(0.4)	2(0.8)	15(1.0)	2(2.1)	2(0.9)
Procedures and Data	17	48(0.7)*	8(1.0)*	19(2.1)	58(1.6)	15(2.3)	22(4.1)
Level 250	9	31(0.7)*	4(0.6)*	9(1.7)	41(1.5)	11(1.4)	11(2.5)
Applies General	13	57(0.9)*	15(1.7)	18(1.8)*	71(1.1)	22(4.3)	32(3.3)
Scientific Information	17	88(0.4)*	41(1.5)*	62(1.7)	92(0.9)	58(3.7)	59(7.4)
Level 200	9	77(0.7)*	27(1.5)*	42(3.1)	86(1.0)	52(2.3)	50(3.1)
Understands Simple	13	92(0.5)*	57(2.4)*	62(2.4)*	98(0.4)	74(3.2)	81(2.5)
Scientific Principles	17	99(0.1)	84(1.3)*	93(1.7)	99(0.3)	91(1.9)	90(3.3)
Level 150	9	98(0.3)*	72(1.8)*	85(1.8)*	99(0.3)	91(1.5)	91(2.3)
Knows Everyday	13	100(0.1)*	93(1.0)*	94(1.3)*	100(0.0)	99(0.6)	99(0.4)
Science Facts	17	100(0.0)	99(0.3)	100(0.2)	100(0.0)	100(0.5)	99(0.5)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The percentages of 9-year-old White and Black students performing at or above Levels 150, 200, and 250 increased between 1977 and 1994. The percentage of 9-year-old Hispanic students at or above Level 150 also increased. In 1994, a greater percentage of White students than Black or Hispanic students performed at or above Levels 150, 200, 250, and 300.

Between 1977 and 1994 the percentages of 13-year-old White and Hispanic students performing at or above Levels 150, 200, and 250 increased. In addition, there was an increase in the percentage of Black 13-year-olds performing at or above Levels 150 and 200. In 1994, the percentage of White students reaching Levels 200, 250, and 300 was higher than the percentage of Black or Hispanic students who did so.

For 17-year-old students, comparisons of performance between 1977 and 1994 yielded mixed results. At Levels 250, 300, and 350, the percentage of White students increased between the first and most recent assessment years. The percentage of 17-year-old Black students performing at or above Levels 200, 250, and 300 also increased. The percentage of Hispanic 17-year-olds performing at or above all five levels did not differ significantly in 1994 when compared to 1977. Comparing the 1994 performance of 17-year-old students in the three racial/ethnic groups, a greater percentage of White students than Black or Hispanic students performed at or above Levels 200, 250, 300, and 350.

Trends in Levels of Science Performance between 1977 and 1994 by Gender

Table 2.3 shows the percentages of male and female students at or above each of the five performance levels in 1977 and 1994.

Between 1977 and 1994, there were increases in the percentages of 9- and 13-year-old males and females performing at or above Levels 150, 200, and 250 and in the percentage of 17-year-old females performing at or above Level 300.

Table 2.3

Trends in Percentages of Students At or Above Five Science Performance Levels by Gender, 1977 and 1994

		ASSESSMENT YEARS			
		1977		1994	
Performance Levels	Age	Male	Female	Male	Female
Level 350	9	0(0.0)	0(0.0)	0(0.1)	0(0.0)
Integrates Specialized	13	1(0.2)*	0(0.1)*	0(0.2)	0(0.1)
Scientific Information	17	12(0.6)	5(0.4)	14(1.2)	6(0.6)
Level 300	9	4(0.3)	3(0.3)	5(0.7)	3(0.4)
Analyzes Scientific	13	13(0.6)	9(0.5)	15(1.1)	9(1.0)
Procedures and Data	17	49(1.1)	35(1.0)*	53(1.8)	42(1.8)
Level 250	9	27(0.9)*	24(0.9)*	35(1.4)	32(1.5)
Applies General	13	52(1.3)*	45(1.2)*	62(1.3)	57(1.4)
Scientific Information	17	85(0.7)	78(1.0)	85(1.3)	82(1.6)
Level 200	9	70(1.2)*	67(1.1)*	78(0.9)	77(1.4)
Understands Simple	13	87(0.8)*	85(0.8)*	92(0.8)	93(0.6)
Scientific Principles	17	98(0.2)	96(0.3)	97(0.6)	97(1.0)
Level 150	9	94(0.5)*	93(0.7)*	97(0.4)	97(0.5)
Knows Everyday	13	99(0.2)*	98(0.2)*	100(0.1)	100(0.2)
Science Facts	17	100(0.0)	100(0.1)	100(0.2)	100(0.1)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The percentages of 9-year-old students at or above Levels 150, 200, and 250 in science showed a pattern of overall increases since 1977. Also, 1994 percentages of 9-year-olds at these performance levels were higher than the 1977 percentages. More White and Black 9-year-olds reached Levels 150, 200, and 250 in 1994 than in 1977; an increase was also seen for Hispanic students at Level 150. In 1994 White students continued to outperform their Black and Hispanic counterparts at Levels 150, 200, 250, and 300. The percentages of males and females reaching Levels 150, 200, and 250 were higher in 1994 than in 1977.

There was a pattern of overall increases in the percentages of 13-year-old students at or above Levels 200, 250, and 300 in science across the assessment years. At Levels 200 and 250, the 1994 percentages were higher than the 1977 percentages. There were increases between 1977 and 1994 in the percentages of White, Black, and Hispanic 13-year-old students at Levels 150 and 200 and in the percentages White and Hispanic students at Level 250. In 1994, White 13-year-old students outperformed Black and Hispanic students except at Level 150, where almost all 13-year-olds from each racial/ethnic group demonstrated knowledge of everyday science facts, and at Level 350, where very few students from any group were able to integrate specialized scientific information. Higher percentages of males and females were at or above Levels 150, 200, and 250 in 1994 than in 1977.

The percentages of 17-year-old students at Levels 250, 300, and 350 in science showed a pattern of increase from 1977 to 1994. However, the 1994 percentage of students was higher than the 1977 percentage only at Level 300. Compared to 1977, the percentages of White students at or above Levels 250, 300, and 350 in science were higher in 1994. The percentages of 17-year-old Black students at or above all levels except Levels 150 and 350 increased between 1977 and 1994. For the 1994 assessment, the percentages of White students at or above Levels 200, 250, 300, and 350 were higher than that of Black and Hispanic students. In 1994, the percentages of 17-year-old Black and Hispanic students able to integrate specialized scientific information (Level 350) remained at 2 percent or less; in contrast, 13 percent of White students reached this level. Seventeen-year-old female students showed improvement at Level 300 between 1977 and 1994. No improvement was seen for male 17-year-olds at any science performance level.

3

Trends in Students' Experiences in Science and Attitudes Toward Science

Introduction

Many experiences, such as the number and types of science courses taken in school, exposure to different modes of teaching and learning, and perception of the role of science in individual lives and in world affairs, are necessary to produce scientifically literate individuals who are ready to meet the challenges of the 21st century.[12] This chapter looks at the relationship between self-reported student experiences and average science scale scores.

[12] *Educating Americans for the 21st Century: A Report to the American People and the National Science Board* (Washington, DC: National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, 1983).

Statewide Systemic Initiatives in Science, Mathematics, and Engineering (Arlington, VA: The National Science Foundation, 1995-1996).

Benchmarks for Science Literacy (Washington, DC: Project 2061, American Association for the Advancement of Science, 1993).

Clinton, William J. & Gore, A., *Science in the National Interest* (Washington, DC: Executive Office of the President, Office of Science and Technology Policy, 1994).

National Science Education Standards (Washington, DC: National Research Council, 1995).

Trends in 9-Year-Olds' Participation in Science Activities and Equipment Use between 1977 and 1994

The central role of investigation in science teaching and learning has received much attention in recent years.[13] Hands-on experiences and the use of common science instruments are necessary parts of scientific investigation. Table 3.1 summarizes trends between 1977 and 1994 in 9-year-old students' reported participation in three types of experiments and their use of specific instruments. The table also displays average scale scores.

In 1994, 66 percent of 9-year-olds reported experimenting with plants, as compared to 70 percent in 1977, and 41 percent reported experimenting with animals, as compared to 55 percent in 1977. In 1994, a high percentage (89 percent to 97 percent) of 9-year-olds reported use of scales, thermometers, and calculators. The percentages of students using thermometers and calculators were higher in 1994 than in 1977, while the percentage who reported using a scale showed no significant difference. Seventy-five percent of 9-year-olds in 1994 reported using stopwatches and compasses, and 58 percent reported using microscopes. These percentages were higher than in 1977. No significant change in the percentage of students who reported experimenting with batteries and bulbs was observed over the same time period.

In 1994, students who answered in the affirmative each question concerning experimentation and equipment had higher average scale scores than those who answered in the negative, with two exceptions: experimentation with living animals and experimentation with batteries and bulbs. (A comparison could not be made between students' responses to the question about calculator use due to the insufficient sample size of students responding "no" in 1994.)

---

[13] *National Science Education Standards* (Washington, DC: National Research Council, 1995).

Benchmarks for Science Literacy (Washington, DC: Project 2061, American Association for the Advancement of Science, 1993).

Table 3.1

Trends in Participation in Science Activities at Age 9, 1977 and 1994

		STUDENTS ANSWERING "YES"		STUDENTS ANSWERING "NO"	
Have you ever . . .	**Year**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**
Experimented with living plants?	1994	66(1.5)	237(1.8)	30(1.4)	226(3.4)
	1977	70(1.4)*	221(2.3)*	27(1.3)	217(2.8)*
Experimented with living animals?	1994	41(1.2)	231(2.7)	55(1.4)	236(1.9)
	1977	55(1.5)*	216(2.8)*	42(1.3)*	227(2.1)*
Experimented with batteries & bulbs?	1994	51(1.7)	237(2.1)	43(1.7)	232(2.6)
	1977	51(1.4)	225(2.8)*	43(1.4)	217(2.1)*
Used a scale to weigh things?	1994	91(1.0)	234(1.5)	7(0.8)	213(4.0)
	1977	89(0.8)	220(2.3)*	9(0.7)	202(4.5)
Used a thermometer?	1994	89(1.1)	234(1.5)	9(0.8)	214(4.7)
	1977	84(1.0)*	222(2.2)*	14(0.9)*	199(2.7)*
Used a microscope?	1994	58(1.8)	236(2.0)	36(1.8)	225(1.9)
	1977	53(1.4)*	222(2.5)*	43(1.5)*	214(2.1)*
Used a calculator?	1994	97(0.4)	232(1.5)	***(***)	***(***)
	1977	87(1.2)*	222(2.2)*	11(1.0)	195(3.4)
Used a compass?	1994	75(1.3)	236(1.6)	21(1.2)	220(2.2)
	1977	61(1.3)*	222(2.3)*	33(1.2)*	214(2.7)
Used a stopwatch?	1994	75(1.3)	235(1.4)	21(1.1)	223(2.7)
	1977	44(1.3)*	223(2.6)*	49(1.2)*	215(2.5)*

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "not certain" by a small percentage of students.

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Science Course Taking at Age 17 from 1986 to 1994

Since 1986, NAEP has gathered information to determine the percentages of students studying certain science subjects. Many school curricula follow the sequence biology, chemistry, physics. Therefore, most students have studied biology by the time they are 17 years old. Table 3.2 presents the percentages of 17-year-old students taking general science, biology, chemistry, and physics, and their average scale scores. Results are presented for the nation and for male and female students.

In 1994, almost all 17-year-olds (93 percent) reported taking biology; slightly more than half (53 percent) reported taking chemistry; and 18 percent reported taking physics. The 1994 percentages for the nation represent increases over 1986 percentages. These results are similar to those from other studies documenting a trend toward more advanced course work among high school seniors.[14] The average scale scores of students taking general science, biology, or physics have increased over the assessment years, from 1986 to 1994.

Reflecting results for the nation, the percentages of male and female students taking biology, chemistry, or physics also have increased since 1986. The only statistically significant difference between male and female 17-year-olds' reports on course taking in 1994 was a higher percentage of male students taking physics. Trend analyses of average scale scores across assessment years 1986 to 1994 indicated a pattern of increasing scale scores for female students taking general science; for males and females taking biology; for females taking chemistry; and for females taking physics.

---

[14] Blank, R. K., & Gruebel, D., *State Indicators of Science and Math Education 1995: State-by-State Trends and New Indicators from the 1993-94 School Year.* (Washington, DC: Council of Chief State School Officers, 1995).

Smith, T. M.; Young, B.A.; Choy, S. P.; Perie, M.; Alsalam, N.; Rollefson, M. R., & Bae, Y., *The Condition of Education 1996.* (Washington, DC: National Center for Education Statistics, 1996).

Table 3.2

Trends in Science Course Taking at Age 17 from 1986 to 1994 for the Nation and by Gender

	TOTAL		MALE		FEMALE	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
General Science						
1994	83(1.3)	296(1.6)†	84(1.5)	302(2.2)	82(1.4)	291(1.8)†
1992	84(1.0)	297(1.3)†	86(1.1)	301(1.6)	83(1.5)	290(1.5)†
1990	82(1.3)	292(1.1)	84(1.3)	298(1.4)	81(1.7)	286(1.4)
1986	83(1.3)	290(1.3)*	84(1.5)	298(1.7)	82(1.6)	283(1.6)*
Biology						
1994	93(0.9)†	300(1.2)†	92(0.9)†	306(1.8)	93(1.0)†	294(1.3)†
1992	92(0.9)†	299(1.1)†	91(1.2)†	305(1.5)	93(1.0)†	293(1.4)†
1990	89(0.9)*	296(1.0)*	87(1.1)*	302(1.3)	91(1.0)	290(1.5)
1986	88(1.0)*	294(1.5)*	87(1.1)*	301(1.8)	88(1.1)*	287(1.7)*
Chemistry						
1994	53(2.1)†	315(1.7)	50(2.6)†	322(2.4)	55(2.3)†	309(1.9)
1992	49(1.7)†	319(1.0)†	47(1.9)	325(1.5)	51(2.0)†	313(1.5)†
1990	45(1.5)*	316(1.4)	45(1.7)	324(1.9)	45(1.7)*	310(1.7)
1986	40(1.6)*	312(2.1)	42(1.8)*	319(2.7)	39(2.1)*	304(2.2)
Physics						
1994	18(1.2)†	314(2.9)†	20(1.5)†	318(4.1)	16(1.3)†	310(3.3)†
1992	14(1.1)*	306(3.9)	15(1.0)*	310(4.7)	12(1.5)	302(4.1)†
1990	14(1.5)	304(3.7)	16(1.8)	311(4.3)	13(1.5)†	295(4.2)*
1986	11(0.9)*	296(4.7)*	14(1.3)*	305(6.8)	8(0.7)*	282(3.8)*

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1986, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 3.3 presents trends in science course taking by race/ethnicity. Increases in the percentages of White students taking biology, chemistry, or physics were observed across the assessment years. The percentages of Black students taking biology and chemistry, and the percentages of Hispanic students taking chemistry, also increased over time. In 1994, a higher percentage of White students than Hispanic students reported taking general science, biology, or chemistry classes. There was no significant difference in course taking between White and Black students.

Trend analyses of scale scores revealed a pattern of increasing average scores for White students taking general science, biology, or chemistry, and for Black students taking physics. In 1994, White students outperformed their Black and Hispanic peers at each level of science course work. In interpreting these findings, it should be considered that science courses with the same topic may vary in content and instructional approach from school to school and from state to state.

Table 3.3

Trends in Science Course Taking at Age 17 from 1986 to 1994 by Race/Ethnicity

	WHITE		BLACK		HISPANIC	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
General Science						
1994	84(1.5)	306(1.7)†	80(1.9)	259(3.3)	77(2.4)	268(5.1)
1992	86(1.0)	304(1.3)†	79(3.6)	259(3.9)	79(3.2)	274(5.4)
1990	84(1.4)	300(1.1)*	76(3.1)	258(4.5)	82(4.4)	266(4.8)
1986	84(1.6)	297(1.5)*	83(2.6)	257(2.8)	82(3.5)	264(4.5)
Biology						
1994	94(0.9)†	310(1.3)†	93(1.8)†	263(2.7)	84(3.4)	273(6.1)
1992	93(1.0)†	308(1.1)†	92(1.9)†	260(3.1)	87(4.1)	276(4.5)
1990	90(0.9)*	304(1.0)*	87(2.2)	260(4.6)	79(4.4)	270(5.0)
1986	89(1.1)*	301(1.8)*	84(2.7)*	260(3.1)	84(3.4)	265(3.7)
Chemistry						
1994	54(2.5)†	324(1.7)†	51(3.6)†	278(3.4)	41(3.0)†	288(6.3)
1992	52(1.8)†	325(1.3)†	36(3.2)*	282(3.6)	36(5.6)	298(4.1)
1990	46(1.7)*	325(1.3)†	46(4.0)†	280(7.3)	31(4.3)	295(6.0)
1986	43(1.8)*	317(2.2)*	29(2.6)*	275(6.4)	24(2.2)*	281(8.7)
Physics						
1994	18(1.4)†	326(3.2)	16(2.0)	268(7.5)†	***(***)	***(***)
1992	13(1.2)*	319(3.5)	14(1.9)	251(7.4)	13(2.3)	282(11.1)
1990	13(1.7)	317(2.6)	15(2.7)	263(11.8)	17(4.5)	253(18.3)
1986	10(0.8)*	316(4.4)	18(3.5)	239(5.4)*	13(2.8)	257(17.6)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1986, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Attitudes about the Value of Science at Ages 13 and 17 between 1977 and 1994

Thirteen- and 17-year-old students were asked whether they agreed or disagreed with three statements about the value of science (Table 3.4). To determine whether attitudes have changed over time, the percentages of students in 1994 who agreed with these statements about the value of science were compared to the percentages of students in 1977 who agreed with the statements. Increases were seen in the percentages of 17-year-olds who agreed that much of what was learned in science classes was useful in the future and that science should be required in school.

Table 3.4

Trends in Attitudes About the Value of Science at Ages 13 and 17, 1977 and 1994

			STRONGLY AGREE OR AGREE		UNDECIDED, DISAGREE, OR STRONGLY DISAGREE	
	Age	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Much of what you learn in science classes is useful in everyday life.	13	1994	54(1.5)	258(1.8)	46(1.5)	257(1.8)
		1977	58(1.4)	249(2.3)*	42(1.4)	256(2.1)
	17	1994	56(1.6)	296(2.1)	44(1.6)	291(2.5)
		1977	53(1.2)	290(2.4)	47(1.2)	293(1.8)
Much of what you learn in science classes will be useful in the future.	13	1994	74(1.4)	258(1.7)	26(1.4)	253(2.2)
		1977	74(1.2)	251(2.1)*	26(1.2)	255(2.8)
	17	1994	70(1.4)	297(1.9)	30(1.4)	288(2.7)
		1977	65(1.3)*	292(2.0)	35(1.3)*	290(2.0)
Science should be required in school.	13	1994	73(1.3)	260(1.8)	27(1.3)	251(2.5)
		1977	70(1.2)	252(2.1)*	30(1.2)	252(2.5)
	17	1994	78(1.3)	298(2.0)	22(1.3)	281(2.6)
		1977	62(1.1)*	292(2.0)	38(1.1)*	291(2.4)*

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

A comparison of the average scores of students who agreed and disagreed with the statements revealed some statistically significant differences. Among 17-year-olds in 1994, students who agreed that science learning will be useful in the future had higher average scores than their peers who disagreed with this statement. For both 13- and 17-year-olds in 1994, the average scores of students who agreed that science should be required in school were higher than the average scores of students who disagreed with this statement.

Thirteen- and 17-year-old students were also asked to respond to questions about applications of science. Table 3.5 depicts the results. The percentages of 13-year-olds who agreed "very much" with the statements were higher in 1994 than in 1977, except for two statements: the statement about finding cures for disease, where no difference was seen; and the statement about preventing world starvation, where a significant decrease in the percentage of 13-year-olds was observed. The percentages of 17-year-olds who thought that application of science could help with energy shortages, birth defects, natural resources, and pollution were higher in 1994 than in 1977. Similar to 13-year-olds, there was a decrease in the percentage of 17-year-olds who agreed that science could prevent world starvation. Also, 17-year-olds in 1994 were less likely than those in 1977 to indicate that science could help reduce overpopulation.

Table 3.5

Trends in Perceived Applications of Science at Ages 13 and 17, 1977 and 1994

How much do you think that the application of science can help . . .	Year	PERCENT RESPONDING "VERY MUCH"			
		Age 13		Age 17	
		Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Prevent world starvation?	1994	18(0.7)	260(3.1)	23(1.4)	310(4.1)
	1977	32(1.5)*	251(2.7)*	51(1.2)*	298(1.9)*
Save us from an energy shortage?	1994	64(1.3)	265(1.8)	75(1.4)	303(2.1)
	1977	54(1.7)*	256(2.4)*	70(1.0)*	295(1.7)*
Find cures for diseases?	1994	71(1.1)	264(2.0)	85(1.3)	301(2.0)
	1977	70(1.5)	253(2.3)*	85(0.8)	292(1.6)*
Control weather?	1994	22(0.9)	263(2.1)	18(1.1)	298(3.1)
	1977	15(0.9)*	246(4.0)*	16(0.8)	296(2.7)
Prevent birth defects?	1994	36(1.6)	267(1.9)	55(1.7)	306(2.0)
	1977	23(1.2)*	257(3.1)*	44(1.2)*	297(1.9)*
Save our natural resources?	1994	60(1.1)	260(1.8)	59(1.3)	298(2.3)
	1977	47(1.1)*	252(2.5)*	48(1.2)*	293(2.0)
Reduce air and water pollution?	1994	55(1.4)	262(2.2)	63(1.3)	302(2.5)
	1977	44(1.2)*	253(2.4)*	54(1.2)*	296(1.8)*
Reduce overpopulation?	1994	15(0.9)	254(4.5)	13(0.9)	301(4.8)
	1977	11(0.8)*	249(3.7)	22(0.8)*	296(2.0)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The percentages of 9-year-old students who reported using some types of science instruments increased between 1977 and 1994, whereas the percentages of 9-year-old students who reported experimenting with plants and animals decreased. The average scale score of all 9-year-olds increased between the first and last assessment years regardless of whether or not they had participated in science activities, with two exceptions: students who reported not using a scale to weigh things, and students who reported not using a compass. The average scale score of students who participated in most of the science activities tended to be higher than those who did not.

Trend analyses across the assessment years showed the following overall increases in the percentages of 17-year-old students studying various sciences: males, females, and White students taking biology, chemistry, or physics; Black students taking biology or chemistry; and Hispanic students taking chemistry. While the percentages of students taking different science courses rose across the years, average scale scores did not always show the same pattern. Increases across the years in average scale scores were found in the following instances: males taking biology; females taking all subjects; White students taking general science, biology, or chemistry; and Black students taking physics.

The percentages of 17-year-olds who believed that much of what was learned in science classes will be useful in the future, and that science should be required in schools, increased since 1977. In 1994, 17-year-olds who agreed that science learning will be useful in the future had significantly higher scores than 17-year-olds who disagreed with this statement. Also, both 13- and 17-year-olds who agreed that science should be required in school had higher scores than their peers who did not agree.

The percentages of 13- and 17-year-old students who believed that science can help solve several societal concerns were higher in 1994 than in 1977, although there were some exceptions. Thirteen- and 17-year-olds were less likely to believe that the application of science can prevent world starvation, and 17-year-olds were less likely to believe that the application of science can help reduce overpopulation. The percentages of students at both ages who believed that science can find cures for diseases did not change significantly between 1977 and 1994, nor did the percentage of 17-year-olds who believed that science could control weather.

Part II

Trends in Mathematics Achievement from 1973 to 1994

Introduction

Since the establishment of standards in school mathematics in 1989, much attention has been given to the discipline and its role in the school curriculum. This attention has resulted in the reworking of school curricula and teaching programs, increased focus on faculty development in mathematics, and advances in assessing student progress in the subject.[15] As we approach the year 2000, eyes are beginning to focus on what effects, if any, these efforts have had on student achievement and improved practices in the classroom.

The 1994 NAEP long-term trend assessment in mathematics is one of many programs that can shed light on these questions. This program, initiated in 1973, provides a baseline look at long-term trends in student mathematics performance. The NAEP 1994 long-term trend assessment in

[15] Garet, M. S. & Mills, V. L., "Changes in Teaching Practices: The Effects of the Curriculum and Evaluation Standards," *Mathematics Teacher, 88*, 380-388 (1995).

Joyner, J. M., "Implementing the Assessment Standards for School Mathematics: NCTM's Assessment Standards: A Document for all Educators," *Teaching Children Mathematics, 2*, 20-22 (1995).

Lindquist, M. M., "Tides of Change: Teachers at the Helm," *Arithmetic Teacher, 41*, 64-68 (1993).

mathematics was the seventh of its kind, with previous assessments conducted in the 1972-73, 1977-78, 1982-83, 1985-86, 1989-90, and 1991-92 school years.[16] Each of these mathematics assessments, which will subsequently be referred to by the last half of the school year in which they occurred, involved a nationally representative sample of 9-, 13-, and 17-year-old students. It should be noted that some of the analyses reported in this section do not go back to the first mathematics trend assessment because the data are not available.

Unlike the NAEP mathematics assessments conducted in 1990 and 1992, to collect national data for students of grades 4, 8, and 12 and state data for grades 4 and 8[17], the long-term trend assessments replicate NAEP's initial data-gathering process of sampling students from across the country at ages 9, 13, and 17. Another difference is that the long-term trend assessments employ a different set of questions, reflecting a more limited view of the curriculum than the assessments newly developed for the 1990, 1992, and 1996 NAEP national- and state-level mathematics assessments.[18] These newly developed assessments focus more heavily on students' performance and associated achievement levels related to use of manipulatives and performance on constructed-response questions. They also contain extended sets of background questions describing the context of students' mathematics learning experiences, both in and out of school. Because the content of the national- and state- level NAEP mathematics series differs from that of the long-term trend assessment, and because the populations differ due to the age-versus-grade sampling methods, the results of the two assessments are not directly comparable.

The present work provides a supporting picture of school achievement in a time of reform and change. While the new short-term trends associated with the national- and state-level NAEP work provides a glimpse of change and progress by grade levels, the long-term trend study provides a picture of how 9-, 13-, and 17-year-olds are performing on a set of questions developed to measure long-held objectives for school mathematics. These mathematics objectives were set in the late 1960s. Today, they represent a somewhat constrained view of mathematics. The resulting assessment is more heavily weighted toward students' abilities to perform basic facts, carry out

---

[16] Mullis, I. V. S.; Dossey, J. A.; Campbell, J. R.; Gentile, C. A.; O'Sullivan, C.; & Latham, A. *NAEP 1992 Trends in Academic Progress* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1994).

[17] Mullis, I. V. S.; Dossey, J. A.; Owen, E. H.; & Phillips, G. W. *NAEP 1992 Mathematics Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1993).

[18] National Assessment of Educational Progress (NAEP). (1988). *Mathematics Objectives: 1990 Assessment.* Princeton, NJ: NAEP. *Mathematics Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, U.S. Department of Education, 1995).

numerical algorithms using paper and pencil, exhibit knowledge of basic measurement formulas as they are applied in geometric settings, and complete questions reflecting the direct application of mathematics to daily-living skills (such as those related to time and money). As such, during a time of change and reform in the mathematics curriculum, the NAEP long-term trend assessment's results provide an index of whether students are losing ground with respect to long-held goals, while classrooms may be placing more emphasis on processes such as problem solving and communicating mathematics. The long-term trend for the three age groups indicates that, although curriculum goals have been altered to focus more heavily on problem solving, conceptual development, reasoning, and communication skills, there has been no downward movement in student performance on questions designed to measure more traditional procedural aspects in the mathematics curriculum.[19]

The computational focus of the long-term trend assessment also provides an anchor for how well our students are measuring up to traditional procedural skills as the calculator plays an increasingly greater role in the mathematics curriculum from kindergarten through the undergraduate level. Calculators are allowed for a few questions on the long-term trend assessment, but most questions are multiple-choice and are completed without the use of a calculator.

The results from the seven NAEP long-term trend assessments in mathematics provide a wide range of information about how students' performance has changed during the 21-year period from 1973 to 1994. The scale used to measure student performance, which ranges from 0 to 500, provides a common metric for comparing average performance across trend assessments, age groups, and demographic subpopulations. NAEP has also developed descriptions for student performance at five levels on the scale: Level 150 – Simple Arithmetic Facts; Level 200 – Beginning Skills and Understandings; Level 250 – Basic Operations and Beginning Problem Solving; Level 300 – Moderately Complex Procedures and Reasoning; and Level 350 – Multistep Problem Solving and Algebra.

NAEP reports the performance of groups and subgroups of students, not individuals. The measures of achievement included in this report are the average performance of groups of students on the NAEP mathematics scale and the percentages of students attaining successive levels of performance on

[19] *Curriculum and Evaluation Standards for School Mathematics* (Reston, VA: National Council for Teachers of Mathematics, 1989).

the scale. Because the average scale scores and the percentages presented in this report are based on samples, they are necessarily estimates. Like all estimates based on surveys, they are subject to sampling as well as measurement error. To compute standard errors, NAEP uses a complex procedure that estimates the sampling error and other random error associated with observed assessment results.

In the tables and figures that present mathematics trend results, statistically significant differences between 1994 and prior assessments are denoted with an asterisk; statistically significant differences between 1973 and subsequent assessments are denoted with a dagger. In addition to comparisons between individual assessment years, a second test of significance was conducted to detect statistically significant linear and quadratic trends across the assessments. (See the Procedural Appendix for a discussion of the procedure.) This type of analysis allows for the discussion of statistically significant patterns that may be missed by year-to-year comparisons. For example, from year-to-year, students' average scale scores may consistently increase by a small amount. Although these small increases between years may not be statistically significant, the overall increasing trend in average scores may be statistically significant and noteworthy. All of the differences and trend patterns discussed in this report are statistically significant at the .05 level.

4

National Trends in Mathematics Scale Scores from 1973 to 1994

The trend lines in Figure 4.1 show changes in average mathematics scores for 9-, 13-, and 17-year-olds across the 21-year period from 1973 to 1994. The patterns of change, except for the dotted lines from 1973 to 1978, are based on recent scalings developed to provide valid pictures of trends in the data. The results for the 1973 assessment were extrapolated from previous NAEP analyses. (Please refer to the Procedural Appendix for details about the scaling methodology and for information about drawing inferences from the trend analyses.)

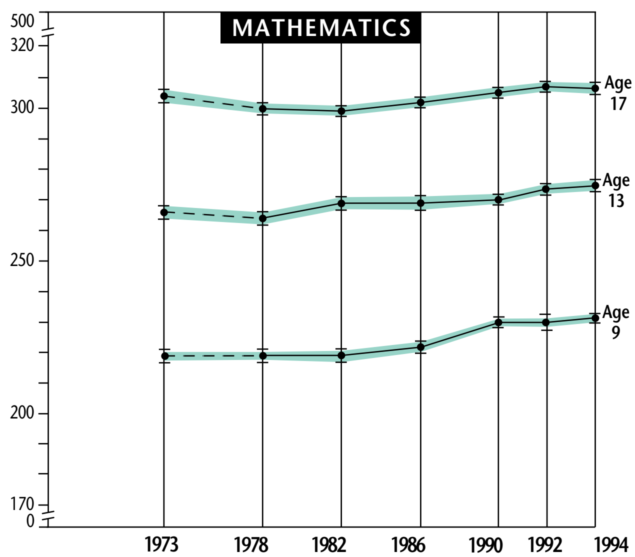
Nine-Year-Olds. The average mathematics scale score of 9-year-olds did not change significantly during the 1970s and early 1980s. Beginning with the 1986 assessment, a pattern of increased performance was observed. Although no change has been observed during the last three assessments, in 1994 the average score of 9-year-olds was higher than in 1973.

Thirteen-Year-Olds. After a small decline in performance during the 1970s, the average mathematics score of 13-year-olds increased. The 1994 average score of 13-year-olds was higher than the average score of their counterparts in 1973.

Seventeen-Year-Olds. The average mathematics score of 17-year-olds declined from 1973 to 1982. Although performance has increased since that time, the average score in 1994 did not differ significantly from the average score in 1973.

Figure 4.1

Trends in Average Mathematics Scale Scores for the Nation, 1973 to 1994



	1973	1978	1982	1986	1990	1992	1994
Age 17	304(1.1)	300(1.0)*	299(0.9)*†	302(0.9)*	305(0.9)	307(0.9)	306(1.0)
Age 13	266(1.1)*	264(1.1)*	269(1.1)*	269(1.2)*	270(0.9)*†	273(0.9)†	274(1.0)†
Age 9	219(0.8)*	219(0.8)*	219(1.1)*	222(1.0)*	230(0.8)†	230(0.8)†	231(0.8)†

95-percent confidence interval [– – –] Extrapolated from previous NAEP analyses.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1973, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1978 to 1994 by Quartiles

Table 4.1 presents trends in mathematics scale scores for 9-, 13-, and 17-year-old students who were in the upper quartile (upper 25 percent), middle two quartiles (middle 50 percent), and the lower quartile (lower 25 percent) of student performance in each assessment. Please note that these trends are not extrapolated back to 1973. As would be expected, standard errors are generally smaller for these more homogeneous groups than for the total group.

The results illustrate a consistent improvement in students' performance for all three age groups at each of the quartiles. These data suggest that improvement on the trend mathematics assessments has not been limited to a particular segment of the performance distribution. This information is particularly relevant in light of Goal 3 of *The National Education Goals Report*, which states that "the academic performance of elementary and secondary students will increase significantly in every quartile. . . ."[20]

For 9-year-olds in each performance range, average scores were relatively stable during the 1970s and early 1980s and then increased. In 1994, the average scores of 9-year-olds across the performance distribution were higher than in 1978.

Thirteen-year-olds in the upper quartile displayed relatively stable performance across most of the trend assessments until 1992. In both 1992 and 1994, these students attained average scores that were higher than in 1978. Among 13-year-olds in the middle two quartiles, increased performance was observed in 1982 and has continued to increase through the most recent assessment. In 1994, the average score for these students continued to be higher than in 1978. In the lower quartile of performance, 13-year-olds' average scores increased in 1982, remained relatively stable through the 1980s and 1990s, and in 1994 continued to be higher than in 1978.

After a decline between 1978 and 1982 in the performance of 17-year-olds in the upper and middle two quartiles, gains were made, and their 1994 average scores were higher than the 1978 average scores of their counterparts. In the lower quartile, 17-year-olds also demonstrated gains in performance from 1978 to 1994.

[20] *The National Education Goals Report*, National Education Goals Panel (Washington, DC: U.S. Government Printing Office, 1992).

Table 4.1

Trends in Average Mathematics Scale Scores by Quartiles, 1978 to 1994

		AVERAGE SCALE SCORE		
Quartile	**Year**	**Age 9**	**Age 13**	**Age 17**
Upper Quartile	1994	267(0.8)†	312(0.9)†	342(1.1)†
	1992	266(0.8)†	309(0.6)*†	342(0.7)†
	1990	266(0.8)†	307(0.6)*	341(0.8)†
	1986	259(0.7)*†	306(0.7)*	340(0.7)
	1982	256(0.6)*	306(0.7)*	336(0.6)*†
	1978	256(0.8)*	305(0.6)*	339(0.4)*
Middle Two Quartiles	1994	233(0.6)†	275(0.5)†	307(0.4)†
	1992	232(0.5)†	274(0.4)†	308(0.4)†
	1990	231(0.4)†	271(0.4)*†	305(0.5)*†
	1986	223(0.5)*†	269(0.5)*†	301(0.5)*
	1982	221(0.5)*	269(0.3)*†	299(0.3)*†
	1978	221(0.5)*	266(0.4)*	302(0.3)*
Lower Quartile	1994	191(0.8)†	236(1.0)†	270(0.8)†
	1992	190(0.8)†	236(1.2)†	270(0.9)†
	1990	190(1.0)†	234(0.8)†	268(0.9)†
	1986	181(0.7)*†	233(0.7)†	265(0.9)*†
	1982	179(0.8)*	230(0.8)*†	260(0.7)*
	1978	178(0.6)*	221(0.7)*	260(0.5)*

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1978, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1973 to 1994 by Race/Ethnicity and by Gender

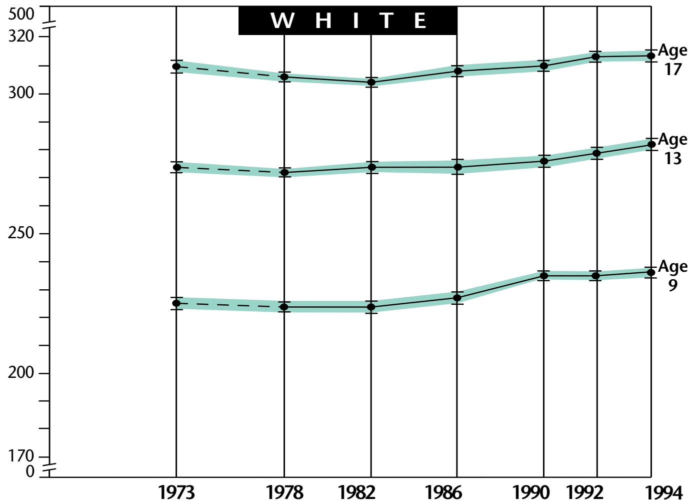
Race/Ethnicity. Figure 4.2 shows trends in average mathematics scale scores for White, Black, and Hispanic students from 1973 to 1994. The performance of White 9- and 13-year olds remained relatively stable throughout the 1970s and 1980s. Beginning with the 1990 assessment for 9-year-olds and the 1992 assessment for 13-year-olds, gains over 1973 averages were observed. In 1994, both 9- and 13-year-old White students had average scores that were higher than that of their counterparts in 1973. The average scores of White 17-year-olds declined from 1973 to 1982. Since that time, gains have been made. However, no significant difference between the 1973 and 1994 average scores was observed for these students.

For Black students at all three ages, significant gains in performance have been observed from 1973 to 1994. Although performance has been relatively stable during the 1990s, the average score of Black students in each age group continued to be higher in 1994 than in 1973.

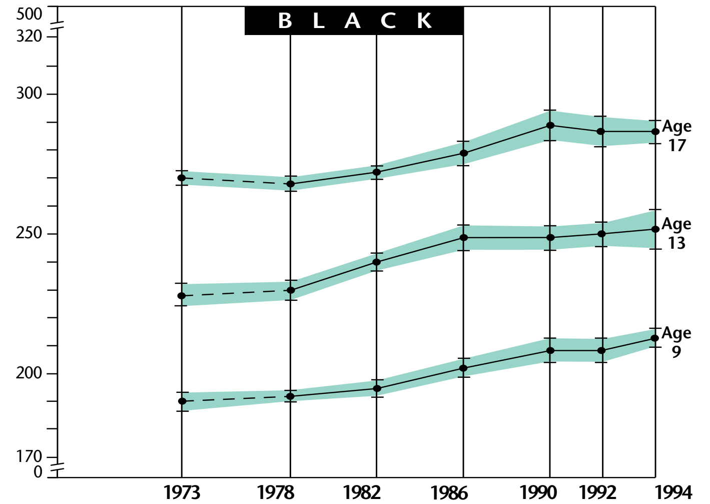
The average score of Hispanic 9-year-olds was higher in 1990 and 1992 than in the first assessment. Although the overall pattern is one of improved performance, there was no significant difference in the average score of these students between 1973 and 1994. Among 13-year-old Hispanic students, gains made during the 1980s have been maintained, and in 1994 the average score was higher than in 1973. The performance of Hispanic 17-year-olds was relatively stable through the 1970s and early 1980s. Since that time, the average score has increased and was higher in 1994 than in 1973.

Figure 4.2

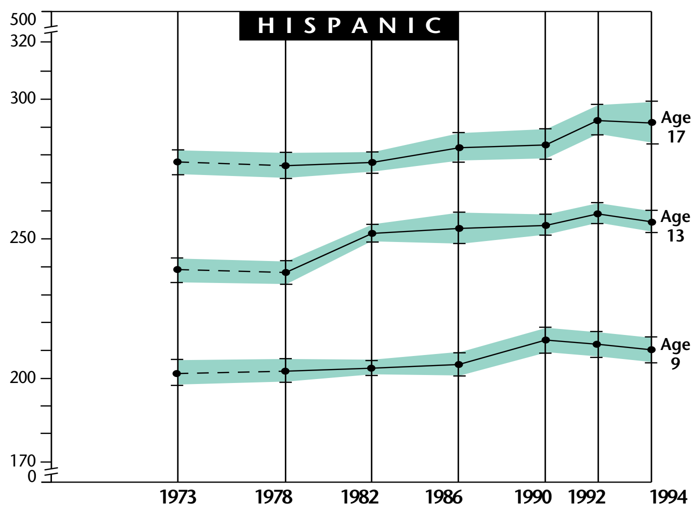
Trends in Average Mathematics Scale Scores by Race/Ethnicity, 1973 to 1994



	1973	1978	1982	1986	1990	1992	1994
Age 17	310(1.1)	306(0.9)*†	304(0.9)*†	308(1.0)*	310(1.0)	312(0.8)	312(1.1)
	—	83(1.3)*	81(2.0)*	78(0.5)*†	73(0.5)†	75(0.5)*†	72(0.5)†
Age 13	274(0.9)*	272(0.8)*	274(1.0)*	274(1.3)*	276(1.1)*	279(0.9)†	281(0.9)†
	—	80(1.7)*	79(2.1)*	77(1.0)*	73(0.7)†	74(0.5)†	73(0.5)†
Age 9	225(1.0)*	224(0.9)*	224(1.1)*	227(1.1)*	235(0.8)†	235(0.8)†	237(1.0)†
	—	79(1.4)*	79(2.5)	77(1.1)	75(1.1)†	75(0.9)	75(0.7)†



	1973	1978	1982	1986	1990	1992	1994
Age 17	270(1.3)*	268(1.3)*	272(1.2)*	279(2.1)†	289(2.8)†	286(2.2)†	286(1.8)†
	—	12(1.1)*	13(1.7)	14(0.3)	16(0.3)†	15(0.3)	16(0.3)†
Age 13	228(1.9)*	230(1.9)*	240(1.6)*†	249(2.3)†	249(2.3)†	250(1.9)†	252(3.5)†
	—	13(1.5)	14(1.8)	14(0.9)	16(0.3)	16(0.3)	15(0.3)
Age 9	190(1.8)*	192(1.1)*	195(1.6)*	202(1.6)*†	208(2.2)†	208(2.0)†	212(1.6)†
	—	14(1.4)	14(2.0)	15(0.5)	16(0.7)	16(0.4)	15(0.5)



	1973	1978	1982	1986	1990	1992	1994
Age 17	277(2.2)*	276(2.3)*	277(1.8)*	283(2.9)	284(2.9)	292(2.6)†	291(3.7)†
	—	4(0.5)*	5(1.0)*	6(0.3)*	7(0.4)*†	7(0.5)†	9(0.3)†
Age 13	239(2.2)*	238(2.0)*	252(1.7)†	254(2.9)†	255(1.8)†	259(1.8)†	256(1.9)†
	—	6(0.9)	5(1.2)	7(1.1)	7(0.5)	7(0.5)	8(0.4)
Age 9	202(2.4)	203(2.2)	204(1.3)	205(2.1)	214(2.1)†	212(2.3)†	210(2.3)
	—	5(0.7)	5(1.1)	6(1.1)	6(0.6)	6(0.8)	6(0.4)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval [– – –] Extrapolated from previous NAEP analyses.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1973 (for scale scores) or 1978 (for percentages), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

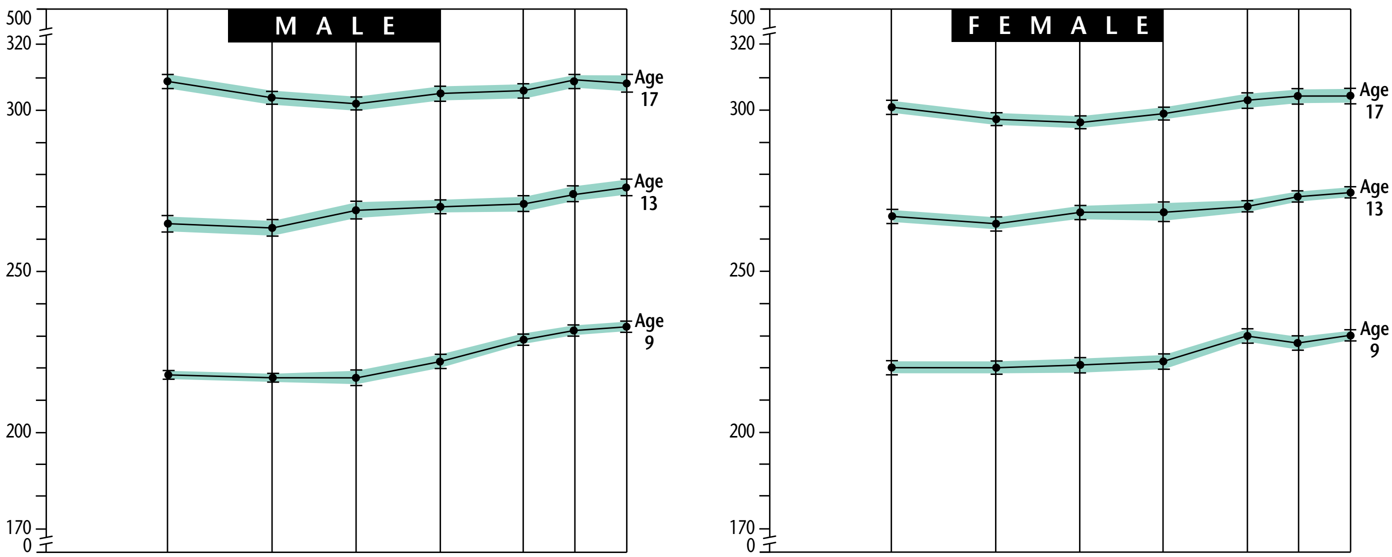
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Gender. Figure 4.3 presents trends in average mathematics scale scores by gender. Among male students, the average score of 9-year-olds was relatively consistent through the 1970s and early 1980s. Beginning in 1986, an increase was observed that has been maintained through the 1994 assessment. Among 13-year-old male students, overall improvement has been observed, and 1994 performance was higher than that observed in 1973. The average score of 17-year-old males declined from 1973 to 1982. Although increased performance has been observed since that time, the average score in 1994 of these students did not differ from the average in 1973.

Among female students, the average scores of 9- and 13-year-olds was relatively consistent through the 1970s and 1980s. An increase over 1973 averages was observed in 1990 for 9-year-olds and in 1992 for 13-year-olds. These increases have been maintained, and the 1994 average scores for both age groups continued to be higher than the 1973 averages. At age 17, female students demonstrated a decline between 1973 and 1982 that has since reversed, resulting in the 1994 average score not being significantly different from the average score of their counterparts in 1973.

Figure 4.3

Trends in Average Mathematics Scale Scores by Gender, 1973 to 1994



MALE

	1973	1978	1982	1986	1990	1992	1994
Age 17	309(1.2)	304(1.0)*†	302(1.0)*†	305(1.2)	306(1.1)	309(1.1)	309(1.4)
	—	49(0.5)	49(0.6)	49(1.2)	49(0.9)	51(1.2)	50(1.3)
Age 13	265(1.3)*	264(1.3)*	269(1.4)*	270(1.1)*†	271(1.2)*†	274(1.1)†	276(1.3)†
	—	50(0.5)	50(0.4)	50(0.9)	50(0.6)	50(0.8)	49(0.8)
Age 9	218(0.7)*	217(0.7)*	217(1.2)*	222(1.1)*†	229(0.9)†	231(1.0)†	232(1.0)†
	—	50(0.5)	49(0.6)	50(0.6)	49(0.6)	49(0.6)	49(0.8)

FEMALE

	1973	1978	1982	1986	1990	1992	1994
Age 17	301(1.1)	297(1.0)*	296(1.0)*†	299(1.0)*	303(1.1)	305(1.1)	304(1.1)
	—	51(0.5)	51(0.6)	51(1.2)	51(0.9)	49(1.2)	50(1.3)
Age 13	267(1.1)*	265(1.1)*	268(1.1)*	268(1.5)	270(0.9)	272(1.0)†	273(1.0)†
	—	50(0.5)	50(0.4)	50(0.9)	50(0.6)	50(0.8)	51(0.8)
Age 9	220(1.1)*	220(1.0)*	221(1.2)*	222(1.2)*	230(1.1)†	228(1.0)†	230(0.9)†
	—	50(0.5)	51(0.6)	50(0.6)	51(0.6)	51(0.6)	51(0.8)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval [– – –] Extrapolated from previous NAEP analyses.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1973 (for scale scores) or 1978 (for percentages), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Differences in Average Scale Scores

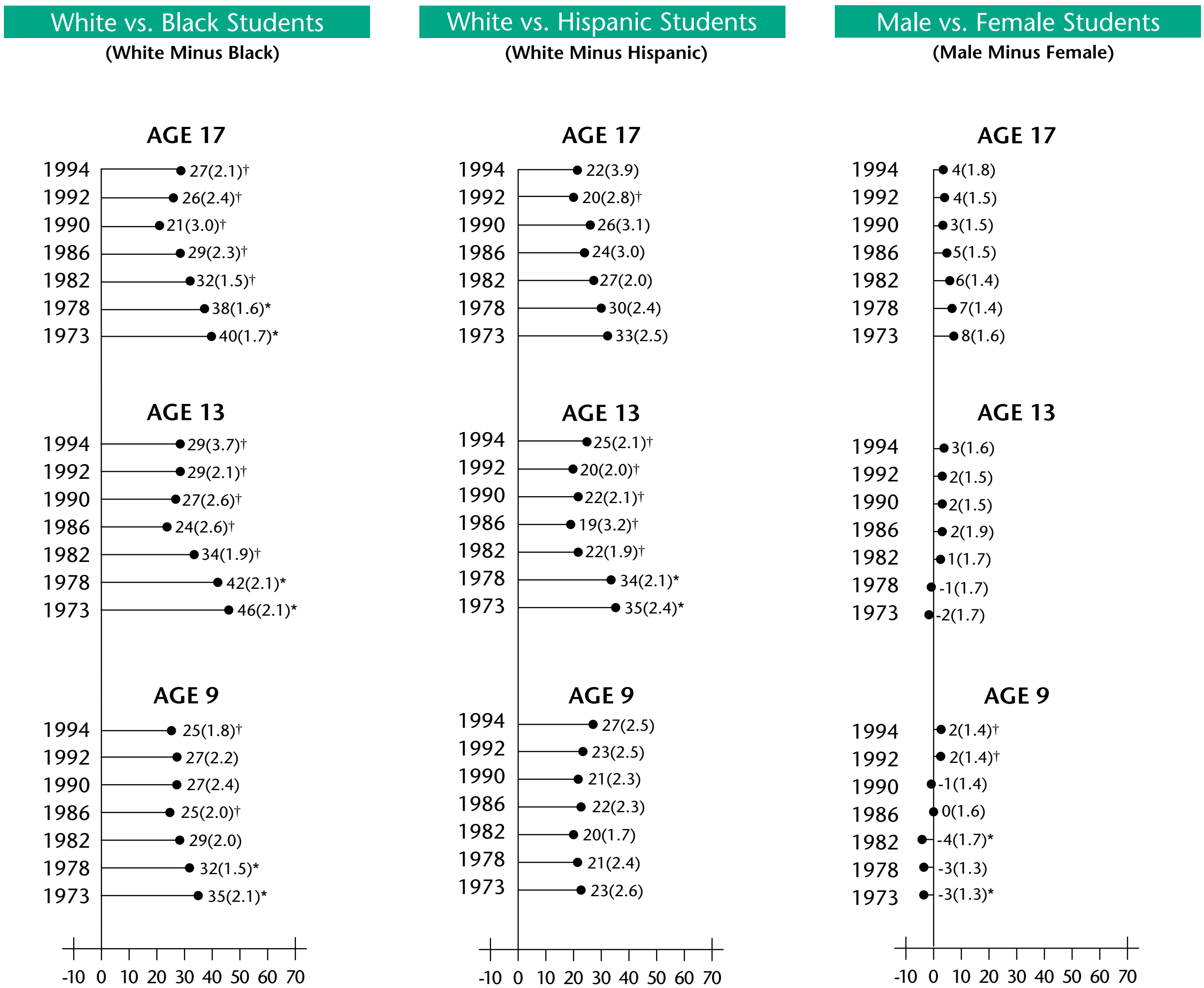
The previous section discussed the trends in mathematics achievement for White, Black, and Hispanic students, and for male and female students. As with past NAEP assessments, significant differences were observed between racial/ethnic subgroups and between males and females. Figure 4.4 presents trends in the differences between the average mathematics scores of selected subgroups of students across the trend assessments.

In 1994, the average scores of White students in each age group were higher than the average scores of their Black peers. At age 9, the gap between White and Black students' scores decreased across the assessment years, and in 1994 it was smaller than the gap observed in 1973. The reduction in the gap for 9-year-olds was due to average score increases among Black students during the first four mathematics trend assessments, while the average scores of White students remained relatively stable during the same time period. At age 13, the gap between scores of White and Black students decreased during the 1970s and 1980s. Although there is evidence that this trend has reversed, the magnitude of the difference between White and Black average performance was smaller in 1994 than in 1973. This trend in score gaps may be attributed to gains made by Black 13-year-olds from 1973 to 1986, while White 13-year-olds showed no significant improvements. Since 1986, White 13-year-olds have made gains, while the average scores of their Black peers have not increased significantly. At age 17, the gap between White and Black students decreased across time and was smaller in 1994 than in 1973. The trend toward smaller gaps was the result of a 17-point gain in average scale scores made by Black students from 1973 to 1994, compared to a 2-point gain made by White students (gains based on unrounded scale scores).

White students outperformed Hispanic students in mathematics at all ages in 1994. Among 9-year-olds the gap between White and Hispanic students has remained relatively constant since 1973. At age 13, the gap between average scores for White and Hispanic students decreased from 1973 to 1986. Although this trend has reversed in recent assessments, the 1994 gap between White and Hispanic 13-year-olds was significantly smaller than the 1973 gap. From 1973 to 1986, Hispanic students made gains in average scores, while White students remained relatively stable. Since 1986, the average score of White 13-year-olds has increased, but the average for their Hispanic peers has not changed significantly. Although a trend toward smaller differences between White and Hispanic 17-year-olds' average scores was observed, the magnitude of the gap in 1994 was not significantly different from that in 1973.

Trend analyses of the mathematics score gaps between male and female students aged 9 and 13 revealed a shift across time. At both ages, the trend was away from females outperforming males toward either no difference or males outperforming females. At age 13, male students outperformed female students in 1994. Among 9-year-olds, there was no significant difference in 1994 between the performance of male and female students. In 1994, 17-year-old males outperformed their female peers in mathematics. Although there is evidence that the gap has decreased across assessment years, in 1994 the gap was not significantly different from that in 1973.

Figure 4.4
Trends in Differences in Average Mathematics Scale Scores



*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores for White vs. Black student and Male vs. Female student differences) or from 1977 (for White vs. Hispanic student differences), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale score differences appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1973 to 1994 by Region

Figure 4.5 shows trends in average mathematics scale scores for each of four geographic regions of the country — Northeast, Southeast, Central, and West.

In the Northeast, the performance of 9-year-olds remained relatively stable from 1973 until 1990 when an increase over the first assessment was observed. This increase has been maintained and the 1994 average score was higher than that in 1973. The performance of 13-year-olds has been consistent across the trend assessments until 1994, when an increase over the 1973 average was observed. At age 17, average scores demonstrated a decline between 1973 and 1982. Performance remained relatively stable from that point until 1992 when a slight increase was observed. However, the 1994 average score for these students did not differ significantly from the 1973 average.

After a period of relative stability during the 1970s and early 1980s, in 1986 the average score of 9-year-olds in the Southeast increased to a level higher than in 1973. Since 1986, further improvement in performance resulted in a 1994 average score that was significantly higher than the average attained in all previous years, except 1990. For 13-year-olds, average scores increased between 1973 and 1994. The performance of 17-year-olds showed some decline during the 1970s and early 1980s. Although this pattern has reversed, the average score in 1994 did not differ significantly from that in 1973.

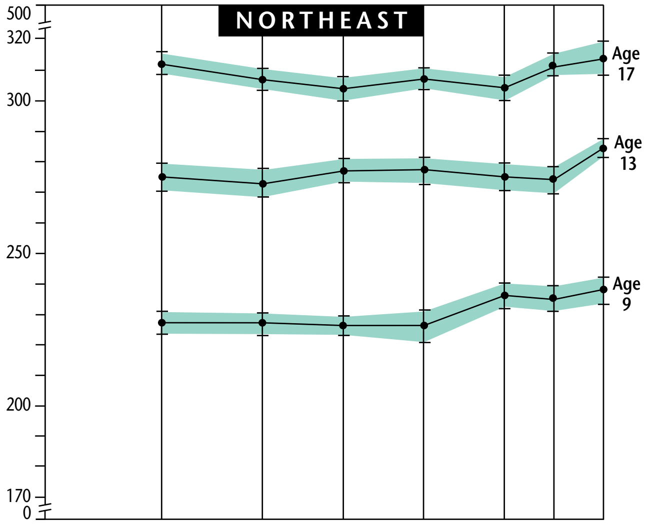
In the Central region, 9- and 17-year-old students demonstrated some decline in performance during the 1970s and early 1980s. Although the average scores of both age groups have since increased, only 9-year-olds attained an average score in 1994 that was higher than the average score in 1973. At age 13, no significant changes have been observed across the assessment years for students in the Central region.

Overall improvement was observed for 9- and 13-year-olds in the West region. In addition, the average scores in 1994 for students in both age groups were higher than those in 1973. After a period of decline between 1973 and 1982 for 17-year-olds, a subsequent pattern of improvement has returned performance in 1994 to a level not significantly different from that in 1973.

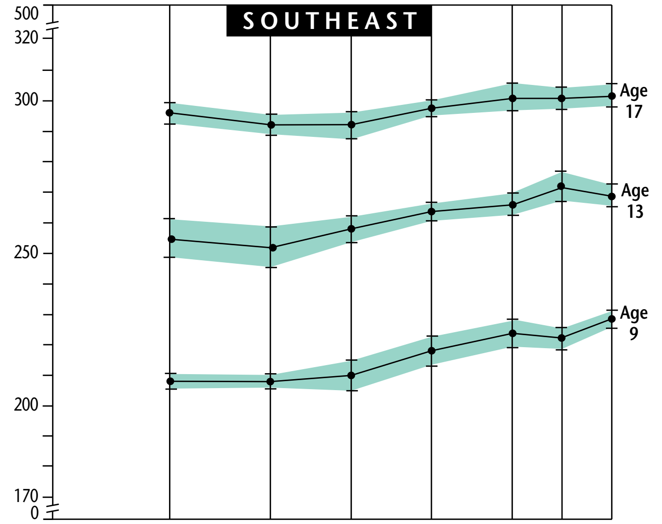
In 1994, comparisons of average scores of students from different regions revealed several differences. At ages 9 and 13, students from the Northeast had higher average scores than students from the Southeast or West regions. Also, 9-year-olds from the Central region outperformed their peers in the West and 17-year-olds in the Northeast outperformed their peers in the Southeast.

Figure 4.5

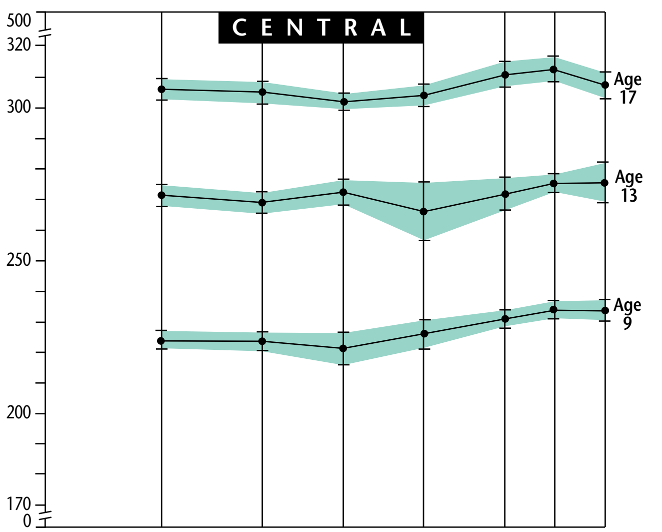
Trends in Average Mathematics Scale Scores by Region, 1973 to 1994



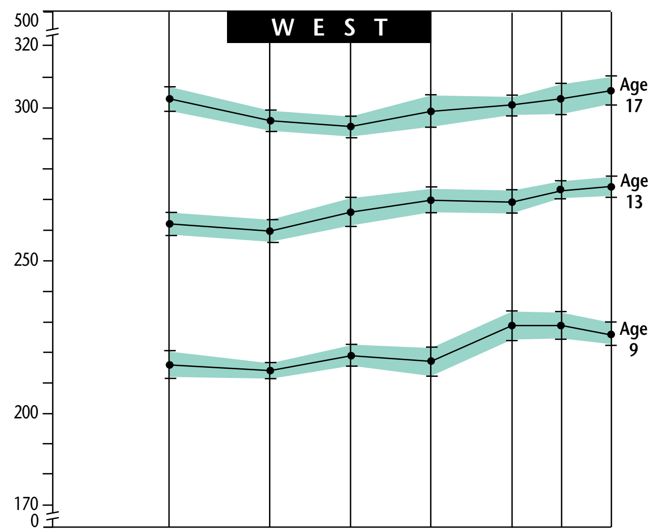
	1973	1978	1982	1986	1990	1992	1994
Age 17	312(1.8)	307(1.8)	304(2.0)†	307(1.9)	304(2.1)†	311(2.0)	313(2.9)
	—	23(2.0)	24(2.8)	24(0.8)	22(1.0)	22(0.9)	23(1.0)
Age 13	275(2.4)*	273(2.4)*	277(2.0)*	277(2.2)*	275(2.3)*	274(2.2)*	284(1.5)†
	—	22(2.2)	24(2.5)	22(1.6)	24(1.0)*	21(0.8)	20(0.9)
Age 9	227(1.9)*	227(1.9)*	226(1.8)*	226(2.7)*	236(2.1)†	235(1.9)†	238(2.2)†
	—	22(2.5)	21(2.7)	21(1.1)	23(0.9)	21(1.2)	22(0.9)



	1973	1978	1982	1986	1990	1992	1994
Age 17	296(1.8)	292(1.7)*	292(2.1)*	297(1.4)	301(2.3)	301(1.9)	301(1.6)
	—	22(2.1)	20(2.1)	23(2.3)	24(0.9)	25(1.4)	25(1.5)
Age 13	255(3.2)*	253(3.3)*	258(2.2)*	264(1.4)	266(1.9)†	271(2.5)†	269(2.0)†
	—	23(2.0)	22(2.2)	25(5.8)	23(0.8)	24(1.6)	22(1.7)
Age 9	208(1.3)*	209(1.2)*	210(2.5)*	218(2.5)*†	224(2.4)†	221(1.7)*†	229(1.4)†
	—	24(2.1)	24(2.7)	22(4.7)	24(0.9)	24(1.3)	24(1.5)



	1973	1978	1982	1986	1990	1992	1994
Age 17	306(1.8)	305(1.9)	302(1.4)	304(1.9)	311(2.1)	312(2.0)	307(2.2)
	—	31(2.1)	30(4.3)	28(2.2)	26(0.9)	25(1.0)†	26(1.1)
Age 13	271(1.8)	269(1.8)	273(2.1)	266(4.5)	272(2.4)	275(1.5)	275(3.4)
	—	31(1.8)	27(3.4)	25(5.0)	24(0.7)†	27(1.0)	27(1.8)
Age 9	224(1.5)*	224(1.5)*	221(2.7)*	226(2.3)	231(1.3)†	234(1.6)†	233(1.8)†
	—	31(1.8)	28(3.8)	29(4.0)	24(0.6)†	28(0.8)	26(1.0)



	1973	1978	1982	1986	1990	1992	1994
Age 17	303(2.0)	296(1.8)*†	294(1.9)*†	299(2.7)	302(1.5)	303(2.3)	305(2.4)
	—	23(2.5)	26(2.8)	26(0.9)	28(0.9)	29(0.9)	27(1.0)
Age 13	262(1.9)*	260(1.9)*	266(2.4)	270(2.1)†	269(1.6)†	272(1.4)†	272(1.7)†
	—	24(2.2)*	27(2.3)	28(1.5)	29(0.9)	29(0.9)	31(1.1)†
Age 9	216(2.2)*	214(1.3)*	219(1.8)*	217(2.4)*	229(1.8)†	229(2.3)†	226(1.6)†
	—	23(2.0)	27(2.0)	28(1.6)	28(0.8)	27(0.8)	28(1.0)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval [– – –] Extrapolated from previous NAEP analyses.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1973 (for scale scores) or 1978 (for percentages), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1978 to 1994 by Parents' Highest Level of Education

Trend results from 1978 to 1994 in average mathematics scores by parents' highest level of education are presented in Table 4.2. (Results by parental education level are not available for extrapolated data.) Across all three age groups, students reported higher levels of parental education in 1994 than in 1978. A higher percentage at each grade reported having at least one parent who had graduated from college, whereas a lower percentage reported that neither parent had graduated from high school, or that high school graduation was their parents' highest level of education.

For 9-year-olds who reported college graduation as their parents' highest level of education, a period of relative stability in performance was followed by an increase in performance. This pattern was also observed for 9-year-olds who reported that at least one parent had received some education after high school. For 9-year-olds who reported the two lower levels of parental education, a pattern of overall improvement was observed across the assessment years. At all levels of parental education for 9-year-olds, the 1994 average scores were higher than those in 1978.

At age 13, students who reported that at least one parent had graduated from college demonstrated a pattern of declining performance from 1978 to 1990, followed by an increase in performance. However, there was no difference between 1978 and 1994 in the average scores for these students. Although a pattern of overall improvement was observed for 13-year-olds with at least one parent who had received some education after high school, there was no significant difference between 1978 and 1994 in average scores. No significant change across the assessment years was observed for 13-year-olds who reported high school graduation as their parents' highest level of education. An increase in average scale scores between 1978 and 1994 was observed for 13-year-olds who reported that neither parent had graduated from high school.

Among 17-year-olds, the average score of students who reported that at least one parent had graduated from college declined in 1982. Since that time the average score has increased, but in 1994 was not significantly different from the 1978 average. The performance of 17-year-olds who reported that at least one parent had received some education after high school has been relatively stable since 1978. Although trend analyses indicated a pattern of overall improvement in average scores for 17-year-olds who reported high school graduation or less as their parents' highest level of education, there was no significant difference between 1978 and 1994 averages.

Table 4.2

Trends in Average Mathematics Scale Score by Parents' Highest Level of Education, 1978 to 1994

		AGE 9		AGE 13		AGE 17	
Level of Education	**Year**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**
Graduated College	1994	45(0.8)†	238(0.8)†	46(1.3)†	285(1.2)	44(1.5)†	318(1.4)
	1992	42(1.2)†	236(1.0)†	44(1.3)†	283(1.0)	43(1.4)†	316(1.0)
	1990	40(1.1)*†	238(1.3)†	41(1.2)*†	280(1.0)*	39(1.4)*†	316(1.3)
	1986	38(1.1)*†	231(1.1)*	38(2.0)*†	280(1.4)*	37(1.2)*†	314(1.4)
	1982	30(1.5)*†	229(1.5)*	32(1.3)*†	282(1.5)	32(1.3)*	312(1.0)*†
	1978	24(1.1)*	231(1.1)*	26(1.2)*	284(1.2)	32(1.1)*	317(1.0)
Some Education After High School	1994	7(0.4)†	239(2.1)†	17(0.6)†	277(1.6)	24(1.1)†	305(1.3)
	1992	8(0.4)	237(1.9)†	18(0.7)†	278(1.0)†	25(0.9)†	308(1.1)
	1990	7(0.4)	236(2.0)	17(0.6)†	277(1.0)†	24(0.9)†	308(1.0)
	1986	7(0.6)†	229(2.1)*	16(0.6)	274(0.8)	24(1.0)†	305(1.2)
	1982	9(0.4)*	225(2.1)*	14(0.4)*	275(0.9)	18(0.5)*	304(0.9)
	1978	9(0.4)*	230(1.7)*	14(0.4)*	273(1.2)	16(0.4)*	305(0.9)
Graduated High School	1994	14(0.6)†	225(1.3)†	23(0.9)†	266(1.1)	22(0.8)†	295(1.1)
	1992	14(0.7)†	222(1.5)	23(0.9)†	263(1.2)	21(0.9)†	298(1.7)
	1990	16(0.7)†	226(1.2)†	27(0.8)*†	263(1.2)	26(1.1)*†	294(0.9)
	1986	16(0.7)†	218(1.6)*	31(1.3)*	263(1.2)	28(1.1)*†	293(1.0)
	1982	25(0.8)*	218(1.1)*	34(0.8)*	263(0.8)	33(0.8)*	293(0.8)
	1978	23(0.8)*	219(1.1)*	33(0.8)*	263(1.0)	33(0.7)*	294(0.8)
Less Than High School	1994	4(0.4)†	210(3.0)†	6(0.4)†	255(2.1)†	7(0.5)†	284(2.4)
	1992	4(0.3)†	217(2.2)†	6(0.5)†	256(1.0)†	8(0.6)†	286(2.3)
	1990	5(0.4)†	210(2.3)†	8(0.5)†	253(1.8)†	8(0.6)†	285(2.2)
	1986	4(0.4)†	201(2.5)	8(1.1)†	252(2.3)†	8(0.4)†	279(2.3)
	1982	8(0.7)*	199(1.7)*	11(0.6)*	251(1.4)†	14(0.9)*	279(1.0)
	1978	8(0.4)*	200(1.5)*	12(0.6)*	245(1.2)*	13(0.6)*	280(1.2)
I Don't Know	1994	30(0.8)†	225(1.1)†	8(0.5)†	252(2.4)†	3(0.3)†	283(3.8)
	1992	33(0.8)	225(1.0)†	8(0.4)†	253(1.8)†	3(0.3)†	290(3.9)†
	1990	32(0.8)†	223(1.0)†	8(0.5)†	248(2.1)†	3(0.4)†	277(2.8)
	1986	35(1.0)*	214(1.4)*	8(0.4)†	247(2.3)†	3(0.3)†	281(2.4)
	1982	27(1.1)†	213(1.5)*	9(0.8)†	252(3.2)†	4(0.3)	272(1.8)*
	1978	37(1.5)*	211(1.1)*	15(0.9)*	240(1.3)*	5(0.4)*	276(1.9)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1978, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1978 to 1994 by Type of School

Comparison of the educational quality of public and nonpublic schools is an issue that has received a great deal of interest in recent years. The NAEP trend assessment samples in such a way that a comparison can be made between public and nonpublic schools. (Results by type of school are not available for extrapolated data). The percentages of students enrolled in the two types of school have remained relatively constant over time, with approximately 88 percent of 9-year-olds, 13-year-olds, and 17-year-olds enrolled in public schools and the remaining 12 percent at each age enrolled in nonpublic schools. The 1994 data displayed in Table 4.3 show that nonpublic schools had higher average scale scores than public schools in all three age groups.

The average scores of public school students at all three ages demonstrated an overall pattern of increased performance across the assessment years. Additionally, the 1994 average score for each age group was higher than the 1978 average score.

Among nonpublic school students, the performance of 9-year-olds was relatively stable during the late 1970s and 1980s. Beginning in 1990, an increase in the average score compared to 1978 was observed and has been maintained through the most recent assessment. Although an overall pattern of improved performance was observed for 13-year-old nonpublic school students, the 1978 and 1994 average scores did not differ significantly. No significant change has been observed across the assessment years in the performance of 17-year-old students attending nonpublic schools.

Table 4.3

Trends in Average Mathematics Scale Score by Type of School, 1978 to 1994

Type of School	Year	AGE 9 Percent of Students	AGE 9 Average Scale Score	AGE 13 Percent of Students	AGE 13 Average Scale Score	AGE 17 Percent of Students	AGE 17 Average Scale Score
Public	1994	88(1.8)	229(0.9)†	88(1.7)	273(1.1)†	88(2.3)†	304(0.9)†
	1992	87(1.6)	228(0.9)†	88(1.9)	272(1.0)†	91(2.2)	305(0.9)†
	1990	89(2.1)	229(0.9)†	90(1.4)	269(1.0)†	93(1.8)	304(0.8)†
	1986	84(2.7)	220(1.2)*	96(1.8)*	269(1.2)*†	96(1.4)*	301(1.0)
	1982	86(2.2)	217(1.1)*	89(1.3)	267(1.3)*†	92(1.6)	297(0.9)*
	1978	89(1.8)	217(0.8)*	91(1.6)	263(1.2)*	94(1.0)*	300(1.0)*
Nonpublic	1994	12(1.8)	245(2.3)†	12(1.7)	285(2.4)	12(2.3)†	319(4.0)
	1992	13(1.6)	242(1.7)†	12(1.9)	283(2.5)	9(2.2)	320(3.0)
	1990	11(2.1)	238(2.3)†	10(1.4)	280(1.7)	7(1.8)	318(6.6)
	1986	16(2.7)	230(2.5)*	4(1.8)*	276(4.9)	4(1.4)*	320(9.8)
	1982	14(2.2)	232(2.1)*	11(1.3)	281(2.1)	8(1.6)	311(1.7)
	1978	11(1.8)	231(1.7)*	9(1.6)	279(1.4)	6(1.0)*	314(3.2)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1978, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Average Mathematics Scale Scores from 1978 to 1994 by Modal Grade

Table 4.4 shows average mathematics scale scores by students enrolled below, at, and above modal grade for each age group. The modal grade is the grade typically attended by students of a particular age. The modal grades are grade 4 for age 9, grade 8 for age 13, and grade 11 for age 17. Students who are below modal grade are older than the expected age of students in their grade. Conversely, students who are above modal grade are younger than the expected age of students in their grade. At all three ages, the percentages of students below modal grade increased across the assessment years. (Results by modal grade are not available for extrapolated data.)

For students in each age group who were below or at modal grade, an overall pattern of improved performance was observed across the assessment years. In addition, the 1994 average scores for these students was higher than the average scores of their counterparts in 1978. Among 17-year-olds who were above their modal grade, an overall pattern of improved performance was observed. However, the 1994 average score for these students did not differ significantly from that of their counterparts in 1978.

Table 4.4

Trends in Average Mathematics Scale Score by Modal Grade, 1978 to 1994

		AGE 9		AGE 13		AGE 17	
Modal Grade	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Below Modal Grade	1994	33(1.3)†	211(1.1)†	38(1.3)†	259(1.2)†	21(1.6)†	284(1.6)†
	1992	38(1.2)†	208(1.2)†	37(1.1)†	258(1.3)†	24(1.1)†	285(1.4)†
	1990	35(1.4)†	207(1.2)†	36(1.3)†	253(1.0)*†	22(1.0)†	282(1.7)†
	1986	34(1.7)†	198(1.0)*†	33(2.1)	251(1.1)*†	17(0.9)	277(1.6)*
	1982	30(1.5)	193(1.4)*	28(1.4)*	247(1.4)*†	16(1.0)*	274(1.6)*
	1978	26(0.9)*	191(1.1)*	27(1.1)*	240(1.4)*	15(0.6)*	273(1.1)*
At Modal Grade	1994	66(1.3)†	241(1.0)†	62(1.3)†	283(1.0)†	73(1.7)	312(0.9)†
	1992	62(1.2)†	243(0.7)†	62(1.0)†	282(0.9)†	70(1.0)†	313(0.8)†
	1990	65(1.4)†	242(1.0)†	63(1.4)†	280(0.9)†	70(1.0)†	311(0.8)†
	1986	66(1.7)†	234(1.0)*†	67(2.1)	278(1.0)*†	75(1.2)	307(0.9)*
	1982	69(1.5)	230(1.0)*	70(1.4)*	277(0.9)*	75(1.0)	303(0.9)*
	1978	72(0.9)*	229(0.9)*	71(1.1)*	274(1.1)*	75(0.7)	305(1.0)*
Above Modal Grade	1994	***(***)	***(***)	***(***)	***(***)	6(0.6)†	316(3.1)
	1992	0(0.1)	***(***)	0(0.1)	***(***)	6(0.5)†	318(2.4)†
	1990	0(0.1)	***(***)	1(0.2)	278(16.5)	8(0.6)	311(1.8)
	1986	0(0.1)	***(***)	1(0.1)	297(7.7)	8(0.7)	309(3.0)
	1982	1(0.1)	258(9.3)	1(0.5)	304(6.3)	10(0.7)*	307(1.4)*
	1978	1(0.2)	241(7.1)	1(0.2)	298(9.1)	10(0.5)*	309(1.0)

NOTE: The modal grades are: grade 4 at age 9, grade 8 at age 13, and grade 11 at age 17.

*Statistically significant difference from 1992, where alpha equals .05 per set of comparisons.

†Statistically significant difference from 1978, where alpha equals .05 per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages of students may not total 100 percent due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessments.

Summary

The overall picture of mathematics achievement provided by the long-term trend results is one of early declines or relative stability followed by a pattern of increased performance. For 9-year-olds, average scale scores began to increase with the 1986 assessment and were higher in 1994 than in 1973. For 13-year-olds, slight declines in the 1970s were followed by a period of increased performance that resulted in a 1994 average score that was higher than the average score in 1973. For 17-year-olds, the increased performance observed after a period of decline from 1973 to 1982 has returned the average scale score in 1994 to a level that did not differ significantly from that in 1973.

Trend results for students in the upper, middle two, and lower quartiles of performance indicate that improvement in mathematics scale scores has not been limited to a particular segment of the performance distribution. Nine-year-olds in each performance range displayed increased average scores since the 1980s and attained average scores in 1994 that were higher than those in 1978. For 13-year-olds in the upper quartile, an increase between 1992 and 1994 resulted in an average score in the most recent assessment that was higher than that in 1973. Early gains in average scores that were observed for 13-year-olds in the middle two and lower quartiles continued across the assessment years and resulted in averages scores that were higher in 1994 than in 1978. Seventeen-year-olds in the upper and lower quartiles demonstrated gains in average scores from 1978 to 1994. In the upper and middle two quartiles, 17-year-olds made gains after an early period of decline and attained an average score in 1994 that was higher than the average score of their counterparts in 1978.

White 9- and 13-year-olds demonstrated relatively stable performance during the 1970s and 1980s. Increases were observed in the 1990s that resulted in 1994 average mathematics scores that were higher than 1978 averages. Although White 17-year-olds have made gains in performance after having declined from 1973 to 1982, their average score in 1994 was not significantly different from that of their counterparts in 1973. For Black students, average scores increased from 1973 to 1994 at all ages. Although gains were observed in recent assessments for Hispanic 9-year-olds, the 1994 average score for this group was not significantly different from that observed in 1973. Hispanic 13- and 17-year-olds demonstrated overall gains and attained average scores in 1994 that were higher than averages attained in 1973.

In 1994, White students outperformed Black and Hispanic students in mathematics at all ages. Across the assessment years, the gap between White and Black students has decreased in each age group and is smaller in 1994 than it was in 1973. Average score differences between White and Hispanic students aged 13 and 17 also decreased across the assessments. However, the gap between White and Hispanic 9-year-olds has not changed significantly since 1973.

Male 9-year-olds demonstrated relatively stable mathematics performance through the 1970s and early 1980s, followed by gains that resulted in a higher average score in 1994 than in 1973. Thirteen-year-old males demonstrated increased performance from 1973 to 1994. For 17-year-old males, recent gains that followed a period of decline have resulted in a 1994 average score that did not differ significantly from the average score in 1973. Female 9- and 13-year-olds made gains in the 1990s and attained average scores in 1994 that were higher than those in 1973. Although the performance of 17-year-old females has improved after declining from 1973 to 1982, the 1994 average score for these students did not differ significantly from that of their counterparts in 1973.

In 1994 in mathematics, male students outperformed their female peers at ages 13 and 17. For 13-year-olds this represented a shift across the assessments away from female students outperforming male students. For 17-year-olds, the difference between males and females has decreased since 1973, but continued to be significant. At age 9, a shift away from female students outperforming male students has also been observed. However, there was no significant difference between males and females at age 9 in 1994.

In the Northeast, the overall gains observed for 9- and 13-year-olds resulted in 1994 average mathematics scores that were higher than those in 1973. After a period of decline, the average score of 17-year-olds returned to a level in 1994 that was not significantly different from that in 1973. The average scores of 9- and 13-year-olds in the Southeast have increased from 1973 to 1994. Although an overall increase was observed for 17-year-olds, no significant difference between 1973 and 1994 average scores was observed. In the Central region, 9- and 17-year-olds showed overall gains across the assessments. However, the 1994 average score was higher than the 1973 average only for 9-year-olds. No significant change across the assessment years was observed for 13-year-olds in the Central region. Increased average scores were observed for 9- and 13-year-olds in the West region. At age 17, a period of declining performance was followed by a pattern of improvement that resulted in a 1994 average score not significantly different from that in 1973.

At all levels of parental education, 9-year-olds demonstrated increases in average mathematics scores from 1978 to 1994. Increases were also observed for 13-year-olds who reported the lowest level of parental education. Although an overall gain was observed for 13-year-olds with at least one parent who had received some education after high school, there was no significant difference between 1978 and 1994 average scores. For 13-year-olds at the highest level of parental education (at least one parent who graduated from college), some increase has been observed in average scores since a period of decline in the 1980s. However, the 1978 and 1994 average scores did not differ significantly. Among 17-year-olds, a pattern of overall improvement was observed at the two lower levels of parental education, although the 1978 and 1994 averages did not differ significantly. After a decline between 1978 and 1982, the average score of 17-year-olds at the highest level of parental education has increased, but did not differ significantly in 1994 from the average score in 1978.

Between 1978 and 1994, significant increases in average mathematics scores were observed for public school students at all ages and for nonpublic school students at age 9. Although an overall increase was observed for 13-year-old nonpublic schools students, their average score in 1994 did not differ significantly from that of their counterparts in 1978. No significant change across the assessments was observed for 17-year-olds attending nonpublic schools.

Increased average mathematics scores were observed from 1978 to 1994 for students in each age group who were below or at modal grade. Although an overall increase was indicated in the results for 17-year-old students who were above modal grade, the 1994 average score for this group did not differ significantly from that observed in 1978.

5

National Trends in Levels of Mathematics Performance from 1978 to 1994

To better understand trends in students' knowledge and skills in mathematics, levels of mathematical performance were created to allow for the observation of the nature of any changes. Five levels were established by "anchoring" the NAEP mathematics scale at five points on the scale: 150, 200, 250, 300, and 350. The anchoring was accomplished by determining which questions students performing at one point on the scale were more likely to answer correctly than students performing at the next lower level. Mathematics educators from schools and universities then carefully studied the sets of questions that make up the assessments to develop descriptions for the five levels. These descriptions outline the concepts, procedures, and processes associated with correct responses to the questions at each level. Figure 5.1 provides these descriptions for the five anchor levels.

Figure 5.1 - Levels of Mathematics Performance

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 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.

The percentages of students at ages 9, 13, and 17 reaching the various performance levels in each of the NAEP long-term trend assessments are shown in Table 5.1. Because these analyses were not possible for data collected for the 1973 mathematics assessment, the results are presented for the 1978 through the 1994 assessments.

Table 5.1

Trends in Percentages of Students At or Above Five Mathematics Performance Levels, 1978 to 1994

		ASSESSMENT YEARS					
Performance Levels	**Age**	**1978**	**1982**	**1986**	**1990**	**1992**	**1994**
Level 350							
Multistep Problem	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Solving and Algebra	13	1(0.2)	1(0.1)	0(0.1)†	0(0.1)†	0(0.2)	1(0.2)
	17	7(0.4)	6(0.4)†	7(0.5)	7(0.6)	7(0.6)	7(0.8)
Level 300							
Moderately Complex	9	1(0.1)	1(0.1)	1(0.2)	1(0.3)	1(0.3)	1(0.4)
Procedures and	13	18(0.7)	17(0.9)	16(1.0)*	17(1.0)	19(1.0)	21(1.4)
Reasoning	17	52(1.1)*	49(1.3)*	52(1.4)*	56(1.4)	59(1.3)†	59(1.4)†
Level 250							
Numerical Operations	9	20(0.7)*	19(1.0)*	21(0.9)*	28(0.9)†	28(0.9)†	30(1.1)†
and Beginning	13	65(1.2)*	71(1.2)*†	73(1.6)†	75(1.0)†	78(1.1)†	78(1.1)†
Problem Solving	17	92(0.5)*	93(0.5)*	96(0.5)†	96(0.5)†	97(0.5)†	97(0.5)†
Level 200							
Beginning Skills and	9	70(0.9)*	71(1.2)*	74(1.2)*	82(1.0)†	81(0.8)†	82(0.7)†
Understandings	13	95(0.5)*	98(0.4)†	99(0.2)†	99(0.2)†	99(0.3)†	99(0.3)†
	17	100(0.1)	100(0.0)	100(0.1)	100(0.1)	100(0.0)	100(0.0)
Level 150							
Simple Arithmetic	9	97(0.3)*	97(0.3)*	98(0.3)*†	99(0.2)†	99(0.2)†	99(0.2)†
Facts	13	100(0.1)	100(0.0)	100(0.0)	100(0.0)	100(0.0)	100(0.0)
	17	100(0.0)	100(0.0)	100(0.0)	100(0.0)	100(0.0)	100(0.0)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1978, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Level 350: No improvement was evident for 17-year-olds at Level 350, characterized by the ability to solve multistep problems. Only 7 percent of the 17-year-old students in 1994 attained this level of performance. These results have remained essentially constant since 1978.

Level 300: Students performing at or above Level 300 demonstrated better numerical reasoning and were able to draw from a wider range of mathematical areas, including algebra and geometry. At age 13, the percentage of students attaining this level of performance or above was relatively stable until the 1990s when some increase was observed. However, there was no significant difference between the 1978 and 1994 percentages. After a slight decline in 1982, the percentage of 17-year-olds at or above Level 300 has increased. In 1994, the percentage of 17-year-old students performing at or above Level 300 was higher than the percentage in 1978.

Level 250: Students performing at or above Level 250 had developed a surface understanding of the four basic operations and were beginning to acquire more developed reasoning skills. At all three ages, higher percentages of students in 1994 compared to 1978 reached at least this level of performance. An increase over the 1978 percentages was first observed in 1990 for 9-year-olds, 1982 for 13-year-olds, and 1986 for 17-year-olds. No significant changes in performance at or above Level 250 have been observed during the last three assessments.

Level 200: Students performing at or above Level 200 demonstrated a greater range and depth of basic mathematical skills than did those who reached only Level 150, but were still developing a grasp of multiplication and division and reasoning ability beyond that required by simple numerical computations. Virtually all 13- and 17-year-olds and more than fourth-fifths of the 9-year-olds performed at or above Level 200 in the 1994 assessment. This represented an increase between 1978 and 1994 at ages 9 and 13.

Level 150: In 1994, as in previous mathematics assessments, virtually all students in each of the three age groups understood simple arithmetic facts as described in Level 150.

Trends in Levels of Mathematics Performance between 1978 and 1994 by Race/Ethnicity

Table 5.2 presents changes in the percentages of White, Black, and Hispanic students reaching various levels of performance on the NAEP mathematics scale.

Table 5.2

Trends in Percentages of Students At or Above Five Mathematics Performance Levels by Race/Ethnicity, 1978 and 1994

		ASSESSMENT YEARS					
		1978			1994		
Performance Levels	Age	White	Black	Hispanic	White	Black	Hispanic
Level 350							
Multistep Problem	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Solving and Algebra	13	1(0.2)	0(0.1)	0(0.1)	1(0.3)	0(0.6)	0(0.1)
	17	9(0.5)	1(0.2)	1(0.6)	9(1.1)	0(0.4)	1(1.0)
Level 300							
Moderately Complex	9	1(0.2)	0(0.1)	0(0.5)	2(0.4)	0(0.1)	0(0.1)
Procedures and	13	21(0.7)*	2(0.5)	4(1.0)	26(1.6)	6(2.4)	6(1.8)
Reasoning	17	58(1.1)*	17(1.6)*	23(2.7)*	67(1.4)	30(3.4)	38(5.5)
Level 250							
Numerical Operations	9	23(0.9)*	4(0.6)*	9(2.5)	35(1.3)	11(1.7)	10(1.8)
and Beginning	13	73(0.9)*	29(2.1)*	36(2.9)*	86(0.9)	51(3.9)	59(2.2)
Problem Solving	17	96(0.3)*	71(1.7)*	78(2.3)*	98(0.4)	91(1.8)	92(3.6)
Level 200							
Beginning Skills and	9	76(1.0)*	42(1.4)*	54(2.8)*	87(0.8)	66(2.6)	64(3.1)
Understandings	13	98(0.3)*	80(1.5)*	86(0.9)*	99(0.2)	96(1.6)	97(1.3)
	17	100(0.0)	99(0.3)*	99(0.4)	100(0.0)	100(0.0)	100(0.0)
Level 150							
Simple Arithmetic	9	98(0.2)*	88(1.0)*	93(1.2)*	100(0.2)	97(1.0)	97(1.2)
Facts	13	100(0.0)	99(0.4)*	100(0.3)	100(0.0)	100(0.1)	100(0.1)
	17	100(0.0)	100(0.0)	100(0.0)	100(0.0)	100(0.0)	100(0.0)

* Statistically significant difference from 1994, at a 5 percent significance level, without application of a multiple comparison procedure. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The percentages of 9-year-old White, Black, and Hispanic students attaining at least Levels 150 and 200 were higher in 1994 than in 1978. Additional increases were observed for White and Black students performing at or above Level 250. In 1994, a greater percentage of White students than Black or Hispanic students performed at or above Levels 200 and 250.

Compared to 1978, the percentages of 13-year-old White, Black, and Hispanic students at or above Levels 200 and 250 were higher in 1994. For White students, an additional increase was observed at or above Level 300. Higher percentages of White students than Black or Hispanic students performed at or above Levels 250 and 300 in 1994.

For 17-year-old students in all three racial/ethnic groups, higher percentages performed at or above Levels 250 and 300 in 1994 compared to 1978. Comparing the 1994 performance of students in different racial/ethnic groups revealed higher percentages for White students than Black or Hispanic students at Levels 300 and 350, and higher percentages for White students than Black students at Level 250. However, the difference in the percentage of White and Hispanic 17-year-olds attaining at least Level 250 was not statistically significant.

Trends in Levels of Mathematical Performance between 1978 and 1994 by Gender

Table 5.3 shows the percentages of males and females attaining each of the five performance levels in both 1978 and 1994. At all three ages, the performance of males and females has been remarkably similar across assessments except at the two highest levels of performance. Between 1978 and 1994, both male and female 9-year-olds showed increased performance at Levels 200 and 250. There were improvements for 13-year-olds of both genders at Levels 200 and 250, and for males at Level 300. At age 17, both male and female students had improved performance at Levels 250 and 300.

In 1994, the only significant difference in the percentage of male and female students at any of the performance levels was observed in the top two levels. At age 13, a higher percentage of males than females was at or above Level 300. Among 17-year-olds, the percentage of male students reaching Level 350 was higher than the percentage of female students.

Table 5.3

Trends in Percentages of Students At or Above Five Mathematics Performance Levels by Gender, 1978 and 1994

		ASSESSMENT YEARS			
		1978		1994	
Performance Levels	Age	Male	Female	Male	Female
Level 350					
Multistep Problem	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Solving and Algebra	13	1(0.2)	1(0.2)	1(0.3)	1(0.3)
	17	10(0.6)	5(0.7)	9(1.0)	6(0.9)
Level 300					
Moderately Complex	9	1(0.2)	1(0.2)	1(0.4)	1(0.4)
Procedures and	13	18(0.9)*	18(0.7)	24(1.6)	19(1.4)
Reasoning	17	55(1.2)*	48(1.3)*	60(2.1)	57(1.4)
Level 250					
Numerical Operations	9	19(0.6)*	20(1.0)*	32(1.6)	28(1.3)
and Beginning	13	64(1.3)*	66(1.2)*	79(1.5)	77(1.0)
Problem Solving	17	93(0.5)*	91(0.6)*	97(0.6)	96(0.6)
Level 200					
Beginning Skills and	9	69(1.0)*	72(1.1)*	82(0.9)	82(0.9)
Understandings	13	94(0.5)*	95(0.5)*	98(0.4)	99(0.3)
	17	100(0.1)	100(0.1)	100(0.0)	100(0.0)
Level 150					
Simple Arithmetic	9	96(0.5)*	97(0.3)*	99(0.3)	99(0.3)
Facts	13	100(0.1)	100(0.1)	100(0.1)	100(0.0)
	17	100(0.0)	100(0.0)	100(0.0)	100(0.0)

*Statistically significant difference from 1994, at a 5 percent significance level, without application of a multiple comparison procedure. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). When percentages of students are exactly either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages of 0.5 or less were rounded to 0 percent.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The percentages of 9-year-old students at or above Levels 200 and 250 in mathematics showed a pattern of overall increases. Also, the 1994 percentages of 9-year-olds at these performance levels were higher than the 1978 percentages. More White, Black, and Hispanic 9-year-olds reached at least Levels 150 and 200 in 1994 than in 1978; an increase was also seen for White and Black students at or above Level 250. In 1994, White 9-year-olds outperformed their Black and Hispanic peers at Levels 200 and 250. The percentage of male and female 9-year-old students at or above Levels 200 and 250 was higher in 1994 than in 1978. In 1994, there was no significant difference in the percentage of males and females at any of the performance levels.

There was an overall pattern of increase in the percentages of 13-year-old students at or above Levels 200, 250, and 300 in mathematics across the assessment years. At Levels 200 and 250, the 1994 percentages were higher than the 1978 percentages. There were increases between 1978 and 1994 in the percentages of White, Black, and Hispanic 13-year-olds performing at or above Levels 200 and 250, and in the percentage of White 13-year-olds at or above Level 300. In 1994, White 13-year-old students outperformed their Black and Hispanic peers at Levels 250 and 300. In 1994, more male than female students performed at or above Level 300 among 13-year-olds.

The percentages of 17-year-old students at or above Levels 250 and 300 in mathematics showed a pattern of increase from 1978 to 1994. In addition, the 1994 percentages of 17-year-olds at or above these two performance levels were higher than the 1978 percentages. Compared to 1978, the percentages of White, Black, and Hispanic 17-year-olds at or above Levels 250 and 300 were higher in 1994. For the 1994 assessment, the percentages of White 17-year-olds at or above Levels 250, 300, and 350 were higher than the percentages of Black or Hispanic students at these performance levels. Both male and female students at this age demonstrated gains between 1978 and 1994 at Levels 250 and 300. Among 17-year-olds in 1994, more male than female students performed at or above Level 350.

6

Patterns and Trends in School and Home Contexts for Learning Mathematics

Introduction

With both professional mathematics groups and federal legislation setting national goals for school mathematics, the focus on school mathematics in the United States has probably never been greater. Recommendations for reform include curriculum revision, more active learning and problem solving by students, encouragement of all students to reach their full potential in course selection and completion, and increased use of technology (calculators and computers) in the learning of mathematics.[21]

[21] *Curriculum and Evaluation Standards for School Mathematics* (Reston, VA: National Council of Teachers of Mathematics, 1989).

Professional Standards for Teaching School Mathematics (Reston, VA: National Council of Teachers of Mathematics, 1991).

Central to these new goals for school mathematics is the increased focus on student mastery of processes: problem solving, reasoning, communication, and connecting mathematical ideas across contexts. Calls for such a focus have come both from the mathematics community and from those who seek to employ the graduates of the nations' schools.[22]

Trends in Classroom Instruction at Age 17 between 1978 and 1994

The NAEP trend assessment results provide one window through which to view the change in curricular and instructional programs as it affects the nation's 17-year-olds. Further, in many cases, data exist to examine the various factors associated with change during the period between 1978 and 1994. Such factors include several in which students might be more actively engaged in mathematics learning, such as participating in discussions, completing reports, or carrying out projects. The activities sampled also cover more passive activities, such as listening to the teacher and watching the teacher work problems on the board. Table 6.1 presents the data which address these factors in mathematics training.

[22] *Learning a Living: A Blueprint for High Performance,* (Washington, DC: Secretary's Commission on Achieving Necessary Skills, Department of Labor, 1992).

A Call for Change: Recommendations for the Mathematical Preparation of Teachers of Mathematics (Washington, DC: Committee on the Mathematical Education of Teachers, Mathematical Association of America, 1991).

Table 6.1

Trends in Mathematics Classroom Activities at Age 17, 1978 and 1994

		OFTEN		SOMETIMES		NEVER	
In your high school mathematics courses, how often did you. . .	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Listen to a teacher explain a mathematics lesson?	1994	85(1.0)	307(1.3)	12(0.8)	300(2.8)	4(0.5)	288(2.9)
	1978	79(1.2)*	304(1.5)	19(1.1)*	294(3.2)	2(0.4)*	282(6.0)
Discuss mathematics in class?	1994	63(1.4)	309(1.5)	30(1.1)	303(1.4)	8(0.6)	295(2.8)
	1978	51(1.5)*	306(1.8)	43(1.4)*	298(1.8)*	7(0.6)	289(4.0)
Watch the teacher work mathematics problems on the board?	1994	85(1.2)	308(1.2)	12(1.0)	299(3.2)	3(0.5)	282(3.9)
	1978	80(1.1)*	304(1.5)*	18(0.9)*	292(2.9)	2(0.4)	282(5.2)
Work mathematics problems on the board?	1994	27(1.3)	304(2.1)	49(1.3)	309(1.4)	24(1.5)	302(1.7)
	1978	28(1.3)	303(1.9)	60(1.2)*	302(1.8)*	12(1.1)*	293(3.9)*
Make reports or do projects on mathematics?	1994	4(0.5)	298(2.9)	29(1.8)	306(2.4)	67(2.1)	307(1.1)
	1978	2(0.2)*	286(8.3)	23(1.2)*	300(2.5)	75(1.3)*	302(1.5)*
Take mathematics tests?	1994	83(0.9)	308(1.2)	16(0.8)	298(2.7)	***(***)	***(***)
	1978	64(1.3)*	308(1.7)	33(1.1)*	292(2.1)	3(0.5)	270(4.7)*

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

In 1994, 85 percent of the nation's 17-year-olds reported often listening to a teacher explain a mathematics lesson and often watching the teacher work mathematics problems on the board. This represents an increase in these more passive classroom behaviors since 1978. An examination of their other possible responses indicates a corresponding decrease in the percentages of students who responded "Sometimes" in 1994. Such shifts in student classroom activity are not consistent with the recommendations of the *NCTM Teaching Standards.*

Responses to the question "Work mathematics problems on the board" showed a decrease in the "Sometimes" category and an increase in the "Never" response. The 17-year-olds' responses about often discussing mathematics in class showed an increase, paralleling a decrease in the "Sometimes" responses. The increase in discussion in the classroom indicates a movement toward meeting the present recommendations for teaching mathematics. The increases in both the "Often" and "Sometimes" response categories for making reports or doing projects on mathematics also reflect a change in practice that corresponds to the suggested teaching approaches for getting students involved in creating, or doing, mathematics.

Another question given 17-year-olds about classroom activities dealt with the frequency of classroom mathematics tests. An increase in the "Often" response category for this question was observed. Whether this trend is toward or against the recommendations for change depends on the type of tests associated with the increased frequency. If the tests were different forms of assessment, providing teachers with information to improve instruction or learning, then the movement would be a positive one. If the tests focus on short-term goals and on procedures, however, the movement would be contrary to current recommendations.

Trends in Mathematics Course Taking at Ages 13 and 17 between 1978 and 1994

Central to moving students to an internationally competitive level in mathematics is making sure that they have had an equal opportunity to learn the same mathematics content as their competitors. Studies across the time span of the NAEP trend assessment have indicated that U.S. students have not had such opportunities as a whole.[23] Others claim that where U.S. schools provide equal opportunity for exposure, neither the focus of instruction nor the expectations match up with those found in the schoolrooms of our economic competitors.[24]

The *NCTM Teaching Standards* emphasize the need for an extension both of the amount of content that students see and the number of courses that students take, as well as for a change in the way learning and teaching occur in these settings. The previous section dealt with a few of the indicators reporting changes in instructional and learning conditions. This section examines the degree to which students are taking more advanced courses in the curriculum. Table 6.2 presents trends in the mathematics classes taken by 13-year-olds. An examination of the results shows that the profile of percentages of 13-year-old students enrolled in algebra, prealgebra, and regular grade-level mathematics in 1994 is different from that in 1986. There have been increases in the percentages of 13-year-old students studying algebra and prealgebra and a concomitant decrease in the percentages studying the regular mathematics curriculum. As would be expected, 13-year-olds pursuing higher levels of mathematics coursework in 1994 attained higher average mathematics scale scores.

---

[23] McKnight, C. C.; Crosswhite, F. J.; Dossey, J. A.; Kifer, E.; Swafford, J. O .; Travers, K. J.; & Cooney, T. J., *The Underachieving Curriculum*, (Champaign, IL: Stipes, 1987).

Westbury, I.; Ethington, C. A.; Sosniak, L. A.; & Baker, D. P.; (Eds.)., *In Search of More Effective Mathematics Education,* (Norwood, NJ: Ablex Publishing, 1994).

[24] Stevenson, H. W., & Stigler, J. W., *The Learning Gap. Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education* (New York, NY: Summit Books, 1992).

Table 6.2

Trends in Mathematics Course Taking at Age 13, 1986 and 1994

	ALGEBRA		PREALGEBRA		REGULAR MATH		OTHER	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
1994	20(1.3)	294(1.6)	32(1.5)	279(1.2)	43(1.7)	265(1.0)	4(0.4)	274(3.9)
1986	16(2.0)*	299(1.6)*	19(1.8)*	280(1.2)	61(3.0)*	261(0.9)*	5(0.5)	262(3.8)*

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 6.3 presents trends in the mathematics course-taking profile for 17-year-olds. The results in the table represent the highest level mathematics course taken by the students to date. Increases in percentages indicate that students are electing, or being required, to take more, or higher level, courses than their counterparts in 1978. Mathematics course taking is compulsory for 13-year-olds, but not always for 17-year-old students. These results are similar to those from other studies documenting a trend toward more advanced course work among high school seniors.[25]

[25] Blank, R. K., & Gruebel, D., *State Indicators of Science and Mathematics Education 1995: State-by-State Trends and New Indications from the 1993-94 School Year.* (Washington, DC: Council of Chief State School Officers, 1995).

Table 6.3

Trends in Highest Level of Mathematics Course Taken at Age 17, 1978 and 1994

		TOTAL	WHITE	BLACK	HISPANIC	MALE	FEMALE
Prealgebra or General Mathematics							
1994:	Percent	9(1.1)	9(1.1)	***(***)	***(***)	11(1.2)	8(1.2)
	Average Score	272(1.2)	275(1.4)	***(***)	***(***)	274(1.8)	268(1.9)
1978:	Percent	20(1.0)*	18(1.1)*	31(1.3)	36(3.1)	21(1.0)*	20(1.1)*
	Average Score	267(0.8)*	272(0.6)	247(1.6)*	256(2.3)*	269(1.0)*	265(0.9)
Algebra I							
1994:	Percent	15(0.9)	14(0.9)	21(2.4)	***(***)	16(1.3)	14(0.9)
	Average Score	288(1.4)	292(1.7)	275(3.3)	***(***)	289(1.6)	286(1.9)
1978:	Percent	17(0.6)	17(0.6)*	19(1.2)	19(2.1)	15(0.6)	18(0.7)*
	Average Score	286(0.7)	291(0.6)	264(1.5)*	273(2.8)*	289(0.9)	284(1.0)
Geometry							
1994:	Percent	15(0.8)	14(1.0)	17(2.2)	***(***)	15(0.9)	15(1.1)
	Average Score	297(1.7)	301(1.5)	283(3.8)	***(***)	301(2.1)	293(1.8)
1978:	Percent	16(0.6)	17(0.7)*	11(0.8)*	12(1.2)	15(0.5)	18(0.8)*
	Average Score	307(0.7)*	310(0.6)*	281(1.9)	294(4.4)	310(1.0)*	304(0.8)*
Algebra II							
1994:	Percent	47(1.6)	48(1.9)	45(3.4)	38(3.5)	44(1.9)	50(2.0)
	Average Score	316(1.0)	320(1.0)	297(2.5)	304(4.1)	320(1.5)	313(1.1)
1978:	Percent	37(1.2)*	39(1.3)*	28(2.1)*	23(2.5)*	38(1.2)*	37(1.3)*
	Average Score	321(0.7)*	325(0.6)*	292(1.4)	303(2.9)	325(0.8)*	318(0.9)*
Precalculus or Calculus							
1994:	Percent	13(1.2)	14(1.5)	***(***)	***(***)	13(1.3)	12(1.5)
	Average Score	340(2.2)	344(2.0)	***(***)	***(***)	343(2.6)	337(2.8)
1978:	Percent	6(0.4)*	6(0.4)*	4(0.6)	3(0.9)	7(0.5)*	4(0.4)*
	Average Score	334(1.4)*	338(1.1)*	297(6.5)*	306(6.1)	337(2.0)	329(1.8)*

*Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages of students may not total 100 because a small percentage of students reported having taken other mathematics courses.

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Since most students in the 17-year-old group are in the eleventh or twelfth grade, one would expect that, if they were enrolled in a typical curriculum with no interruptions in their pursuit of mathematics courses, they would be enrolled in algebra II or higher. Results for the nation show that in 1994, approximately 60 percent of the students met this expectation. The overall 1994 percentages for algebra II and precalculus or calculus reflected gains over the percentages found in 1978. The percentage of students ending their mathematics study with algebra I or geometry showed no significant difference between 1978 and 1994. However, there was a decrease in the percentage of students ending their mathematics studies with prealgebra or general mathematics.

The increase between 1978 and 1994 in the percentage of 17-year-olds reaching the algebra II level appeared to be consistent across racial/ethnic groups. White 17-year-olds displayed an increase in the percentage taking precalculus or calculus. The percentages of Black and Hispanic 17-year-olds was insufficient to statistically establish this for the precalculus or calculus level. There was a decrease in the percentage of White students ending their mathematics study at the prealgebra, algebra I, and geometry levels. Black 17-year-olds showed an increase at the geometry level.

An examination of the results by gender groups shows that there were gains in the percentages reaching the algebra II and precalculus or calculus levels. For female 17-year-olds, the percentages at the prealgebra, algebra I, and geometry levels have decreased. The corresponding percentages for males remained relatively constant, showing a decrease only in prealgebra.

Among 17-year-olds in 1994, higher average mathematics scale scores were associated with more advanced mathematics course work. Where sample sizes were sufficient for making comparisons, the differences between White and Black or White and Hispanic students were consistent across all level of mathematics course work — White students outperformed their Black or Hispanic peers. Similarly, male students outperformed female students at all levels of mathematics course work, although the difference was not significant among students whose highest level of mathematics course work was algebra I or precalculus/calculus.

Combined, the results in Tables 6.2 and 6.3 reflect a general upward movement in course taking as students face more challenging work, from the regular mathematics courses for 13-year-olds through the precalculus or calculus levels for 17-year-olds.

Trends in the Use of Technology in Mathematics Classes at Ages 13 and 17 between 1978 and 1994

Thirteen- and 17-year-olds were also asked a number of questions about the availability and use of computers in mathematics instruction. The results indicated that approximately half of the nation's students, for both 13- and 17-year-olds, had access to a computer in learning mathematics in 1994. As shown in Table 6.4, the percentages responding "Yes" to the questions in 1994 are higher than those observed in 1978 for all questions in both age groups.

At age 13, average scale scores increased between 1978 and 1994 at each level of computer use. Among 17-year-olds, no significant change in average scale scores between 1978 and 1994 was observed, execpt for a decline in the average scores for students who reported taking a computer programming course. In interpreting these findings it should be noted that the content of computer instruction and the approaches used may vary widely across schools and states.

Table 6.4

Trends in Availability and Use of Computers at Ages 13 and 17, 1978 and 1994

		STUDENTS REPORTING "YES"			
		Age 13		Age 17	
	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Had Access to Computer to Learn Mathematics	1994	48(2.0)	276(1.6)	52(2.2)	308(1.6)
	1978	12(1.8)*	262(4.1)*	24(2.7)*	314(2.9)
Studied Mathematics through Computer Instruction	1994	50(1.8)	276(1.8)	34(1.7)	309(2.6)
	1978	14(0.9)*	267(3.2)*	12(1.1)*	309(4.7)
Used a Computer to Solve Mathematics Problems	1994	70(1.3)	276(1.7)	62(1.3)	307(1.5)
	1978	56(1.4)*	268(1.8)*	46(1.5)*	303(2.1)
Took a Course in Computer Programming	1994	Question not asked at age 13.		29(1.1)	307(1.6)
	1978			10(0.9)*	318(2.0)*

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Attitudes Toward Mathematics at Ages 13 and 17 between 1978 and 1994

Students' attitudes toward mathematics, their ability to use it, and its usefulness in their world are key goals stated for the K-12 curriculum in the *NCTM Teaching Standards.* Students' beliefs about the nature of mathematics may be key to their decisions to pursue mathematics, participate in classroom activities designed to provide opportunities to learn, and persist in applying mathematics to solve problems. To explore students' views about mathematics, students were given statements and were asked to indicate their degree of agreement or disagreement with each. Table 6.5 contains a summary of 13- and 17-year-old students' responses to these statements.

Table 6.5

Trends in Attitudes Toward Mathematics at Ages 13 and 17, 1978 and 1994

			STRONGLY AGREE OR AGREE		UNDECIDED, DISAGREE, OR STRONGLY DISAGREE	
	Age	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
I would like to take more mathematics.	13	1994	42(1.3)	276(1.9)	58(1.3)	272(1.7)
		1978	50(1.5)*	263(2.6)*	51(1.5)*	268(1.4)*
	17	1994	39(1.4)	313(2.2)	61(1.4)	303(1.7)
		1978	39(1.7)	304(2.0)*	61(1.7)	295(1.7)*
I am taking mathematics only because I have to.	13	1994	24(1.2)	264(1.8)	76(1.2)	277(1.6)
		1978	29(1.4)*	256(2.4)*	71(1.4)*	270(1.9)*
	17	1994	29(1.3)	299(1.8)	71(1.3)	310(1.7)
		1978	27(1.5)	287(2.5)*	73(1.5)	302(1.8)*
I am good at mathematics.	13	1994	69(1.4)	278(1.5)	31(1.4)	264(2.0)
		1978	65(1.3)*	270(2.0)*	35(1.3)*	258(1.9)*
	17	1994	61(1.2)	313(1.7)	39(1.2)	298(1.8)
		1978	54(1.5)*	307(2.0)*	46(1.5)*	289(1.5)*
I usually understand what we are talking about in mathematics.	13	1994				
		1978	Question not asked at age 13			
	17	1994	73(1.2)	310(1.9)	27(1.2)	299(1.9)
		1978	67(1.1)*	303(1.8)*	33(1.1)*	290(2.1)*
Mathematics helps a person think logically.	13	1994	72(1.3)	276(1.6)	29(1.3)	268(2.1)
		1978	74(1.1)	268(1.9)*	26(1.1)	261(2.4)*
	17	1994	73(1.5)	310(1.6)	27(1.5)	298(2.0)
		1978	77(1.1)	301(1.7)*	23(1.1)	289(2.2)*
New discoveries are seldom made in mathematics.	13	1994	33(1.4)	268(2.0)	67(1.4)	277(1.7)
		1978	36(1.5)	255(2.2)*	64(1.5)	272(1.5)*
	17	1994	32(1.4)	300(2.1)	68(1.4)	310(1.8)
		1978	19(1.2)*	284(3.2)*	81(1.2)*	302(1.5)*

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The first statements dealt with students' experience with mathematics itself. Most 13- and 17-year-olds reported that, if they had their choice, they were either undecided or would rather not take any more mathematics courses. This was an increase in apparent dislike by 13-year-olds, but no change for 17-year-olds. It may be that the value of mathematics in their lives has not been made clear. However, nearly three-fourths of 13- and 17-year-olds reported that they were undecided or tended to disagree with the statement that they were only taking mathematics because they had to. Again, the 13-year-olds had changed their views since 1978. Here, more tended to be undecided or to disagree with the statement. This represents a move in the positive direction toward a favorable view of mathematics as a subject. For both 13- and 17-year-olds, increases were noted in students responding that they agreed or strongly agreed with the statement, "I am good at mathematics." There was also an increase in the percentage of 17-year-olds who said they agreed or strongly agreed with the statement, "I usually understand what we are talking about in mathematics."

The last two statements dealt with students' perceptions of mathematics as a discipline itself. The statement, "Mathematics helps a person think logically," was agreed to by nearly three-fourths of 13- and 17-year-olds, indicating a fairly consistent view across adolescents that mathematics provides a rational base for thinking through problems and situations. These responses were also fairly stable between 1978 and 1994. The final statement, "New discoveries are seldom made in mathematics," sampled students' views about the dynamic nature of the subject itself. Is it a changing and growing field, or is it a static body of facts and rules? A response of undecided, disagree, or strongly disagree would indicate a view that mathematics may be changing over time (that is, a view somewhat more toward the dynamic side of the balance). The percentage of 13-year-olds that responded this way in 1994 was not significantly different than the percentage in 1978. Among 17-year-olds, however, the decrease from 1978 to 1994 was statistically significant, indicating that fewer students in 1994 affirmed the dynamic nature of mathematics.

Trends in Television Watching at Ages 9, 13, and 17 between 1978/1982 and 1994

Table 6.6 presents students' reports about the amount of time they spend watching television per day. For 9- and 13-year-olds, 1982 was the first year this question was asked. For 17-year-olds, the first year was 1978. Students were asked to select the number of hours they watched television, and the data were aggregated into three categories: 0-2 hours, 3-5 hours, and 6 or more hours.

Table 6.6

Trends in Television Watching at Ages 9 and 13, 1982 and 1994; and at Age 17, 1978 and 1994

	NUMBER OF HOURS WATCHED PER DAY					
	0-2 Hours		3-5 Hours		6 or More Hours	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Age 9						
1994	43(1.0)	232(0.9)	38(0.9)	235(1.0)	19(0.8)	222(1.3)
1982	44(1.1)	218(1.4)*	29(0.6)*	227(1.1)*	26(1.0)*	215(1.2)*
Age 13						
1994	38(1.3)	282(1.3)	49(1.1)	274(1.0)	13(0.6)	257(2.4)
1982	45(0.8)*	273(1.2)*	39(0.4)*	269(1.1)*	16(0.8)*	256(1.8)
Age 17						
1994	53(1.7)	314(1.3)	39(1.3)	301(1.1)	8(0.7)	286(2.5)
1978	69(0.7)*	305(1.0)*	26(0.6)*	296(1.1)*	5(0.2)*	279(2.1)

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding. Data from 1978 are not available at ages 9 and 13.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

At age 9, there were indications that hours spent watching television had decreased between 1982 and 1994. A smaller percentage of students reported watching television 6 or more hours each day. Correspondingly, a greater percentage of 9-year-olds reported watching 3 to 5 hours each day. Little or no change was observed in the percentage who reported only 0 to 2 hours. Although there was a decrease in the percentage of 13-year-olds who reported 6 or more hours of television viewing each day, there was also a decrease in the percentage who reported watching only 0 to 2 hours. Consequently, the percentage of 13-year-olds in the middle range of television viewing (3 to 5 hours) increased between 1982 and 1994. For 17-year-olds, the results indicated increased television viewing between 1978 and 1994. A smaller percentage reported watching only 0 to 2 hours each day, and the percentages who reported 3 to 5, or 6 or more hours increased during the same period.

Across the three age groups in 1994, the relationship between hours spent watching television and average mathematics scale scores was fairly consistent. More time devoted to television viewing was associated with lower mathematics scores. For 9-year-olds, however, there was no difference between the scale scores of students who reported 0 to 2 hours and students who reported 3 to 5 hours.

Since 1986, NAEP has tracked students' responses to a question about whether their family has any rules about watching television. Table 6.7 shows there was an increase in parental rule making for 9-year-olds. The percentage of 9-year-olds responding "Yes" increased from 37 to 42 percent. The percentages of 13- and 17-year-olds responding "Yes" did not change significantly between 1986 and 1994.

Table 6.7

Trends in Students' Reports About Family Rules For Television Watching at Ages 9, 13, and 17, 1986 and 1994

	YES		NO	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Age 9				
1994	42(0.9)	231(1.2)	58(0.9)	231(1.0)
1986	37(0.7)*	220(1.0)*	63(0.7)*	223(1.2)*
Age 13				
1994	27(0.9)	275(1.5)	73(0.9)	274(1.0)
1986	27(1.2)	270(1.6)*	74(1.2)	269(1.1)*
Age 17				
1994	12(0.6)	307(2.4)	88(0.6)	306(0.9)
1986	11(0.6)	300(2.4)*	89(0.6)	303(0.8)*

*Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Doing Mathematics Homework at Age 17, 1978 and 1994

One issue often brought up in discussions about increased expectations in the classroom is the amount of homework assigned. One of the questions asked 17-year-olds was how often, in general, they do mathematics homework. Table 6.8 provides the results. The results show that the percentage of 17-year-olds who reported doing mathematics homework "often" increased between 1978 and 1994. The percentage who reported doing mathematics homework "sometimes" decreased over this time period.

In 1994, 17-year-olds who reported doing mathematics homework "often" had average mathematics scale scores that were higher than their peers who reported spending less time on homework. No significant changes between 1978 and 1994 were observed in 17-year-olds' average scale scores at any level of homework frequency. In interpreting these findings, it should be considered that students' reports on the frequency of doing homework is not an indication of the content or difficulty of the assigned homework.

Table 6.8

Trends in Frequency of Doing Mathematics Homework at Age 17, 1978 and 1994

	OFTEN		SOMETIMES		NEVER	
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
1994	74(1.5)	310(1.4)	21(1.3)	296(2.2)	5(0.6)	289(4.5)
1978	59(2.0)*	309(1.6)	35(1.9)*	291(2.1)	6(0.7)	284(3.5)

* Statistically significant difference between 1994 and the earliest year with available data, at a 5 percent combined significance level per set of comparisons. This notation indicates statistical significance without application of a multiple comparison procedure. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

Throughout the past decade, considerable effort has been expended in attempts to significantly change the context for the teaching and learning process for school mathematics and to renew the curriculum provided for students. The shift to a process-oriented curriculum, one having affective goals as well as cognitive ones, has received a great deal of attention. Accompanying these efforts has been the attempt to get *all* students involved in learning more and better mathematics. Finally, a great deal of emphasis has been given to the roles of social and family influences on mathematics.[26]

The pattern of change noted in this chapter shows (as in previous chapters) a process of slow, but overall generally positive, change toward the goals of the reform movement in school mathematics. This chapter opened with a look at trends in classroom activities. Increases in the frequency with which students listen to a teacher explain mathematics and watch a teacher work mathematics problems on the board were observed. Although these two activities may represent more passive instructional techniques, there were some modest increases reported by students in discussing mathematics in class and in making reports or doing projects. In this last category, however, the increase was from only 2 to 4 percent.

For both the 13- and 17-year-old age groups, there were reports of increased student selection of more advanced courses, with an increase in algebra and prealgebra enrollment for 13-year-olds and an increase in algebra II and precalculus or calculus enrollment for 17-year-olds. The overall course work data did indicate increased exposure to mathematics for students across the board.

The results on computer usage indicated an increase in student exposure to and use of computers in mathematics-related activities. Even in the area of computer-programming courses, students showed increases over time.

Many 13- and 17-year-olds still tend to think of mathematics as something to be endured, although they have relatively good self-concepts about their capabilities. Students' reports indicated an increase in their perceived ability to understand what was being discussed in their mathematics classes. One-third of the students see mathematics as an

---

[26] *Curriculum and Evaluation Standards for School Mathematics* (Reston, VA: National Council of Teachers of Mathematics, 1989).

Mullis, I. V. S.; Jenkins, F.; & Johnson, E. G., *Effective Schools in Mathematics* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1994).

undynamic discipline. This may not be an encouraging finding if one goal of mathematics instruction is for students to value the power of mathematics in their lives as a problem-solving tool through its representational, considerational, transformational, and computational aspects.

Trends in television watching were somewhat mixed. Results indicate a decrease in television watching among 9-year-olds but an increase among 17-year-olds. The results for 13-year-olds were mixed. However, a greater percentage of 17-year-olds were working on homework "often" in 1994, as compared to 17-year-olds in 1978.

Part III

Trends in Reading Achievement from 1971 to 1994

Introduction

To monitor progress across time in the reading achievement of American students, NAEP has conducted eight national assessments of reading performance involving nationally representative samples of 9-, 13-, and 17-year-old students. These assessments were conducted in the 1970-71, 1974-75, 1979-80, 1983-84, 1987-88, 1989-90, 1991-92, and 1993-94 school years. They will subsequently be referred to by the latter half of the school year in which they occurred — 1971, 1975, 1980, 1984, 1988, 1990, 1992, and 1994.

Concern about the literacy achievement of our nation's students continues to be a major educational, social, and political issue. Although helping students to read beyond only surface understanding has long been a goal of reading instruction, research indicates that students of all ages have difficulty reading and responding thoughtfully.[27] Because of such findings, reading

[27] Campbell, J. R.; Donahue, P. L.; Reese, C. M.; & Phillips, G. W., *NAEP 1994 Reading Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1992).

Mullis, I. V. S.; Campbell, J. R.; & Farstrup, A. E., *NAEP 1992 Reading Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1993).

Foertsch, M. F., *Reading In and Out of School* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1992).

achievement has become vital to educational reform efforts throughout the country, at both the national and state levels. *The National Education Goals Report* has highlighted goals for competency in English and increased levels of literacy and lifelong learning.[28] Clearly, these goals are attainable only if students develop the literacy abilities, attitudes, and habits that characterize expert readers.

During the last two decades, the field of reading has witnessed tremendous debate over issues related to reading theory and reading instruction.[29] These discussions have resulted in an increased awareness of the complexities of the reading process and of how important the context of learning to read is to the developing reader. In addition, these concerns about how reading is viewed and taught have led to intense interest in how achievement in reading is assessed.[30] What has emerged predominantly is a recognition that reading instruction and reading assessment must be approached from an integrative and contextual perspective in order to promote the more advanced literacy skills necessary for today's society.[31] These current issues provide a dynamic context for examining and interpreting the results of NAEP's reading trend assessments.

As a whole, Part III of this report is intended to serve as a resource for groups concerned with improving students' reading achievement — not only reading teachers and researchers, but also educators in other subjects, policy makers, school administrators, and parents. Together with information from other sources, the findings provide a basis for discussing the adequacy of students' current reading achievement, in light of the factors that seem to be related to reading ability. These discussions may then lead to the development of means for improving reading performance in the years ahead.

The NAEP reading trend assessments incorporate a wide range of text materials, from simple narrative passages to complex articles on specialized topics.[32] The selections include stories, poems, essays, reports, and passages from textbooks of varying levels of difficulty, as well as sample train schedules, telephone bills, and advertisements. Students' comprehension is assessed with a variety of question types. Some multiple-choice questions

---

[28] *The National Education Goals Report: Building a Nation of Learners.* (Report of the National Education Goals Panel: U.S. Government Printing Office, 1993).

[29] Smith, F., "Learning to Read: The Never-Ending Debate" *Phi Delta Kappan,* 432-441, February, 1992.

[30] Winograd, P.; Paris, S.; & Bridge, C., "Improving the Assessment of Literacy" *The Reading Teacher, 45*, 108-115, October, 1991.

[31] Hiebert, E. H. (Ed.), *Literacy for a Diverse Society: Perspectives, Practices, and Policies.* (New York: Teachers College, Columbia University, 1991).

[32] *Reading Objectives, 1983-84 Assessment* (Princeton, NJ: Educational Testing Service).

require students to identify particular information presented in the text. Constructed-response questions require students to restructure and interpret what they have read and to present their responses in writing. In order that performance trends over time can be measured, the same sets of reading materials and questions are administered in each assessment.

Students participating in each assessment were asked to provide information on their demographic characteristics, instructional experiences, and reading attitudes and behaviors. The relationships observed between reading performance and self-reported background information can help educators, reading researchers, and policy makers to identify and discuss central issues and concerns and can guide further inquiries.

The NAEP 1994 long-term trend reading assessment measuring trends since 1971 is separate from another reading assessment conducted in 1992 and 1994. The assessment conducted in 1992 and 1994 is based on a new framework representing an innovative approach to measuring reading achievement.[33] The 1992 and 1994 assessment made use of instructionally relevant activities involving the use of authentic reading materials and contained predominantly constructed-response questions.[34] The 1992 and 1994 assessments were designed both to be responsive to the needs of the Trial State Assessments and to reflect current research on effective reading instruction and assessment. Students participating in the newly developed 1992 and 1994 reading assessments were selected by grade definitions and completed the assessment at a different time of year than did students participating in the trend assessment. The 1994 results from the newly developed assessment are published in a previously released reading report, *NAEP 1994 Reading Report Card for the Nation and the States*.[35] Because of the many differences between the two reading assessments, the results are not directly comparable.

NAEP uses analysis techniques based on item response theory (IRT) to estimate students' reading ability on a scale ranging from 0 to 500. The NAEP reading scale is useful in making comparisons across assessments for the three age groups and among subpopulations of students. (The Procedural

---

[33] This newly developed reading assessment includes longer reading materials, such as those actually used by readers in their everyday reading, and a large number of constructed-response questions. The newly developed materials were also used for the 1992 and 1994 Trial State Assessment Programs in reading at Grade 4.

[34] *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress*, National Assessment Governing Board (Washington DC: U.S. Government Printing Office).

[35] Campbell, J. R.; Donahue, P. L.; Reese, C. M.; & Phillips, G.W., *NAEP 1994 Reading Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1996).

Appendix contains more detailed information about analysis procedures and student subgroups.) To provide a basis for interpreting the results, the report describes what students attaining different performance levels on the scale are able to do. Based on the assessment results, five levels of reading performance were defined: Level 150 – Simple, Discrete Reading Tasks; Level 200 – Partially Developed Skills and Understanding; Level 250 – Interrelate Ideas and Make Generalizations; Level 300 – Understand Complicated Information; and Level 350 – Learn from Specialized Reading Materials. Essentially, students performing at Level 150 were able to carry out simple, discrete reading tasks. At Level 200, students demonstrated partial skills and basic understanding of what they read. Performance at Level 250 suggests the ability to search for specific information, interrelate ideas, and make generalizations. Students performing at Level 300 were able to find, understand, summarize, and explain relatively complicated information. Those performing at Level 350 showed some ability to synthesize and learn from specialized reading materials.

NAEP reports the performance of groups and subgroups of students, not individuals. The measures of achievement included in this report are the average performance of groups of students on the NAEP reading scale and the percentages of students attaining successive levels of performance on the scale. Because the average scale scores and the percentages presented in this report are based on samples, they are necessarily estimates. Like all estimates based on surveys, they are subject to sampling as well as measurement error. To compute standard errors, NAEP uses a complex procedure that estimates the sampling error and other random error associated with observed assessment results.

In the tables and figures that present reading trend results, statistically significant differences between 1994 and prior assessments are denoted with an asterisk; statistically significant differences between 1971 and subsequent assessments are denoted with a dagger. In addition to comparisons between individual assessment years, a second test of significance was conducted to detect statistically significant linear and quadratic trends across the assessments. (See the Procedural Appendix for a discussion of the procedure.) This type of analysis allows for the discussion of statistically significant patterns that may be missed by year-to-year comparisons. For example, from year-to-year, students' average scale scores may consistently increase by a small amount. Although these small increases between years may not be statistically significant, the overall increasing trend in average scores may be statistically significant and noteworthy. All of the differences and trend patterns discussed in this report are statistically significant at the .05 level.

Each chapter in Part III provides a somewhat different perspective on trends in students' reading abilities. Chapter 7 describes changes in the average reading performance of 9-, 13-, and 17-year-olds across the eight reading trend assessments conducted by NAEP since 1971. In Chapter 8, levels of reading performance are defined, and the percentages of students attaining successive levels in each assessment are presented. Chapter 9 summarizes trends in students' responses to questions about their reading instruction and experiences and investigates the relationships between these background factors and reading achievement.

7

National Trends in Reading Scale Scores

An examination of the trends in performance of students participating in NAEP provides valuable information about literacy development in the United States over the past 23 years. During the same period, the reading curriculum in this country has changed in response to the ever-growing body of research on reading and reading instruction. For example, the past 20 years have seen a movement away from a mostly skills-oriented view of reading to a more cognitive view emphasizing processes of comprehension.[36] Considering the performance of students on the NAEP reading trend assessment in the context of changing curricular emphases can enhance the relevance and interpretation of these data.

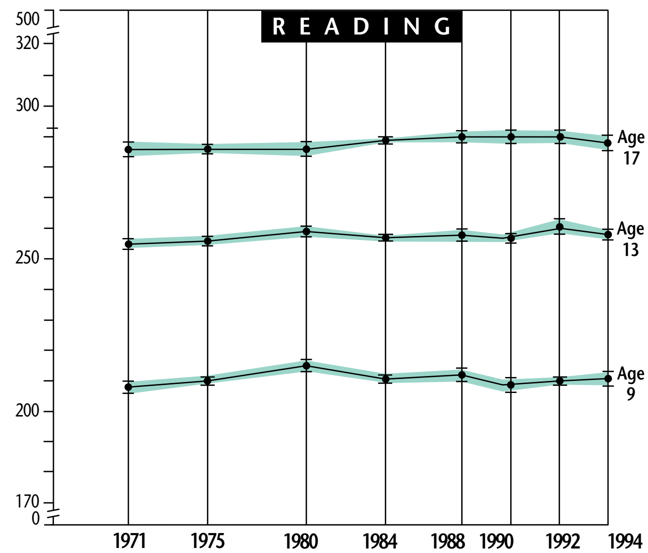
The results of the eight trend assessments conducted from 1971 to 1994 are presented in Figure 7.1. For all three age groups, only minimal changes have been observed across the trend assessments.

---

[36] Dole, J. A.; Duffy, G. G.; Roehler, L. R.; & Pearson, D. P. "Moving from the Old to the New: Research on Reading Comprehension Instruction." *Review of Educational Research, 61,* 239-264, 1991.

Figure 7.1

Trends in Average Reading Scale Scores for the Nation, 1971 to 1994



	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	285(1.2)	286(0.8)	286(1.2)	289(0.6)	290(1.0)†	290(1.1)†	290(1.1)†	288(1.3)
Age 13	255(0.9)	256(0.8)	259(0.9)	257(0.5)	258(1.0)	257(0.8)	260(1.2)†	258(0.9)
Age 9	208(1.0)	210(0.7)	215(1.0)†	211(0.7)†	212(1.1)†	209(1.2)	211(0.9)	211(1.2)

95-percent confidence interval

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Nine-Year-Olds. Compared to 1971, significantly higher average scores were attained by 9-year-olds during the 1980s. However, performance declined somewhat in the early 1990s, resulting in a 1994 average score that was not significantly different from the 1971 average score.

Thirteen-Year-Olds. Results for 13-year-olds show some improvement in performance across the assessment years. In 1992, these students attained an average score that was higher than the average score in 1971. However, performance in 1994 returned to a level not significantly different from that observed in 1971.

Seventeen-Year-Olds. The performance of 17-year-olds improved slightly across the assessment years. From 1988 to 1992, these students had average scores that were higher than the average score in 1971. However, the 1994 average score returned to a level that did not differ significantly from the 1971 average score.

Trends in Reading Scale Scores from 1971 to 1994 by Quartiles

Table 7.1 presents average reading scale scores for 9-, 13-, and 17-year-old students who were in the upper quartile (upper 25 percent), the middle two quartiles (middle 50 percent), and the lower quartile (lower 25 percent) of student performance in each assessment. These data reveal changes that have occurred in the last 23 years for students at different points along the distribution, demonstrating whether overall gains or losses were evident across the full range of reading performance in the student population. This information is particularly relevant in light of Goal 3 of *The National Educational Goals Report*, which states that "the academic performance of elementary and secondary students will increase significantly in every quartile. . . ."[37] The report emphasizes that students of all abilities should be granted access to educational opportunities and should demonstrate gains in educational achievement. As can be seen by the data in Table 7.1, varied patterns of change were evident for the three age groups across the performance distribution.

---

[37] *The National Education Goals Report,* National Education Goals Panel (Washington, DC: U.S. Government Printing Office, 1992).

At age 9, students in the upper quartile had higher average scores in 1994 than did their counterparts in 1971. It is worth noting, however, that while the overall trend for 9-year-olds in the upper quartile was up, scores have declined since 1990. Nine-year-olds in the middle two and lower quartiles demonstrated gains from 1971 to 1980. Since that time, their average scores have declined, but remained higher in 1994 than they were in 1971.

At age 13, students in the upper and middle two quartiles made gains across the assessment years and attained an average score in 1994 that was higher than that in 1971. Thirteen-year-olds in the lower quartile in 1980 and 1988 scored at higher levels than did their counterparts in 1971. However, recent declines in performance resulted in a 1994 average score that did not differ significantly from the 1971 average.

The scores of 17-year-olds in the upper quartile followed an interesting pattern. Between 1971 and 1980 scores for this population declined. Since that time scores have shown a pattern (though not a steady pattern) of improvement. The end result was that scores in 1994 were higher than were scores in 1971. Although a slight upward trend was suggested by the results for 17-year-olds in the middle two quartiles, the 1994 average score was not significantly different from the 1971 average. In the lower quartile, the average score of 17-year-olds increased from 1971 to 1988 and, despite a decline over the last three assessments, remained higher in 1994 than in 1971.

Table 7.1

Trends in Average Reading Scale Scores by Quartiles, 1971 to 1994

		AVERAGE SCALE SCORES		
Quartile	Year	Age 9	Age 13	Age 17
Upper Quartile	1994	256(1.0)†	301(0.8)†	337(1.3)†
	1992	256(0.9)†	303(1.1)†	335(0.9)
	1990	261(1.1)*†	297(0.8)*†	336(1.1)
	1988	259(1.6)†	296(1.0)*	330(1.3)*
	1984	258(0.4)†	296(0.5)*†	331(0.5)*
	1980	255(0.8)	294(0.5)*	327(0.8)*†
	1975	251(0.7)*	296(0.4)*†	334(0.5)
	1971	253(0.5)*	293(0.4)*	333(0.6)*
Middle Two Quartiles	1994	213(0.7)†	260(0.8)†	291(0.7)
	1992	212(0.7)	262(0.6)†	293(0.7)†
	1990	209(0.6)*	258(0.5)	292(0.5)†
	1988	213(0.7)†	259(0.7)	292(0.7)†
	1984	212(0.3)	258(0.2)	291(0.3)
	1980	218(0.3)*†	261(0.3)†	289(0.4)
	1975	213(0.3)†	258(0.4)	288(0.4)*
	1971	211(0.4)*	258(0.4)*	289(0.5)
Lower Quartile	1994	162(1.4)†	210(0.9)	234(1.3)†
	1992	162(1.0)†	212(1.4)	237(1.3)†
	1990	157(1.5)	215(0.9)*	241(1.6)*†
	1988	163(1.6)†	217(1.0)*†	246(1.1)*†
	1984	162(0.6)†	215(0.5)*	241(0.3)*†
	1980	169(1.0)*†	219(0.7)*†	238(1.0)†
	1975	163(0.5)†	211(0.5)	232(1.0)
	1971	157(0.7)*	212(0.7)	230(0.8)*

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Reading Scale Scores from 1971 to 1994 by Race/Ethnicity and Gender

Race/Ethnicity. Figure 7.2 shows trends in average reading scale scores for White, Black, and Hispanic students.[38] For White and Black students, results are reported from the first trend reading assessment in 1971. For Hispanic students, results are reported from 1975, the first year in which the sample allowed an accurate estimate of the scores for this population.

Among White students, the average score of 9-year-olds increased from 1971 to 1980. Since that time, performance declined slightly in the early 1980s and then remained relatively stable through the remainder of the 1980s and early 1990s, so that in 1994 the average score was not significantly different from the average in 1971. The average score of 13-year-old White students showed a general though uneven pattern of increase between 1971 and 1994. The average score in 1994 was significantly higher than was the score in 1971. Although a pattern of increased performance across the assessment years was observed in the results for 17-year-old White students, the 1994 average score did not differ significantly from the 1971 average score.

In all age groups, Black students demonstrated a pattern of increased performance through the 1970s and 1980s followed by a period of decline in the early 1990s. However, the 1994 average score in each age group remained higher than the 1971 average.

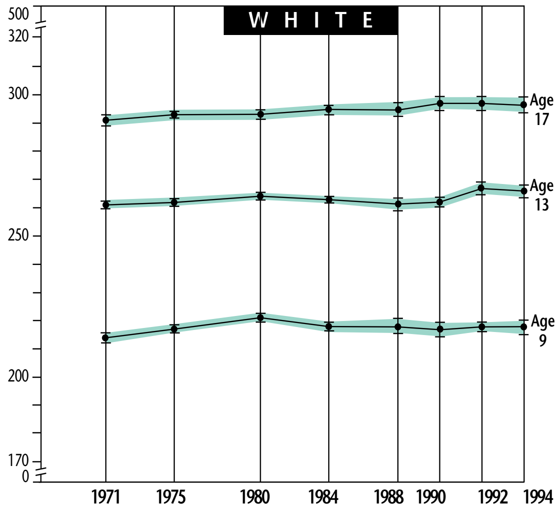
The average score of 9-year-old Hispanic students has not changed significantly across the trend assessments. Although there was evidence of a slight increase in the performance of 13-year-old Hispanic students during the 1980s, this trend has not continued. There was no significant difference between 1994 and 1975 average scores for these students. The average score of 17-year-old Hispanic students increased from 1975 to 1990. Since that time, a decline has been observed, and the 1994 average score returned to a level not significantly different from that in 1975.

---

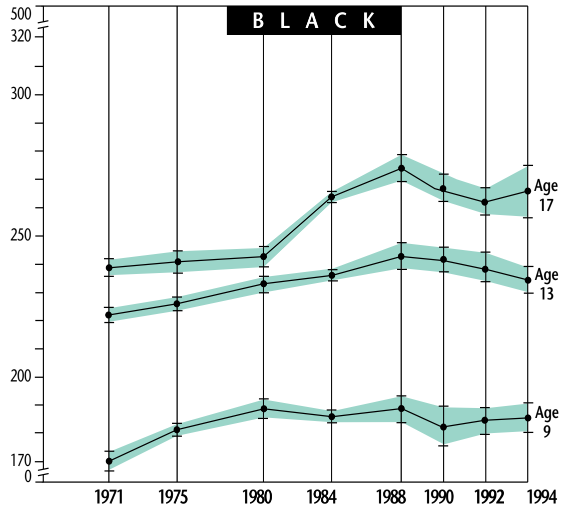
[38] For Asian/Pacific Islander students and American Indian students, the sample sizes were insufficient to permit reliable trend estimates.

Figure 7.2

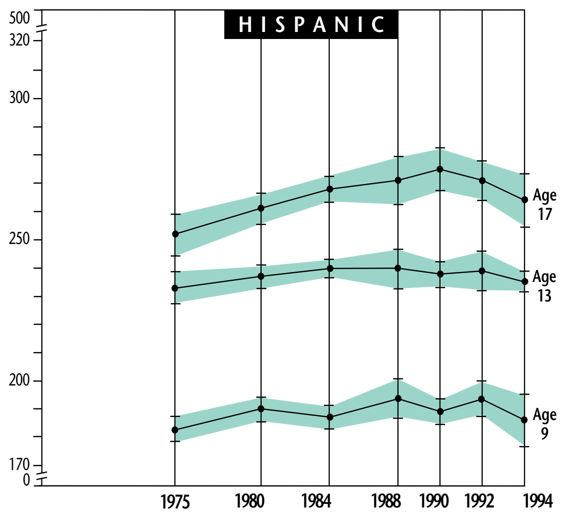
Trends in Average Reading Scale Scores by Race/Ethnicity, 1971 to 1994



	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	291(1.0)	293(0.6)	293(0.9)	295(0.7)†	295(1.2)	297(1.2)†	297(1.4)†	296(1.5)
	87(1.3)*	84(1.0)*	83(1.6)*	77(0.6)*†	77(0.6)*†	74(0.5)†	75(0.6)*†	72(0.8)†
Age 13	261(0.7)*	262(0.7)	264(0.7)†	263(0.6)	261(1.1)	262(0.9)	266(1.2)†	265(1.1)†
	84(1.4)*	81(1.2)*	80(1.8)*	77(0.6)*†	76(0.7)*†	74(0.8)†	73(0.7)†	74(0.5)†
Age 9	214(0.9)	217(0.7)	221(0.8)†	218(0.8)†	218(1.4)	217(1.3)	218(1.0)†	218(1.3)
	84(1.4)*	80(1.2)*	79(1.3)†	75(1.2)†	75(1.0)†	74(1.0)†	74(0.9)†	75(0.7)†



	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	239(1.7)*	241(2.0)*	243(1.8)*	264(1.0)†	274(2.4)†	267(2.3)†	261(2.1)†	266(3.9)†
	11(1.2)*	11(0.8)*	12(1.4)	14(0.2)*	15(0.3)†	16(0.3)†	15(0.3)	15(0.3)†
Age 13	222(1.2)*	226(1.2)*	233(1.5)†	236(1.0)†	243(2.4)†	242(2.2)†	238(2.3)†	234(2.4)†
	15(1.4)	13(0.9)	14(1.3)	14(0.2)	15(0.3)	15(0.2)	16(0.4)	15(0.5)
Age 9	170(1.7)*	181(1.2)†	189(1.8)†	186(1.1)†	189(2.4)†	182(2.9)†	185(2.2)†	185(2.3)†
	14(1.3)	13(0.8)	14(1.0)	16(0.5)	16(0.7)	16(0.6)	16(0.4)	15(0.6)



	1975	1980	1984	1988	1990	1992	1994
Age 17	252(3.6)	261(2.7)	268(2.2)†	271(4.3)†	275(3.6)†	271(3.7)†	263(4.9)
	3(0.6)*	4(0.6)*	7(0.7)†	6(0.5)*†	7(0.4)†	8(0.6)†	8(0.5)†
Age 13	233(3.0)	237(2.0)	240(1.7)	240(3.5)	238(2.3)	239(3.5)	235(1.9)
	5(0.8)*	6(1.0)	7(0.7)	6(0.6)	8(0.5)†	7(0.4)†	8(0.4)†
Age 9	183(2.2)	190(2.3)	187(2.1)	194(3.5)	189(2.3)	192(3.1)	186(3.9)
	5(0.8)	6(0.8)	7(1.4)	6(1.0)	6(0.6)	7(0.8)	6(0.4)

Below each average score, the corresponding percentage of students is presented.

95-percent confidence interval

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1971, (for Hispanics, 1975) at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

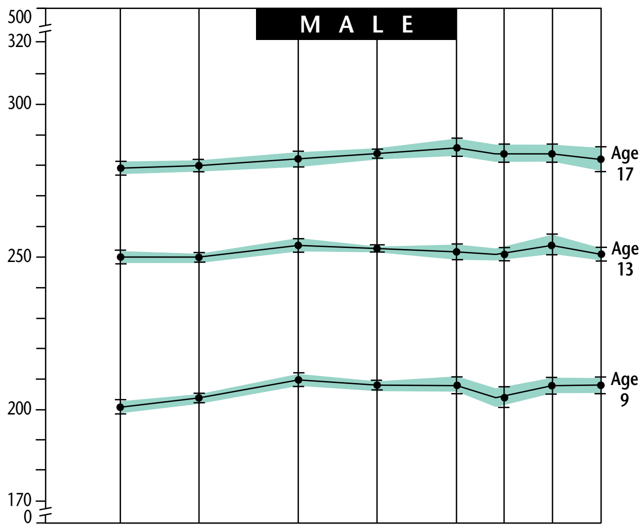
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

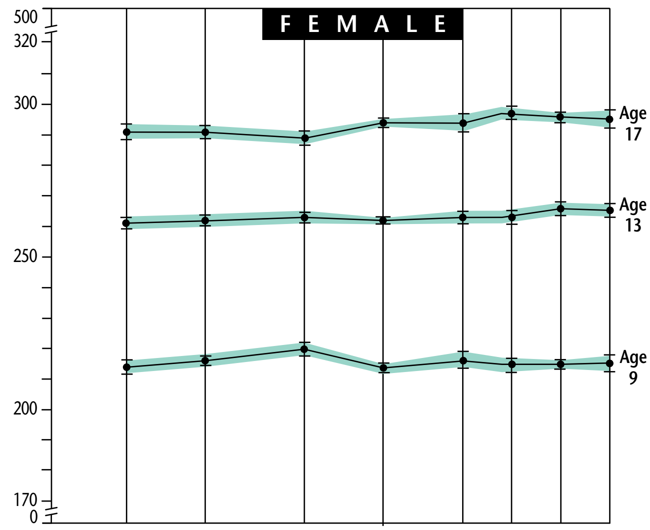
Gender. Figure 7.3 presents trends in average reading scale scores by gender. The performance of male 9-year-olds was, on average, higher in the 1980s than in 1971. Despite some fluctuation since that time, the average score of these students in 1994 remained higher than the average score of their counterparts in 1971. In the early 1980s, 13-year-old male students demonstrated increased performance over 1971. Their scores have fluctuated since that time, and in 1994 the average was not significantly different from that in 1971. The average score of 17-year-old males increased from 1971 to 1988. A slight decrease in performance since that time resulted in a 1994 average score that did not differ significantly from the 1971 average.

Among female students, 9-year-olds displayed an increase between 1971 and 1980, followed by a decrease in performance that resulted in a 1994 average score that was not significantly different from the 1971 average. The pattern of performance for 13-year-old females was one of overall improvement resulting in a higher average score in 1994 than in 1971. Although increased performance was the overall pattern for 17-year-old females, the average score in 1994 did not differ significantly from the 1971 average.

Figure 7.3

Trends in Average Reading Scale Scores by Gender, 1971 to 1994





Male

	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	**279(1.2)** 49(0.5)	**280(1.0)** 49(0.5)	**282(1.3)** 51(0.6)	**284(0.6)†** 51(0.7)	**286(1.5)†** 48(1.3)	**284(1.6)** 50(0.6)	**284(1.6)** 52(0.9)†	**282(2.2)** 50(1.4)
Age 13	**250(1.0)** 50(0.4)	**250(0.8)** 50(0.5)	**254(1.1)†** 49(0.3)	**253(0.6)†** 51(0.5)	**252(1.3)** 50(0.8)	**251(1.1)** 50(0.9)	**254(1.7)** 49(0.9)	**251(1.2)** 51(0.8)
Age 9	**201(1.1)*** 50(0.4)	**204(0.8)** 50(0.4)	**210(1.1)†** 50(0.4)	**208(0.8)†** 50(0.5)	**208(1.4)†** 50(0.8)	**204(1.7)** 51(0.9)	**206(1.3)†** 51(0.7)	**207(1.3)†** 50(0.8)

Female

	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	**291(1.3)** 51(0.5)	**291(1.0)** 51(0.5)	**289(1.2)*** 50(0.6)	**294(0.8)** 49(0.7)	**294(1.5)** 52(1.4)	**297(1.2)†** 50(0.6)	**296(1.1)** 48(0.9)†	**295(1.5)** 50(1.4)
Age 13	**261(0.9)*** 50(0.4)	**262(0.9)** 50(0.5)	**263(0.9)** 51(0.3)	**262(0.6)*** 49(0.5)	**263(1.0)** 51(0.8)	**263(1.1)** 50(0.9)	**265(1.2)†** 51(0.9)	**266(1.2)†** 49(0.8)
Age 9	**214(1.0)** 50(0.4)	**216(0.8)** 50(0.4)	**220(1.1)*†** 50(0.4)	**214(0.8)** 50(0.5)	**216(1.3)** 50(0.8)	**215(1.2)** 49(0.9)	**215(0.9)** 49(0.7)	**215(1.4)** 50(0.8)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The previous section discussed the trends in reading achievement for White, Black, and Hispanic students, and for male and female students. As with past NAEP assessments, significant differences were observed between racial/ethnic subgroups and between males and females. Figure 7.4 presents trends in differences between the average scale scores for selected subgroups of students across the assessment years.

An examination of performance in 1994 among the three ethnic groups showed that at all ages, White students outperformed their Black and Hispanic peers. The gap between average scores of White and Black students at all three ages narrowed between 1971 and 1988. This trend was the result of average scores for Black students increasing 18 points for 9-year-olds, 21 points for 13-year-olds, and 36 points for 17-year-olds (increases based on unrounded scale scores). In comparison, the average scores for White students increased no more than 4 points or remained relatively stable. During the 1990s, there was evidence that this trend toward smaller gaps has reversed due to decreasing scores for Black students. However, in 1994 the gap between White and Black students at ages 9 and 17 remained significantly smaller than the gap in 1971.

The gap between White and Hispanic students aged 9 and 13 was relatively consistent across the assessment years. At age 17, the magnitude of the gap decreased from 1975 to 1990 as the average score for Hispanic students increased 22 points and the average for White students increased 4 points. Due to decreasing performance among Hispanic students since 1990, however, the gap between White and Hispanic 17-year-olds' average scores returned to a level in 1994 that was not significantly different from that in 1975.

Consistent with other studies documenting differences in literacy development between males and females, the NAEP trend assessments revealed a continued disparity favoring female students.[39] Across the assessment years, the difference between the average scores of 9-year-old males and females decreased slightly, as males gained 6 points and females gained 1 point between 1971 and 1994. However, in 1994 the gap between

---

[39] Campbell, J. R.; Donahue, P. L.; Reese, C. M.; & Phillips, G. W., *NAEP 1994 Reading Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1996).

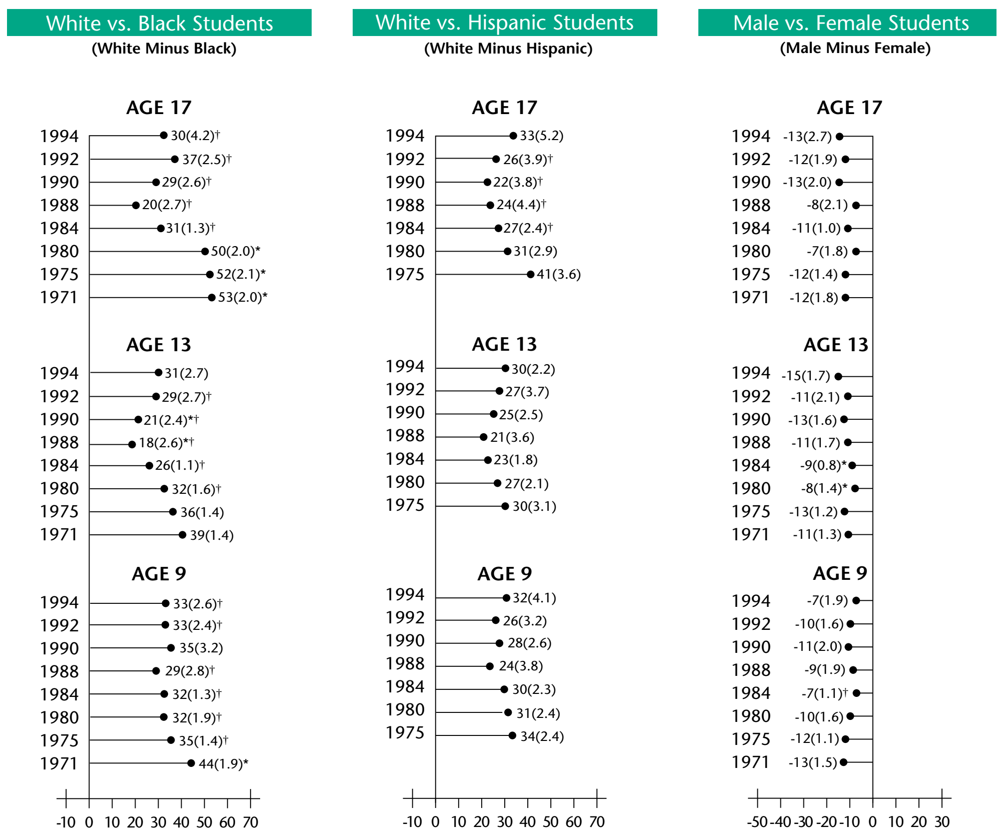
Mullis, I. V. S.; Campbell, J. R.; & Farstrup, A. E., *NAEP 1992 Reading Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, U.S. Government Printing Office, 1993).

Plewis, I., "Pupils' Progress in Reading and Mathematics During Primary School: Associations with Ethnic Group and Sex," *Educational Research, 33,* 133-140 (Summer, 1992).

9-year-old males and females was not significantly different than the gap observed in 1971. At ages 13 and 17, there were indications that the gaps between males and females had decreased slightly between 1971 and 1980, but have since increased or remained stable, so that the gaps in 1994 did not differ significantly from those in 1971.

Figure 7.4

Trends in Differences in Average Reading Scale Scores



*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores for White vs. Black student differences and Male vs. Female student differences) or from 1977 (for White vs. Hispanic student differences), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale score differences appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Reading Scale Scores from 1971 to 1994 by Region

Figure 7.5 presents trends in average reading scale scores for students from the Northeast, Southeast, Central, and West regions of the country.

Except for an increase in 1980 compared to 1971, the average score of 9-year-olds in the Northeast has changed little across the assessment years. After a period of relative stability during the 1970s and 1980s, the average score of 13-year-olds in the Northeast increased in 1992 and 1994. By 1994, scores were higher than during the 1970s and 1980s. An overall pattern of increased performance was observed in the scores of 17-year-olds, although the 1994 average score was not significantly different from the 1971 average.

In the Southeast, the average score of 9-year-olds was higher during the 1980s than it had been in 1971. Despite some fluctuation over the last three assessments, the 1994 average score for these students remained higher than the 1971 average score. Among 13-year-olds, a pattern of increased performance was observed from 1971 to 1988. Subsequent declines in performance resulted in a 1994 average score that did not differ significantly from the average score in 1971. The average score of 17-year-olds also increased between 1971 and 1988. Although the pattern of increase did not continue in the early 1990s, the 1994 average remained higher than the 1971 average score.

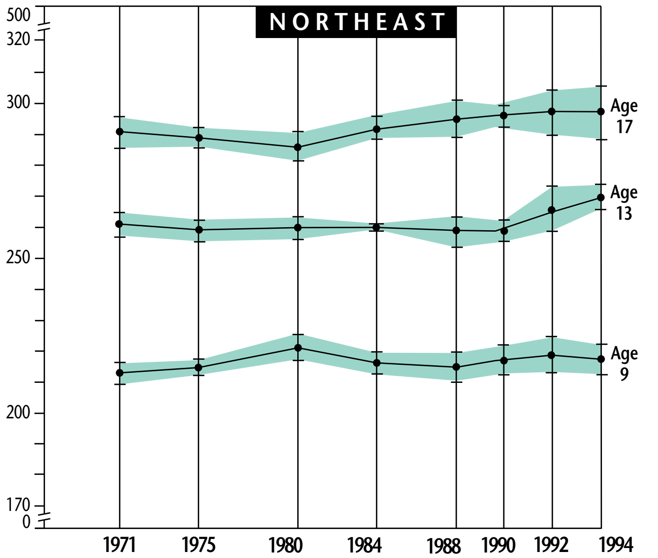
Despite slight fluctuations, the average scores of students in the Central region have not changed significantly across the assessment years. The 1994 average scores of students in each group did not differ from those observed in 1971.

In the West region, 9-year-olds attained an average score in 1980 that was higher than that in 1971. Since that time the average score has decreased, and no significant difference was observed between the 1994 and 1971 average scores. No significant changes were observed across the assessment years for 13-year-olds in the West region. At age 17, despite a pattern of improved performance across the assessment years, the 1994 average score remained at a level not significantly different from that in 1971.

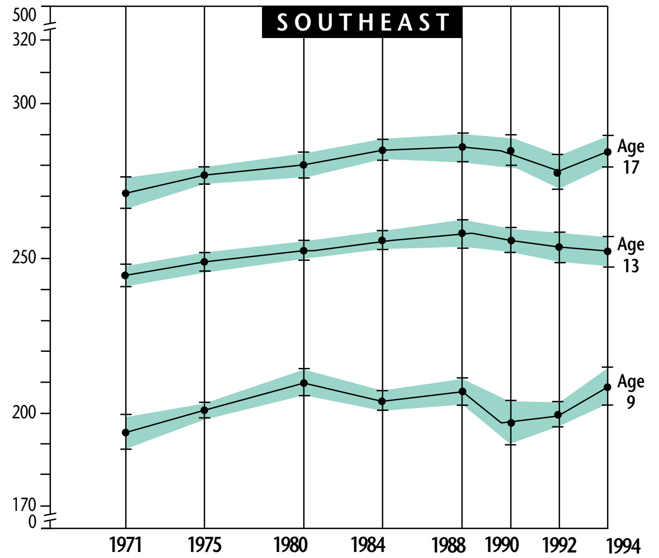
In 1994, comparisons of average scale scores for each age group indicated four instances of differences between regions. At age 9, students in the Northeast performed at a higher level than their peers in the West. At age 13, students in the Northeast performed at a significantly higher level than their peers in the Southeast and the West. And among 17-year-olds, students from the Northeast outperformed their peers in the Southeast.

Figure 7.5

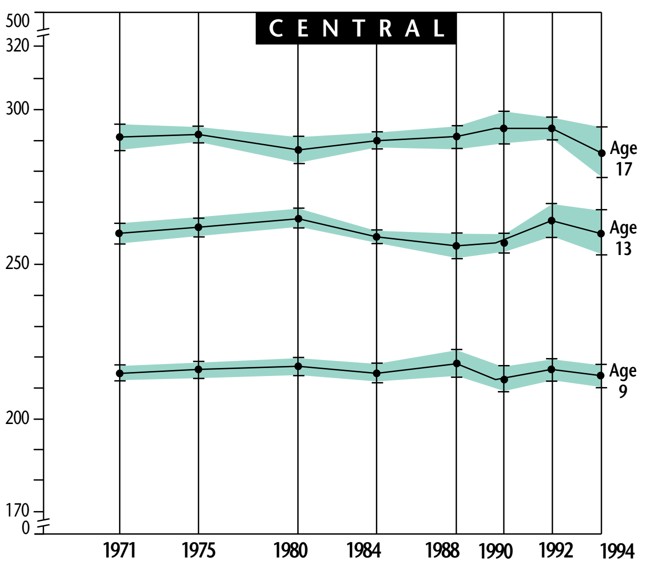
Trends in Average Reading Scale Scores by Region, 1971 to 1994



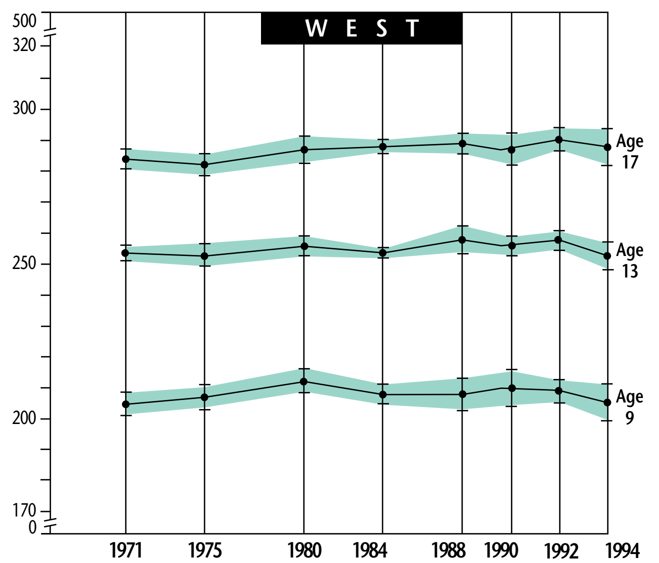
	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	**291(2.8)**	**289(1.7)**	**286(2.4)**	**292(1.9)**	**295(2.9)**	**296(1.8)**	**297(3.2)**	**297(4.2)**
	26(1.9)	26(1.2)	22(2.9)	24(0.4)	23(0.8)	23(1.1)	22(0.9)	22(1.2)
Age 13	**261(2.0)***	**259(1.8)***	**260(1.8)***	**260(0.6)***	**259(2.4)***	**259(1.8)***	**265(3.2)**	**269(2.0)†**
	24(1.5)	25(1.1)*	24(2.2)	23(0.4)*	23(2.2)	23(1.0)	22(0.8)	20(0.7)
Age 9	**213(1.7)**	**215(1.3)**	**221(2.1)†**	**216(1.7)**	**215(2.6)**	**217(2.2)**	**218(2.6)**	**217(2.9)**
	26(1.5)	27(1.1)*	23(2.1)	22(0.4)†	22(1.5)	22(1.1)	21(1.3)	22(1.0)



	1971	1975	1980	1984	1988	1990	1992	1994
	271(2.4)*	**277(1.4)**	**280(2.2)†**	**285(1.6)†**	**286(2.1)†**	**285(2.5)†**	**278(2.9)**	**284(2.8)†**
	20(1.8)	20(0.8)*	21(1.9)	22(1.8)	23(1.0)	23(0.9)	25(1.4)	25(1.7)
	245(1.7)	**249(1.5)**	**253(1.6)†**	**256(1.5)†**	**258(2.2)†**	**256(2.2)†**	**254(2.5)†**	**253(2.5)**
	24(2.2)	22(1.1)	24(2.7)	23(1.7)	24(1.4)	23(0.8)	23(1.6)	22(1.6)
	194(2.9)*	**201(1.2)**	**210(2.3)†**	**204(1.6)†**	**207(2.1)†**	**197(3.2)**	**199(2.0)**	**208(3.0)†**
	22(1.7)	22(0.8)	24(2.3)	24(1.2)	25(1.7)	24(1.2)	23(1.4)	24(1.5)



	1971	1975	1980	1984	1988	1990	1992	1994
Age 17	**291(2.1)**	**292(1.4)**	**287(2.2)**	**290(1.4)**	**291(1.9)**	**294(2.4)**	**294(2.1)**	**286(3.7)**
	30(2.4)	30(1.2)	28(3.2)	27(1.7)	26(1.8)	27(1.0)	25(0.8)	26(1.4)
Age 13	**260(1.8)**	**262(1.4)**	**265(1.4)**	**259(1.0)**	**256(2.0)**	**257(1.5)**	**264(3.0)**	**260(3.3)**
	29(2.2)	28(1.3)	26(3.5)	27(1.6)	26(2.0)	24(0.7)	26(1.0)	27(1.6)
Age 9	**215(1.2)**	**216(1.2)**	**217(1.4)**	**215(1.5)**	**218(2.2)**	**213(2.0)**	**216(1.6)**	**214(2.3)**
	29(2.0)	26(1.2)	25(3.4)	27(1.3)	24(0.6)	25(0.7)	28(1.1)	26(1.0)



	1971	1975	1980	1984	1988	1990	1992	1994
	284(1.8)	**282(1.9)**	**287(2.1)**	**288(1.1)**	**289(1.8)**	**287(2.6)**	**290(2.3)**	**288(2.8)**
	24(2.4)	25(1.1)	28(2.3)	26(0.6)	28(1.1)	28(0.9)	29(1.0)	27(0.9)
	254(1.3)	**253(1.7)**	**256(2.0)**	**254(0.9)**	**258(2.1)**	**256(1.6)**	**258(1.6)**	**253(2.1)**
	23(2.3)*	26(1.1)*	26(2.3)	27(0.5)*	28(1.0)	30(0.9)	29(1.2)	31(1.1)†
	205(2.0)	**207(2.0)**	**213(1.8)†**	**208(1.5)**	**208(2.6)**	**210(2.8)**	**209(2.3)**	**205(2.8)**
	23(1.9)	25(1.1)	28(2.1)	28(0.5)	29(1.1)	29(0.8)†	27(0.9)	28(1.0)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Reading Scale Scores from 1971 to 1994 by Parents' Highest Level of Education

Recent concerns for the literacy development of "at-risk" students (that is, students who are in circumstances that have been associated with educational difficulties) has focused attention on family and environmental influences.[40] One important indicator of the nature of children's literacy environment may be the level of their parents' education.

Table 7.2 presents information regarding levels of parents' education reported by students and the average student reading scale scores associated with them. It is noteworthy that there has been an overall decline since 1971 in the percentages of students at all ages who reported that both of their parents had not finished high school. A corresponding increase was observed in the percentage of students at all ages who reported that at least one of their parents had pursued post-high school education.

For 9-year-olds who reported the highest level of parental education (post-high school), the overall pattern was one of slight decreases in performance. However, the 1994 average score for this group did not differ significantly from the 1971 average score. The average score of 9-year-olds who reported that high school graduation was their parents' highest level of education increased from 1971 to 1980. Since that time, performance has declined, and in 1994 the average score was not significantly different from that in 1971. Except for a higher average score in 1984 compared to 1971, the performance of 9-year-olds who reported the lowest level of parental education (less than high school) has remained relatively stable across the assessment years.

The overall pattern observed for 13-year-olds at the highest level of parental education has been one of slightly decreasing performance, especially from 1971 to 1988. However, the 1994 average score did not differ significantly from that in 1971. Similar results were observed for 13-year-olds who reported that the highest level of parental education was high school graduation.[41] For 13-year-olds at the lowest level of parental education, a pattern of increased performance that occurred from 1971 to 1988 has reversed, resulting in a 1994 average score that was not significantly different from the 1971 average.

---

[40] Langer, J. (Ed.) *Language, Literacy, and Culture: Issues of Society and Schooling.* (Norwook, NJ: Ablex, 1987).

Snow, C.; Barnes, W.; Chandler, J.; Goodman, I.; & Hemphill, L., *Unfilled Expectations: Home and School Influences on Literacy.* (Cambridge, MA: Harvard University Press, 1991).

[41] Except that among this group, the 1990 average was significantly lower than the 1971 average.

Table 7.2

Trends in Average Reading Scale Scores by Parents' Highest Level of Education, 1971 to 1994

Level of Education	Year	AGE 9 Percent of Students	AGE 9 Average Scale Score	AGE 13 Percent of Students	AGE 13 Average Scale Score	AGE 17 Percent of Students	AGE 17 Average Scale Score
Post-High School	1994	46(1.2)†	221(1.3)	57(1.5)†	269(1.2)	62(1.4)†	299(1.4)
	1992	45(0.9)†	220(1.4)	57(1.6)†	270(1.4)	61(1.4)†	299(1.4)
	1990	42(1.3)†	218(2.0)†	50(1.5)*†	267(1.0)	58(1.3)†	300(1.1)
	1988	45(1.4)†	220(1.7)	52(1.5)†	265(1.4)†	58(1.6)†	300(1.3)
	1984	37(1.0)*†	223(0.9)	46(1.1)*†	268(0.7)	50(1.2)*†	301(0.7)
	1980	40(1.5)*†	226(1.1)*	49(1.3)*†	271(0.8)	51(1.3)*†	299(1.0)
	1975	34(0.7)*	222(0.9)	40(0.9)*	270(0.8)	46(0.8)*†	301(0.7)
	1971	33(0.9)*	224(1.1)	38(1.1)*	270(0.8)	42(1.3)*	302(1.0)
Graduated High School	1994	16(0.8)†	207(2.6)	27(1.2)†	251(1.4)	27(1.1)†	276(1.9)†
	1992	16(0.8)†	207(1.5)	28(1.2)	252(1.7)	28(0.9)	281(1.6)
	1990	17(0.8)†	209(1.8)	31(1.2)	251(0.9)†	30(1.0)	283(1.4)*
	1988	16(0.6)†	211(2.2)	31(1.1)	253(1.2)	30(1.2)	282(1.3)
	1984	20(0.6)*†	209(1.0)	36(1.0)*†	253(0.7)	35(1.0)*†	281(0.7)
	1980	25(0.8)*†	213(1.3)†	31(0.7)	254(0.9)	32(0.9)*	278(1.0)†
	1975	24(0.4)*	211(0.9)	33(0.6)*	255(0.7)	34(.05)*	281(1.1)
	1971	22(0.5)*	208(1.2)	32(0.7)*	256(0.8)	31(0.8)*	283(1.2)*
Less than High School	1994	4(0.4)†	189(4.0)	7(0.6)†	237(2.4)	7(0.5)†	268(2.7)
	1992	5(0.4)†	195(4.5)	6(0.5)†	239(2.6)	8(0.8)†	271(3.9)
	1990	5(0.5)†	193(3.2)	8(0.6)†	241(1.8)	9(0.6)†	270(2.8)
	1988	5(0.6)†	193(4.9)	8(0.6)†	247(2.1)*†	9(0.8)†	267(2.0)
	1984	6(0.2)*†	195(1.4)†	9(0.4)†	240(0.9)	12(0.6)*†	269(1.1)†
	1980	7(0.5)*†	194(1.6)	10(0.6)*†	239(1.1)	13(0.7)*†	262(1.5)
	1975	10(0.4)*	190(1.3)	14(0.6)*†	239(1.2)	16(0.6)*†	263(1.3)
	1971	10(0.4)*	189(1.5)	16(0.6)*	238(1.3)	20(0.8)*	261(1.5)
I Don't Know	1994	34(0.8)	202(1.3)†	9(0.7)†	230(3.0)	3(0.4)†	243(5.2)
	1992	34(1.3)	204(1.2)†	9(0.5)†	236(2.6)	3(0.3)†	255(5.9)
	1990	36(1.1)	201(1.5)	11(0.6)†	238(1.9)	3(0.3)†	246(5.7)
	1988	34(1.3)	204(1.5)†	9(0.7)†	240(3.0)	2(0.3)†	255(6.2)
	1984	38(0.9)*	204(0.7)†	10(0.4)†	237(1.3)	3(0.2)†	257(2.0)
	1980	28(1.0)*†	206(1.0)†	10(0.7)†	233(1.7)	4(0.4)†	250(3.5)
	1975	32(0.8)	203(0.8)†	13(0.6)*	235(1.1)	4(0.2)†	240(2.8)†
	1971	35(0.7)	197(1.0)*	14(0.8)*	233(1.0)	7(0.8)*	261(5.0)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Among 17-year-olds, students at the highest level of parental education displayed a pattern of slightly decreased performance across the assessment years, although no significant difference was observed between 1994 and 1971 average scores. The average score of 17-year-olds who reported high school graduation as their parents' highest level of education decreased between 1971 and 1980. The 1994 average score for this group remained below that observed in 1971. The overall pattern for 17-year-olds at the lowest level of parental education was one of increased performance. However, the 1994 average score for these students was not significantly different from the average score of their counterparts in 1971.

Trends in Reading Scale Scores from 1980 to 1994 by Type of School

Students' average reading scale scores by type of school attended are shown in Table 7.3. Trend information concerning school type was first reported in 1980. Examination of data collected from 1980 through 1994 indicates that the relative percentages of students attending nonpublic versus public schools have remained relatively stable since 1980. In 1994 students at all three ages attending nonpublic schools demonstrated higher average reading scores than did students attending public schools.

Among public school students, the overall pattern among 9-year-olds was one of slightly decreased performance. However, there was no significant difference between 1994 and 1980 average scores. At age 13, public school students had relatively stable performance across the trend assessments, with no significant changes observed in any year. The average score of 17-year-old public school students increased from 1980 to 1988, but then declined and remained at a level in 1994 that did not differ significantly from that in 1980.

Among nonpublic school students, no significant changes between 1980 and 1994 were observed in the average scores of 9- and 13-year-olds. For 17-year-old students attending nonpublic schools, the overall pattern observed was one of slightly increased performance. However, the 1994 average score for these students was not significantly different from the 1980 average.

Table 7.3

Trends in Average Reading Scale Scores by Type of School, 1980 to 1994

Type of School	Year	AGE 9 Percent of Students	AGE 9 Average Scale Score	AGE 13 Percent of Students	AGE 13 Average Scale Score	AGE 17 Percent of Students	AGE 17 Average Scale Score
Public	1994	89(2.1)	209(1.4)	89(1.5)	256(1.0)	89(2.3)	286(1.5)
	1992	88(1.7)	209(1.0)†	86(1.9)	257(1.3)	92(1.9)	288(1.0)
	1990	92(1.9)	208(1.4)†	88(1.9)	255(0.8)	93(1.5)	289(1.1)
	1988	88(2.7)	210(1.2)	89(2.5)	256(1.0)	88(3.5)	289(1.0)†
	1984	87(1.7)	209(0.8)†	88(1.1)	255(0.6)	89(1.7)	287(0.6)
	1980	89(1.4)	214(1.1)	88(1.3)	257(1.1)	93(1.2)	284(1.2)
Nonpublic	1994	11(2.1)	225(2.7)	11(1.5)	276(3.4)	11(2.3)	306(5.8)
	1992	12(1.7)	225(2.3)	14(1.9)	276(2.6)	8(1.9)	310(4.2)
	1990	8(1.9)	228(3.3)	12(1.9)	270(2.9)	7(1.5)	311(4.2)
	1988	12(2.7)	223(3.0)	11(2.5)	268(2.8)	12(3.5)	300(3.8)
	1984	13(1.7)	223(1.6)	12(1.1)	271(1.7)	11(1.7)	303(2.0)
	1980	11(1.4)	227(1.8)	12(1.3)	271(1.5)	7(1.2)	298(2.7)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1980, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Reading Scale Scores from 1971 to 1994 by Modal Grade

The percentages of students below, at, and above modal grade, and their average reading scale scores, are presented in Table 7.4. The modal grade is the grade typically attended by students of a particular age. The modal grades are: grade 4 for age 9, grade 8 for age 13, and grade 11 for age 17. Students who are below modal grade are older than the expected age of students in their grade. Conversely, students who are above modal grade are younger than the expected age of students in their grade.

For all three age groups, the percentage of students who were below modal grade increased from 1971 to 1994. Correspondingly, there was a decrease in the percentage of students at modal grade. As discussed in Chapter 1 of this report, increased school entrance ages and retention of children in early grades to allow for additional maturation and development may be reflected in these shifts.

At all three ages, students who were below modal grade made gains through the 1970s and 1980s. Although performance has been somewhat stable during the early 1990s, the average scores in 1994 for these students were higher than the average scores of their counterparts in 1971. Nine-year-olds who were at modal grade demonstrated improved performance between 1971 and 1980. Since that time, the average score has remained relatively consistent, and in 1994 it continued to be higher than in 1971. The average scores of 13- and 17-year-olds who were at modal grade have increased across the assessment years and were higher in 1994 than they were in 1971.

Table 7.4

Trends in Average Reading Scale Scores by Modal Grade, 1971 to 1994

		AGE 9		AGE 13		AGE 17	
Modal Grade	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Below Modal Grade	1994	40(0.5)†	194(1.6)†	44(0.5)†	244(1.6)†	29(0.8)†	261(2.4)†
	1992	43(0.5)*†	192(1.4)†	43(0.7)†	243(1.9)†	28(0.6)†	261(1.5)†
	1990	42(0.5)*†	189(1.8)†	39(0.4)*†	243(1.1)†	26(0.6)*†	261(1.9)†
	1988	37(0.3)*†	193(1.8)†	39(0.5)*†	243(1.4)†	24(1.2)*†	265(1.9)†
	1984	34(0.3)*†	187(0.9)*†	37(0.2)*†	239(0.7)†	22(0.7)*†	259(0.9)†
	1980	28(1.5)*	189(1.3)†	28(1.2)*	240(1.5)†	14(0.7)*	244(2.1)*
	1975	23(0.8)*	183(1.1)*†	28(0.9)*	232(0.9)*	15(0.7)*	242(1.8)*
	1971	24(0.8)*	178(1.2)*	28(0.9)*	230(1.0)*	14(0.6)*	238(1.5)*
At Modal Grade	1994	60(0.5)†	222(1.3)†	56(0.5)†	269(0.8)†	63(0.3)†	299(1.2)†
	1992	57(0.5)*†	224(1.0)†	56(0.5)†	272(1.1)†	64(0.2)†	301(1.3)†
	1990	58(0.5)*†	224(1.5)†	60(0.2)*†	266(0.9)	65(0.2)*†	299(1.0)†
	1988	63(0.3)*†	223(1.5)†	61(0.3)*†	267(1.1)	65(0.2)*†	297(1.1)†
	1984	65(0.2)*†	223(0.8)†	62(0.2)*†	267(0.5)	68(0.2)*†	296(0.6)†
	1980	71(1.4)*	225(0.8)†	70(1.3)*	266(0.8)	77(0.6)*†	291(1.0)*
	1975	75(0.9)*	218(0.7)	72(0.9)*	265(0.7)*	73(0.7)*	292(0.7)*
	1971	75(0.8)*	217(1.1)*	71(0.9)*	265(0.8)*	73(0.7)*	291(1.0)*
Above Modal Grade	1994	0(0.1)†	***(***)	0(0.0)†	***(***)	7(0.8)†	305(5.3)
	1992	0(0.1)	***(***)	1(0.6)	312(3.9)*†	8(0.6)†	300(3.2)
	1990	0(0.1)	***(***)	1(0.3)	290(16.0)*	9(0.6)†	310(2.3)
	1988	1(0.2)	262(11.0)	1(0.5)	272(10.8)*	12(1.3)*	305(3.0)
	1984	0(0.1)	***(***)	1(0.2)*†	295(7.5)*	10(0.7)†	304(1.2)
	1980	0(0.1)	***(***)	1(0.1)*†	274(4.9)*	9(0.6)†	301(1.7)
	1975	1(0.1)†	226(4.3)*	1(0.1)*	278(4.2)*	12(0.4)*	302(1.0)
	1971	1(0.1)*	232(4.1)	1(0.2)*	278(2.4)	13(0.7)*	303(1.6)

NOTE: The modal grades are: grade 4 at age 9, grade 8 at age 13, and grade 11 at age 17.

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The overall picture of trends in reading achievement is one of only minimal changes across the assessment years. At age 9, higher levels of performance that were observed in the 1980s compared to 1971 have not been maintained, and in 1994 the average score returned to a level not significantly different from that in 1971. The average score of 13-year-olds increased in 1992 to a level that was higher than in 1971. However, in 1994 scores returned to a level that did not differ significantly from that in 1971. Seventeen-year-olds had higher average scores from 1988 to 1992 compared to the average in 1971. However, the 1994 average score returned to a level not significantly different from that in 1971.

Despite a decrease since 1990 in the performance of 9-year-olds in the upper quartile, gains were observed across the assessment years, and the 1994 average was higher than the 1971 average. In the middle two and lower quartiles, 9-year-olds displayed an increase in performance from 1970 to 1980. Although their average reading scores subsequently declined, they remained higher in 1994 than in 1971. The average scores of 13-year-olds in the upper and middle two quartiles increased across the assessment years. For 13-year-olds in the lower quartile, increases in the 1980s were followed by a decline in performance that resulted in a 1994 average score that was not different than the 1971 average. Seventeen-year-olds in the upper quartile made gains across the assessment years and attained a higher average score in 1994 than in 1971. Although the pattern of performance for 17-year-olds in the middle two quartiles was one of slight improvement, there was no significant difference between 1971 and 1994 average scores. In the lower quartile, the average score of 17-year-olds increased between 1971 and 1994.

For White students in all three age groups, the overall pattern was one of increased performance in reading. However, only 13-year-olds attained an average score that was higher in 1994 than in 1971. Black students in each age group demonstrated increases in average scores during the 1970s and 1980s. Although declines have been observed during the early 1990s, the average scores of these students continued to be higher than those in 1971. Nine-year-old Hispanic students had average scores that fluctuated, but did not change significantly across the assessments. At age 13, there was evidence that a pattern of slight increases observed for Hispanic students in the 1980s was reversed in the early 1990s. Similarly, the gains made by 17-year-old Hispanic students during the 1980s were not maintained. For both 13- and 17-year-old Hispanic students, the 1994 and 1971 average scores did not differ significantly.

In 1994, White students in each group outperformed their Black and Hispanic peers on the reading assessment. The performance gap between White and Black students declined across the assessment years for ages 9 and 13. Despite some evidence that it has begun to increase in recent assessments, the gap in 1994 was smaller than the gap in 1971. No change in the magnitude of the difference between White and Hispanic students was observed at ages 9 and 13. Although the gap narrowed somewhat among 17-year-olds in the 1980s, there was no significant difference between 1994 and 1975.

Despite some fluctuation, the average reading score of 9-year-old male students increased between 1971 and 1994. The gains demonstrated by 13- and 17-year-old male students in the 1980s subsequently reversed. In 1994, their average scores were not significantly different from the average in 1971. Among female students, the gains made by 9-year-olds between 1971 and 1980 have not been maintained since that time, resulting in a 1994 average score that did not differ significantly from the 1971 average score. The pattern of performance for 13- and 17-year-old female students was one of overall improvement. However, only 13-year-old females attained an average score in 1994 that was higher than in 1971.

In 1994, females outperformed males on the reading assessment. At age 9, the gap between males and females decreased slightly across the assessment years, but did not change to a statistically significantly degree between 1971 and 1994. At ages 13 and 17, a small decrease in the gap between males and females that was observed in the early 1980s has not been maintained, so that the gaps in 1971 and 1994 did not differ significantly.

In the Northeast, little or no change was observed since 1971 in the reading performance of 9-year-olds. Thirteen-year-olds displayed increased performance in the early 1990s, resulting in a higher score in 1994 than in 1971. Although gains have been made by 17-year-olds during the 1980s and early 1990s, there was no significant difference between 1994 and 1971 average scores. In the Southeast, students in each age group showed some improvement in performance during the 1970s and 1980s. Although there have been declines since 1988, the average score for 9- and 17-year-olds in 1994 was higher than in 1971. For 13-year-olds, the average score returned to a level in 1994 that did not differ significantly from that in 1971. In the Central region, despite some fluctuations, there were no significant changes in the average scores of students in any age group. In the West region, improvement that was observed from 1971 to 1980 for 9-year-olds subsequently reversed, resulting in a 1994 average score that did not differ significantly from that in 1971. Little or no change was observed in the performance of 13-year-olds. At age 17, a pattern of overall improvement was observed, although the 1994 average score did not differ significantly from that in 1971.

Nine-year-olds who reported the highest level of education (post-high school) displayed decreased performance across the assessment years, although 1994 and 1971 average reading scores were not different to a statistically significant degree. The average score of 9-year-olds who reported high school graduation as their parents' highest level of education increased between 1971 and 1980, but subsequently decreased to a level in 1994 that did not differ significantly from 1971. Little or no change in performance was observed among 9-year-olds who reported the lowest level of education (less than high school). The average scores of 13-year-olds who reported the highest and the middle levels of parental education declined across the assessment years, but were not significantly different in 1994 than in 1971. At the lowest level of parental education, 13-year-olds had an average score in 1988 that was higher than in 1971. However, subsequent declines resulted in no significant difference between 1994 and 1971 averages. The overall pattern for 17-year-olds at the highest level of education was one of decreased performance, although 1994 and 1971 average scores did not differ significantly. At the middle level of parental education, 17-year-olds' average score fluctuated across assessments, but was lower in 1994 than in 1971. Seventeen-year-olds at the lowest level of parental education showed an overall pattern of improved performance, although no significant difference between 1994 and 1971 averages was observed.

For 9-year-olds in public schools, an overall decline in performance was observed, although the 1994 average reading score did not differ significantly from the 1980 average. No significant change was observed for 13-year-old public school students. At age 17, an increase observed in 1988 was not maintained, and the 1994 average score did not differ significantly from the 1971 average. The performance of 9- and 13-year-old students attending nonpublic schools has remained relatively stable across the assessments. At age 17, there was some indication that performance has improved for nonpublic schools students, although the 1994 and 1971 average scores did not differ significantly.

At all three ages, the average reading scores of students who were below or at modal grade increased between 1971 and 1994.

8

National Trends in Levels of Reading Performance from 1971 to 1994

This chapter expands on the discussion of students' reading achievement presented in Chapter 7 by focusing on trends in levels of reading performance across the eight reading assessments. To provide more specific information about the types of reading abilities displayed by students, five levels of performance have been identified and described along the NAEP scale — 150, 200, 250, 300, and 350. The procedure for describing, or "anchoring," performance at the five levels on the scale involved an empirical process that delineated sets of questions and passages more likely to be answered successfully by students whose overall performance was at a particular anchor level and much less likely to be answered successfully by students performing at the next lower level.[42] In this way, a detailed picture of the reading skills displayed by students at each of the five levels could be developed.

[42] In theory, performance levels above 350 or below 150 could have been defined; however, so few students in the assessment performed at the extreme ends of the scale that it was not practical to do so.

A panel of reading experts analyzed the sets of questions and passages that "anchored" at each of the five scale levels, carefully considering the various reading processes and skills required to answer the questions and the degree of understanding demonstrated by students' responses. In conducting their analyses, the panelists considered how the interaction between characteristics of the text, the students' background knowledge, and the type of question asked affected students' ability to gain meaning from the text and to demonstrate their understanding.

A general finding from the analyses was that most students in each age group could read and understand uncomplicated, short passages consisting of a few simple sentences. As passages became more complex, with less familiar narrative or expository structures, students experienced more difficulty, particularly at age 9. Also, the familiarity of the topic played an important role in students' ability to understand, with general, "everyday" topics being easier than more specialized subject matter.

The nature of the tasks in the assessment ranged from identifying facts in a passage and inferring meaning, to connecting, interpreting, and extending ideas across the text. It was clear that the complexity of the passages, as well as the nature of the tasks, had an impact on students' success with the questions. In some instances, students could make generalizations about ideas within short, simple passages, while in other cases, students had difficulty answering questions about facts that were embedded in rather dense texts. Furthermore, it was apparent that constructing their own response was more difficult for students than selecting a response from among options in a multiple-choice question. This was particularly true when the constructed-response question asked students to provide an interpretation of information in the text.

These findings are consistent with other research suggesting that the nature of the text and students' competencies with various reading processes influence their ability to understand certain texts.[43] A wide range of reading abilities and interactions with text are portrayed across the five levels of performance on the NAEP scale. The descriptions in Figure 8.1 characterize the reading abilities of most students at each of the five levels.

---

[43] Dole, J. A.; Duffy, G. C.; Roehler, L. R.; & Pearson, D. P., "Moving From the Old to the New: Research on Reading Comprehension Instruction," *Review of Educational Research*, 61(2), 239-264 (1991).

McKeown, M. G.; Beck, I. L.; Sinatra, G. M.; & Loxterman, J. A., "The Contribution of Prior Knowledge and Coherent Text to Comprehension," *Reading Research Quarterly*, 27, 78-93 (1992).

Figure 8.1 — Levels of Reading Performance

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: 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: 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.

Table 8.1 presents the percentages of students who performed at or above each reading performance level in the eight reading assessments conducted by NAEP since 1971. It is expected that older students will have more success with the increasingly difficult reading tasks reflected in the higher performance level descriptions. In fact, this was the case as students showed a clear pattern of increased reading abilities from ages 9 to 17.

Table 8.1

Trends in Percentages of Students At or Above Five Reading Performance Levels, 1971 to 1994

		YEARS							
Skills and Strategies	Age	1971	1975	1980	1984	1988	1990	1992	1994
LEVEL 350									
Learn from	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.1)	0(0.0)	0(0.0)
Specialized	13	0(0.0)	0(0.0)	0(0.0)	0(0.1)	0(0.1)	0(0.1)	1(0.3)	1(0.1)
Texts	17	7(0.4)	6(0.3)	5(0.4)†	6(0.3)	5(0.6)*†	7(0.5)	7(0.6)	7(0.7)
LEVEL 300									
Understand	9	1(0.1)	1(0.1)	1(0.1)	1(0.1)	1(0.3)	2(0.3)	1(0.2)	1(0.3)
Complicated	13	10(0.5)*	10(0.5)*	11(0.5)*	11(0.4)*	11(0.8)*	11(0.6)*	15(0.9)†	14(0.8)†
Information	17	39(1.0)	39(0.8)	38(1.1)	40(0.8)	41(1.5)	41(1.0)	43(1.1)†	41(1.2)
LEVEL 250									
Interrelate Ideas	9	16(0.6)	15(0.6)	18(0.8)	17(0.6)	18(1.1)	18(1.0)	16(0.8)	17(1.2)
and Make	13	58(1.1)	59(1.0)	61(1.1)	59(0.6)	59(1.3)	59(1.0)	62(1.4)	60(1.2)
Generalizations	17	79(0.9)	80(0.7)	81(0.9)	83(0.5)†	86(0.8)*†	84(1.0)†	83(0.8)†	81(1.0)
LEVEL 200									
Partial Skills and	9	59(1.0)	62(0.8)†	68(1.0)†	62(0.7)	63(1.3)	59(1.3)	62(1.1)	63(1.4)
Understanding	13	93(0.5)	93(0.4)	95(0.4)*†	94(0.3)*	95(0.6)*	94(0.6)	93(0.7)	92(0.6)
	17	96(0.3)	96(0.3)	97(0.3)†	98(0.1)*†	99(0.3)*†	98(0.3)†	97(0.4)	97(0.5)
LEVEL 150									
Simple,	9	91(0.5)	93(0.4)†	95(0.4)*†	92(0.3)†	93(0.7)	90(0.9)	92(0.4)	92(0.7)
Discrete	13	100(0.0)	100(0.1)	100(0.1)*	100(0.0)	100(0.1)	100(0.1)	100(0.3)	99(0.2)
Reading Tasks	17	100(0.1)	100(0.1)	100(0.1)†	100(0.0)†	100(0.0)†	100(0.1)	100(0.1)	100(0.1)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1971, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Level 350: The percentage of students demonstrating the more advanced reading abilities outlined at Level 350, such as learning from specialized reading materials, continued to be quite small in 1994. For 17-year-olds, a decline in the percentage of students at this level during the 1980s has reversed, so that in 1994 the percentage was not significantly different from that in 1971.

Level 300: The percentage of 13-year-olds performing at or above Level 300 (understanding of complicated information) increased across the assessments, and was higher in 1994 than in 1971. An overall increase was observed at age 17 as well, although the 1994 percentage did not differ significantly from the 1971 percentage.

Level 250: Interrelating ideas and making generalizations were characteristic of performance at Level 250. The percentage of 9- and 13-year-olds reaching Level 250 or above has changed little since 1971. At age 17, an increase in the percentage of students at or above this level was observed from 1971 to 1988. Since that time, the percentage decreased and was not significantly different in 1994 than in 1971.

Level 200: As in past assessments, nearly all of the 17-year-old students and the overwhelming majority of 13-year-old students in 1994 performed at or above Level 200, demonstrating at least partially developed skills and understanding. However, less than two-thirds (63 percent) of 9-year-olds demonstrated performance at or above this level. For 9- and 13-year-olds, the percentages of students at or above Level 200 increased during the late 1970s and early 1980s, but subsequently decreased. For all ages in 1994, the percentages of students attaining this level of performance or above were not significantly different than the percentages in 1971.

Level 150: Across the assessments, nearly all 13- and 17-year-old students and the overwhelming majority of 9-year-olds were successful with the simple, discrete reading tasks representative of this level.

These results reveal some indications of increased performance, particularly at Level 300, for both ages 13 and 17. For the most part, however, the increases that were observed were relatively small and left room for growth at all age levels. Notably, very small percentages of 17-year-olds approaching the end of their high school education have reached the upper level of reading performance defined by NAEP.

Trends in Levels of Reading Performance between 1971 and 1994 by Race/Ethnicity

Table 8.2 shows the percentages of 9-, 13-, and 17-year-old White, Black, and Hispanic students performing at or above each of the reading performance levels in the first assessment year and again in 1994.[44] Initial results for White and Black students are from 1971. The first assessment results for Hispanic students are from 1975.

The percentages of 9-year-old White and Black students who demonstrated mastery at or above the lower two performance levels (Levels 150 and 200) in 1994 were higher than in 1971. Also, a higher percentage of Hispanic 9-year-olds attained at least Level 250 in 1994 than did in 1975. In 1994, a greater percentage of White 9-year-olds than Black or Hispanic 9-year-olds performed at or above Levels 150, 200, and 250.

Among 13-year-olds, increases were demonstrated between 1971 and 1994 in the percentage of Black students attaining at least Levels 200, 250, and 300. A higher percentage of White students performed at or above Levels 250 and 300 in 1994 than in 1971. In 1994, the percentages of White students reaching Levels 200, 250, and 300 or above were higher than the percentages of Black or Hispanic students doing so.

The percentages of Black 17-year-old students performing at or above Levels 200, 250, and 300 were higher in 1994 than in 1971. White 17-year-olds also demonstrated an increase at Level 300. Comparing the 1994 performance of students in the three racial/ethnic groups, a greater percentage of White students than Black or Hispanic students performed at or above Levels 250, 300, and 350. In addition, a greater percentage of White than Black 17-year-olds performed at or above Level 200.

[44] Trends in percentages of students performing at or above each of the five levels in all assessments by race/ethnicity and gender are presented in the Data Appendix.

Table 8.2

Trends in Percentages of Students At or Above Five Reading Performance Levels by Race/Ethnicity, 1971/1975 and 1994

Levels	Age	1971 White	1971 Black	1971 (1975) Hispanic	1994 White	1994 Black	1994 Hispanic
LEVEL 350							
Learn from Specialized	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Reading Materials	13	0(0.1)*	0(0.0)	0(0.0)	1(0.2)	0(0.1)	0(0.2)
	17	8(0.4)	0(0.1)	1(0.6)	9(0.9)	2(1.2)	2(1.1)
LEVEL 300							
Understand Complicated	9	1(0.2)	0(0.0)	0(0.0)	1(0.3)	0(0.2)	0(0.0)
Information	13	11(0.5)*	1(0.2)*	2(1.0)	17(1.0)	4(1.2)	4(1.8)
	17	43(0.9)*	8(0.9)*	13(2.7)	48(1.4)	22(3.7)	20(3.0)
LEVEL 250							
Interrelate Ideas and	9	18(0.7)	2(0.5)	3(0.5)*	20(1.5)	4(1.5)	6(1.6)
Make Generalizations	13	64(0.9)*	21(1.2)*	32(3.6)	68(1.3)	36(3.5)	34(3.9)
	17	84(0.7)	40(1.6)*	53(4.1)	86(1.1)	66(4.1)	63(4.4)
LEVEL 200							
Partial Skills and	9	65(1.0)*	22(1.5)*	35(3.0)	70(1.5)	38(2.8)	37(4.6)
Understanding	13	96(0.3)	74(1.7)*	81(2.3)	95(0.7)	81(2.3)	82(2.7)
	17	98(0.2)	82(1.5)*	89(2.4)	98(0.4)	93(2.0)	91(3.4)
LEVEL 150							
Simple, Discrete	9	94(0.4)*	70(1.7)*	81(2.5)	96(0.5)	79(2.4)	80(4.6)
Reading Tasks	13	100(0.0)*	99(0.3)	100(0.3)	100(0.2)	99(1.0)	99(0.9)
	17	100(0.0)	98(0.4)*	99(0.4)	100(0.1)	100(0.6)	99(1.3)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). The first assessment results for Hispanic 9-, 13-, and 17-year-old students are from 1975.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Levels of Reading Performance between 1971 and 1994 by Gender

Table 8.3 presents the percentages of students at or above each reading performance level by gender. Across the assessments, females have continued to outperform males at all levels of performance. However, in two instances gender differences were not statistically significant: at Level 250 for 9-year-olds, and at Level 200 for 17-year-olds. As shown in Table 8.3, the percentages of male or female students attaining specific levels show some increases between 1971 and 1994. In comparison to percentages observed in 1971, an increased percentage of males at age 9 achieved Levels 200 and 250 or above in 1994. At age 13, higher percentages of both males and females attained at least Level 300 in 1994 than in 1971. No changes were observed between 1971 and 1994 in the percentage of 17-year-old male or female students attaining any of the performance levels.

Table 8.3

Trends in Percentages of Students At or Above Five Reading Performance Levels by Gender, 1971 and 1994

Skills and Strategies	Age	1971 Male	1971 Female	1994 Male	1994 Female
LEVEL 350					
Learn from Specialized Reading Materials	9	0(0.0)	0(0.0)	0(0.0)	0(0.0)
	13	0(0.0)	0(0.1)*	0(0.2)	1(0.2)
	17	5(0.4)	8(0.5)	5(0.9)	9(0.9)
LEVEL 300					
Understand Complicated Information	9	1(0.2)	1(0.2)	1(0.4)	1(0.3)
	13	7(0.5)*	12(0.6)*	10(0.7)	18(1.1)
	17	34(1.1)	44(1.2)	36(1.9)	47(1.5)
LEVEL 250					
Interrelate Ideas and Make Generalizations	9	12(0.6)*	19(0.8)	15(1.2)	18(1.5)
	13	52(1.2)	64(1.1)	53(1.9)	68(1.7)
	17	74(1.0)	83(1.0)	76(1.5)	86(1.2)
LEVEL 200					
Partial Skills and Understanding	9	53(1.2)*	65(1.1)	59(1.5)	67(1.9)
	13	91(0.7)	95(0.4)	89(1.1)	95(0.6)
	17	95(0.4)	97(0.3)	96(0.9)	98(0.5)
LEVEL 150					
Simple, Discrete Reading Tasks	9	88(0.7)	93(0.5)	90(1.0)	94(0.8)
	13	100(0.1)	100(0.1)	99(0.3)	100(0.2)
	17	99(0.1)	100(0.1)	100(0.2)	100(0.1)

* Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Reading Performance between 1984 and 1994 on Constructed-Response Questions

To determine students' ability to construct their own thoughtful responses to text, a subset of the questions included in the NAEP reading trend assessment asked students to analyze, interpret, and evaluate what they had read. These tasks not only required students to demonstrate understanding, but also to express their ideas and reflections in a written response that could be understood by others. The tasks are described in Figure 8.2.

Figure 8.2

Constructed-Response Assessment Tasks, 1984 and 1994

Task 1: (Science passage) Students were asked to discuss the main idea.

Task 2: (Biographical piece) Students were asked to interpret the writer's view.

Task 3: (Informative piece) Students were asked to make comparisons between what they had read and their own experiences.

Task 4: (Informative piece) Students were asked to interpret the way the writer conveyed a particular impression.

Task 5: (Historical piece) Students were asked to compare and contrast information contained in the article.

Task 6: (Humorous piece) Students were asked to evaluate the piece.

Task 7: (Literary piece) Students were asked to discuss the mood.

For each task, students' responses were evaluated according to their overall success in responding to the question and the extent to which the ideas were substantiated with evidence from the passage. Responses were rated as being either unsatisfactory, minimal, satisfactory, or elaborated. Responses rated as unsatisfactory did not address the task or provided irrelevant or inappropriate comments or information. Minimal responses indicated a partial understanding of the task and generally reflected an incomplete comprehension of the passage. Satisfactory responses included enough detail to indicate that students had comprehended the passage successfully, provided an appropriate response to the task, and offered some support for the ideas discussed. Elaborated responses went beyond a basic understanding of the passage by restructuring or extending ideas in the passage and providing relevant support. Such responses identified relationships among ideas, even when relationships were not stated explicitly.

Table 8.4 displays the percentages of students at each score level for constructed-response tasks in the reading trend assessment. At age 9, most students provided unsatisfactory or minimal responses to the four tasks. Comparison of 1994 data with those collected in 1984 indicates only one change in 9-year-olds' performance on these constructed-response questions. In 1994, more students provided unsatisfactory responses when they were asked to discuss the main idea of a science passage in Task 1, with a resulting drop in the number of minimal responses.

Few changes from 1984 to 1994 were observed for 13-year-olds, except for their performance on Tasks 3 and 5. Task 3 required students to compare their own experiences with the information contained in an article. In 1994, fewer 13-year-old students' responses were scored as minimal and more responses were scored as satisfactory than in 1984. For Task 5, which required students to compare and contrast information contained in a historical article, the percentage of 13-year-old students providing satisfactory responses was higher in 1994 than in 1984. The increased percentage of students attaining the satisfactory score level on this task was accompanied by a decrease, though not a statistically significant one, in the percentage of students providing minimal and unsatisfactory responses.

In comparison to those in 1984, 17-year-old students in 1994 showed indications of improvement on some tasks and indications of declines or no changes in performance on others. In 1994, a smaller percentage of students' responses were scored as satisfactory and a greater percentage of responses were scored lower, at the minimal level, for Task 1. This task required students to read and discuss the main idea contained in a science passage.

On Task 2, which required students to provide an interpretation of a writer's view, a greater percentage of 17-year-olds' responses were scored satisfactory in 1994 compared to 1984, while a smaller percentage were scored at the minimal level. On Task 3, an informative reading task in which students were asked to compare text information with prior experiences, fewer 17-year-olds' responses were scored as minimal in 1994 than in 1984. The resulting shift of scores for this task seemed to have been spread across the other three score levels. For Task 5, a compare-and-contrast task, the percentage of responses scored at the satisfactory level was greater in 1994 than in 1984; the percentage scored as minimal was smaller. Student performance in 1994 on Task 6, which required students to provide an evaluation of a passage, was marked by an increase in the percentage of responses scored at the highest level (elaborated) relative to 1984.

Table 8.4

Trends in Students' Responses to Constructed-Response Questions, 1984 and 1994

		1984			1994		
Tasks	Rating	Age 9	Age 13	Age 17	Age 9	Age 13	Age 17
Task 1	Unsatisfactory	69(1.8)*	47(1.8)	34(1.1)	78(2.1)	45(2.0)	35(2.0)
	Minimal	29(1.6)*	41(1.5)	43(1.4)*	20(2.2)	43(1.7)	50(2.1)
	Satisfactory	2(0.5)	11(1.0)	21(0.8)*	2(0.8)	11(1.5)	12(1.2)
	Elaborated	0(0.0)	1(0.3)	2(0.3)	0(0.0)	1(0.5)	2(0.6)
Task 2	Unsatisfactory	28(1.6)	–	10(0.7)	34(3.1)	–	14(1.7)
	Minimal	57(1.6)	–	63(1.2)*	50(3.0)	–	53(2.5)
	Satisfactory	14(1.0)	–	24(0.9)*	15(2.3)	–	29(2.4)
	Elaborated	0(0.1)	–	3(0.5)	1(0.5)	–	4(1.0)
Task 3	Unsatisfactory	56(1.8)	21(1.2)	17(1.0)	57(3.2)	24(1.6)	21(1.8)
	Minimal	40(2.0)	63(1.3)*	59(1.1)*	38(3.2)	53(2.0)	52(2.6)
	Satisfactory	4(0.7)	15(0.8)*	22(1.1)	5(1.1)	20(1.7)	23(2.0)
	Elaborated	0(0.1)	2(0.4)	2(0.3)	0(0.1)	4(0.9)	5(1.2)
Task 4	Unsatisfactory	–	53(1.4)	45(1.5)	–	52(2.5)	49(2.5)
	Minimal	–	40(1.3)	47(1.6)	–	39(2.3)	44(2.5)
	Satisfactory	–	7(0.8)	8(0.8)	–	9(1.2)	6(1.1)
	Elaborated	–	1(0.2)	0(0.2)	–	1(0.4)	1(0.4)
Task 5	Unsatisfactory	–	24(0.9)	15(0.8)	–	21(2.1)	12(1.9)
	Minimal	–	41(1.2)	51(1.2)*	–	35(2.3)	38(2.4)
	Satisfactory	–	33(1.3)*	33(1.5)*	–	44(1.8)	49(2.2)
	Elaborated	–	2(0.3)	1(0.2)	–	1(0.5)	1(0.4)
Task 6	Unsatisfactory	13(1.2)	4(0.5)	2(0.5)	11(1.5)	2(0.7)	3(1.0)
	Minimal	83(1.2)	82(1.0)	76(1.2)	83(1.8)	79(1.7)	74(2.4)
	Satisfactory	4(0.5)	14(1.1)	19(1.0)	0(0.0)	18(1.6)	18(1.8)
	Elaborated	0(0.1)	1(0.2)	2(0.4)*	0(0.0)	1(0.3)	5(1.0)
Task 7	Unsatisfactory	–	–	3(0.4)	–	–	5(1.0)
	Minimal	–	–	18(0.9)	–	–	15(1.6)
	Satisfactory	–	–	18(0.6)	–	–	19(1.2)
	Elaborated	–	–	62(1.1)	–	–	61(2.3)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

Across the reading trend assessments, few changes were observed in the percentages of students at any of the performance levels. An overall pattern of increasing percentages was observed for 9- and 17-year-olds performing at or above Level 250. A similar pattern was also observed for 13- and 17-year-olds at or above Level 300. However, the only change between 1971 and 1994 at any of the performance levels was an increase in the percentage of 13-year-olds attaining at least Level 300.

A comparison of 1994 results by racial and ethnic groups with those obtained in 1971 or 1975 indicates that there have been almost no changes in the percentages of students at the top level of reading performance by racial or ethnic group. There have been some gains made at the intermediate and lower levels, especially by White and Black students. Increased percentages of White students performed at or above Levels 150 and 200 among 9-year-olds, and Levels 250 and 300 among 13-year-olds. Among Black students, gains were observed in the percentages of 9-year-olds performing at or above Levels 150 and 200, the percentages of 13-year-olds performing at or above Levels 200, 250, and 300, and the percentages of 17-year-olds performing at or above Levels 200, 250, and 300. The percentages of Hispanics achieving each of the performance levels remain unchanged between 1975 and 1994, with the exception of an increase in the percentage of 9-year-olds reaching Level 250. In 1994, Whites continued to outperform their Black and Hispanic counterparts at Levels 150, 200, and 250 among 9-year-olds, Levels 200, 250, and 300 among 13-year-olds, and Levels 200, 250, 300, and 350 among 17-year-olds (although the difference between White and Hispanic 17-year-olds at Level 200 was not statistically significant).

A contrast of results by gender in 1994 and 1971 shows few changes for either males or females in overall reading performance. Some improvement was observed for males at age 9 and for both males and females at age 13. Among 9-year-olds, a greater percentage of males performed at or above Levels 200 and 250 in 1994 as compared to 1971. The percentages of male and female 13-year-olds performing at or above Level 300 in 1994 were greater than the percentages in 1971. Female students continued to outperform male students at most performance levels in 1994.

Comparing the data collected on constructed-response questions in 1994 with those collected in 1984 shows one decline in reading performance for 9-year-old students, some improvements for 13-year-olds, and mixed results for 17-year-olds. In 1994, the majority of student responses at all ages for most of the tasks were rated as either unsatisfactory or minimal, indicating that many students had difficulty in providing thoughtful analysis and interpretations of what they read. Two notable exceptions were Tasks 5 and 7. With Task 5, nearly one-half of 13- and 17-year-olds provided responses that were rated at least satisfactory. On Task 7, 80 percent of 17-year-olds wrote responses rated as satisfactory or better, and nearly two-thirds provided elaborated responses.

9

Trends in School and Home Influences on Literacy Development

Introduction

This chapter examines trends in the ways in which students' home and school environments relate to literacy development. The opportunities provided for students to read a variety of materials and engage in positive literacy activities are considered by reading specialists, researchers, and classroom professionals to be important contributors to the development of reading abilities. Since 1984, and in some cases since 1971, NAEP has asked students to provide information about these opportunities and resources as well as about their attitudes and habits related to reading. This information is valuable in helping parents, educators, and policy makers understand how literacy develops and what aspects of a student's experience are most related to advanced abilities in reading.

Trends in Reading Across the Curriculum between 1984 and 1994

The amount of reading that students complete as a part of instruction has been identified as a key factor in promoting literacy abilities. With an increasing emphasis on reading across the curriculum, reading as a process and tool for learning has taken on expanded importance in students' education.[45] Because of the importance placed on reading in most content areas, NAEP assessments have asked students to report the total amount of pages they read as assigned schoolwork per day, including reading at both school and home.

Table 9.1 presents trends in students' reports on the number of pages they read per day in school and for homework and their average reading scale scores. Comparison of students' reports in 1994 with those in 1984 indicates that in 1994 students at age 9 and age 13 reported reading more pages per day, whereas the reports of students at age 17 have not changed. Among 9-year-olds a greater percentage of students reported reading more than 20 pages each day, while a smaller percentage reported reading only 5 or fewer pages. Results for 13-year-olds indicated a greater percentage of students reading 16 to 20 and more than 20 pages each day. Correspondingly, there was a decline in percentages of 13-year-olds who reported reading 6 to 10 pages daily.

[45] Simonsen, S. & Singer, H., "Improving Reading Instruction in the Content Areas." In S. J. Samuels, & A. E. Farstrup (Eds.). *What Research Has to Say About Reading Instruction, 2nd edition* (Newark, DE: International Reading Association, 1992).

Table 9.1

Trends in Pages Read Per Day in School and for Homework, 1984 and 1994

		AGE 9		AGE 13		AGE 17	
Number of Pages	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
More than 20	1994	17(1.0)	217(2.4)	14(0.8)	261(2.0)	23(1.5)	304(2.2)
	1984	13(0.4)*	215(1.4)	11(0.4)*	261(1.2)	20(1.0)	299(1.0)
16 - 20	1994	14(0.9)	209(2.9)	13(0.5)	263(2.3)	13(0.6)	298(2.8)
	1984	13(0.5)	215(1.2)	11(0.2)*	263(1.0)	14(0.4)	296(0.9)
11 - 15	1994	14(0.5)	217(2.3)	17(0.6)	266(1.8)	18(0.6)	288(1.9)
	1984	14(0.5)	220(1.2)	18(0.4)	264(0.9)	18(0.3)	294(0.8)*
6 - 10	1994	26(0.6)	214(1.7)	31(0.9)	261(1.4)	25(0.9)	284(2.1)
	1984	25(0.5)	215(1.0)	35(0.5)*	261(0.6)	26(0.6)	287(0.8)
5 or fewer	1994	28(1.4)	203(2.3)	26(0.9)	249(1.7)	21(1.2)	271(1.9)
	1984	35(1.0)*	208(0.8)	27(0.6)	250(0.7)	21(0.8)	273(0.8)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

As in past assessments, a relationship between reading scale scores and the number of pages read each day was observed. At all three age groups, students who reported reading five or fewer pages per day for school and homework had lower average scores than students who reported reading more pages. Although the difference between average scale scores of 9-year-olds reading 16 to 20 pages and of their peers reading five or fewer pages was not significant. These results support the assertions of many educators and researchers that reading across the curriculum is an important aspect of students' overall reading development.[46]

[46] Chall, J.; Jacobs, V. A.; & Balwin, L. E. *The Reading Crisis: Why Poor Children Fail* (Cambridge, MA: Harvard University Press, 1990).

Another factor that seems to influence students' literacy development is exposure to a wide variety of reading materials. Educators and parents have reached a consensus that varied reading experiences with different types of materials helps readers develop their skills. Since 1984, NAEP has asked students to identify which of several types of texts they read a few times a year or more. The types of texts included are: poems, plays, biographies, science books, and books about other places. Table 9.2 presents students' responses and the average reading scale scores associated with exposure to each type of text.

While some positive trends in exposure to reading materials were observed among 13- and 17-year-olds, this was not the case among 9-year-old students. At age 9, a smaller percentage of students reported reading poems and plays in 1994 as compared to 1984. The reports of 13- and 17-year-olds indicated an increase between 1984 and 1994 in the percentage of students reading poems and biographies. In addition, a greater percentage of 17-year-olds reported reading science books and plays.

Table 9.2

Trends in Reading Certain Types of Materials a Few Times a Year or More Frequently, 1984 and 1994

		AGE 9		AGE 13		AGE 17	
Types of Materials	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Poems	1994	62(2.3)	210(2.9)	79(1.4)	261(2.3)	85(2.2)	293(2.1)
	1984	70(1.5)*	211(1.9)	68(1.3)*	260(1.2)	76(1.1)*	290(1.5)
Plays	1994	45(2.2)	207(3.0)	63(2.3)	263(2.1)	70(2.1)	294(2.4)
	1984	56(1.4)*	211(2.5)	59(1.4)	260(1.3)	63(1.0)*	290(1.7)
Biographies	1994	47(2.1)	210(3.4)	68(1.7)	261(2.1)	69(1.8)	293(2.4)
	1984	45(1.5)	213(2.4)	62(1.3)*	261(1.3)	59(1.2)*	292(1.4)
Science Books	1994	87(1.8)	211(2.6)	92(1.4)	260(2.1)	84(1.9)	293(2.4)
	1984	84(1.3)	212(1.6)	90(1.1)	259(1.2)	70(1.1)*	289(1.4)
Books About Other Places	1994	79(2.0)	211(2.6)	83(1.8)	260(2.2)	82(2.0)	293(2.3)
	1984	79(1.2)	211(1.7)	83(1.1)	259(1.1)	81(0.9)	289(1.4)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Time Spent on Homework for All Subjects between 1984 and 1994

Past NAEP assessments in reading and other subjects have found relationships between achievement and time spent on homework in all subjects. Table 9.3 presents students' responses regarding time spent on homework and their average reading scale scores. Variations in the relationship between homework time and reading scores were observed across all age groups.

Between 1984 and 1994, some changes were observed in time spent on homework by students at ages 9 and 13; although, it is unclear if the results suggest more or less time spent on homework overall. At age 9, there was an increase in the percentage of students reporting that they completed less than 1 hour of homework per day. This change was accompanied by a decline in the percentage of students reporting that they completed more than 2 hours of homework per day. At age 13, there was an increase in the percentage of students reporting that they did not do their assigned homework. From 1984 to 1994, there were no significant changes in the amount of time that 17-year-olds reported spending on homework.

In 1994 the relationship between amount of time spent on homework and average reading scale scores varied across the three age groups. Among 9-year-olds, students who reported doing more than 2 hours of homework on average each day had lower scale scores than students who reported spending less time on homework. These results may reflect the additional homework assigned to lower achieving students, or the additional time that may be required for these students to complete the regularly assigned homework. In contrast, the relationship between amount of time spent on homework and reading scores was reversed for 13- and 17-year-olds. Students in the two older age groups who reported spending 1 to 2 hours or more than 2 hours doing homework each day had higher scale scores than students who reported spending less time or not doing assigned homework.

Table 9.3

Trends in the Amount of Time Spent on Homework for All Subjects, 1984 and 1994

		AGE 9		AGE 13		AGE 17	
Amount of Homework	**Year**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**	**Percent of Students**	**Average Scale Score**
None	1994	32(2.1)	213(2.0)	23(1.4)	250(1.7)	23(1.4)	273(2.3)
	1984	36(1.3)	213(0.9)	23(0.8)	254(0.8)	22(0.9)	276(0.7)
Didn't Do Assigned Homework	1994	5(0.4)	200(4.3)	6(0.6)	243(5.6)	11(0.6)	285(2.1)
	1984	4(0.3)	199(2.1)	4(0.2)*	247(1.7)	11(0.3)	287(1.2)
Less than 1 Hour	1994	48(1.7)	212(1.4)	34(1.0)	261(1.3)	27(0.9)	288(1.6)
	1984	42(1.0)*	218(0.7)*	36(0.7)	261(0.6)	26(0.4)	290(0.8)
1-2 Hours	1994	12(0.7)	214(3.0)	28(1.0)	268(1.7)	26(1.2)	297(1.7)
	1984	13(0.5)	216(1.3)	29(0.5)	266(0.7)	27(0.5)	296(0.8)
More than 2 Hours	1994	4(0.4)	193(6.1)	9(0.7)	270(2.4)	13(0.9)	306(3.1)
	1984	6(0.2)*	201(1.8)	9(0.3)	265(1.2)	13(0.6)	303(1.1)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in the Extent of Reading in the Home between 1984 and 1994

Social and cultural influences on reading abilities and attitudes are receiving increasing attention among educators. Some researchers have suggested that the home environment is central among these influences.[47] One way in which the home environment can support literacy development is the modeling of reading habits by parents or other adults in the home. Children may come to value the use of literacy materials by observing the important people in their lives engaged in such activities.[48] Furthermore, some research has highlighted the significant effects of home reading activities on both students' reading achievement and their attitudes toward reading.[49] Since 1984, NAEP has asked 13- and 17-year-olds about the extent of reading in their homes. Students were asked to report how often adults they lived with read newspapers, magazines, or books. Students were grouped in three categories: those who reported that the adults they lived with never read newspapers, magazines, or books, or read these materials very infrequently (i.e., yearly or monthly); those who reported that the adults they lived with read these materials on a weekly basis; and those who said they lived with an adult who read these materials on a daily basis. Table 9.4 presents trend results on this important aspect of students' home environment.

No significant differences between 1984 and 1994 were observed in 13- and 17-year-olds' reports about the extent of reading in their homes. In 1994, the reports of students in both age groups were quite similar; over 80 percent reported that reading newspapers, magazines, or books occurred in their homes on at least a weekly basis. However, 19 percent of 13-year-olds and 14 percent of 17-year-olds reported that reading took place in their homes monthly or less frequently. These groups of students had lower average reading scores than their peers who reported weekly or daily reading activities in the home.

---

[47] Chall, J. S.; Jacobs, V. A.; & Baldwin, L. E., *The Reading Crisis: Why Poor Children Fall Behind* (Cambridge, MA: Harvard University Press, 1990).

Stevenson, J. & Fredman, G., "The Social Environmental Correlates of Reading Ability." *Journal of Child and Psychiatry*, 681-698 (July, 1990).

[48] Rogoff, B. *Apprenticeship in Thinking: Cognitive Development in Social Context.* (New York, NY: Oxford University Press, 1990).

[49] Rowe, K. J., "The Influence of Reading Activity at Home on Students' Attitudes Toward Reading, Classroom Attentiveness, and Reading Achievement: An Application of Structural Equation Modeling." *British Journal of Educational Psychology*, 61, 19-35 (February, 1991).

Table 9.4

Trends in the Extent of Reading in the Home, 1984 and 1994

		AGE 13		AGE 17	
Extent of Reading In the Home	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Never/Yearly/Monthly	1994	19(1.7)	237(7.0)	14(1.5)	268(5.5)
	1984	16(1.0)	245(2.0)	14(0.8)	268(2.3)
Weekly	1994	41(1.9)	259(2.7)	44(3.0)	287(4.3)
	1984	43(1.1)	259(2.0)	44(1.1)	288(1.5)
Daily	1994	40(1.9)	267(2.7)	41(2.8)	297(3.3)
	1984	41(0.9)	263(1.8)	42(1.4)	292(1.6)

The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Exposure to Reading Materials in the Home between 1971 and 1994

The availability of reading materials in the home increases opportunities for students to develop as readers and also demonstrates for students the importance of literacy in our daily lives. Because of the potentially significant effects of access to reading materials on students' reading development, NAEP has since 1971 asked students whether they have access to newspapers, magazines, books, and encyclopedias in their homes. The long-term trend results on access to reading materials and the related reading scale scores are displayed in Table 9.5.

Trend results indicate a decline between 1971 and 1994 in the percentages of students who reported having all four types of reading materials in their homes. Corresponding to the decline in access to all four types of materials, a greater percentage of 13- and 17-year-olds in 1994 reported only three of these types of reading materials in their homes. In addition, a greater percentage of students in each age group reported having access to two or fewer types of materials. If having access to reading materials outside the schools is considered important for growth in literacy abilities, then, clearly, those students who have few or no reading materials in their homes may be considered "at-risk" in the sense that they may lack the enrichment and stimulation that is so important for developing readers.[50]

Data from 1994 relating students' access to different types of reading materials to their average scale scores show a clear pattern across all three age groups: increased diversity of reading materials in the home was associated with higher average reading scores.

---

[50] Snow, C.; Barnes, W.; Chandler, J.; Goodman, I.; & Hemphill, L. *Unfulfilled Expectations: Home and School Influences on Literacy* (Cambridge, MA: Harvard University Press, 1991).

Table 9.5

Trends in Numbers of Reading Materials in the Home, 1971 and 1994

		AGE 9		AGE 13		AGE 17	
Numbers of Types of Materials in the Home	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
0-2	1994	38(1.1)	197(1.5)	22(0.9)	238(1.8)	18(0.9)	263(2.0)
	1971	28(0.8)*	186(1.0)*	17(0.6)*	227(1.3)*	11(0.6)*	246(1.8)*
3	1994	32(0.8)	215(1.5)	32(0.9)	258(1.7)	29(0.8)	287(1.8)
	1971	33(0.4)	208(1.0)*	25(0.5)*	249(0.9)*	22(0.5)*	274(1.4)*
4	1994	30(1.2)	225(1.4)	46(1.3)	269(0.9)	53(1.0)	298(1.5)
	1971	39(0.9)*	223(0.9)	58(1.0)*	267(0.7)	67(0.9)*	296(1.0)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Independent Reading Habits between 1984 and 1994

Educators and researchers continue to call for increased encouragement of and support for students' recreational reading. Reading research has documented a positive relationship between reading ability and the likelihood of seeking pleasure in reading experiences.[51] NAEP has asked 9-, 13-, and 17-year-old students about recreational reading habits since 1984. Table 9.6 contrasts the responses of students in 1994 and 1984 to questions about the amount of time they spent reading for fun.

[51] Anderson, R. C.; Wilson, P. T.; & Fielding, L. G., "Growth in Reading and How Children Spend Their Time Outside of School," *Reading Research Quarterly*, 285-303 (Summer, 1988).

Table 9.6

Trends in Amount of Time Spent Reading for Fun, 1984 and 1994

		AGE 9		AGE 13		AGE 17	
Amount of Reading	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Daily	1994	58(1.6)	215(2.3)	32(1.8)	272(3.2)	30(2.6)	302(4.2)
	1984	53(1.0)*	214(1.1)	35(1.0)	264(1.4)*	31(0.8)	297(1.5)
Weekly	1994	25(1.5)	214(3.1)	32(2.1)	255(3.1)	31(1.9)	286(4.1)
	1984	28(0.8)	212(1.7)	35(1.2)	255(1.4)	34(1.1)	290(1.7)
Monthly	1994	5(0.6)	213(5.8)	14(1.7)	255(5.7)	15(1.5)	286(4.5)
	1984	7(0.6)*	204(3.3)	14(0.8)	255(2.1)	17(0.5)	290(1.8)
Yearly	1994	3(0.6)	***(***)	10(1.2)	252(5.4)	12(1.5)	281(8.2)
	1984	3(0.3)	197(4.2)	7(0.5)*	252(3.6)	10(0.5)	280(2.7)
Never	1994	9(0.8)	193(3.9)	12(1.7)	237(5.1)	12(1.4)	258(5.2)
	1984	9(0.5)	198(2.7)	9(0.6)	239(2.5)	9(0.6)	269(2.4)

*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

***Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Few changes were observed between 1984 and 1994 in the reports of students about their recreational reading. A higher percentage of 9-year-olds reported reading for fun daily in 1994 than in 1984. Correspondingly, there was a decline in the percentage of 9-year-olds who reported reading for fun on a monthly basis. Among 13-year-olds, the only change between 1984 and 1994 was an increase from 7 percent to 10 percent in students reporting reading for fun on a yearly basis. No significant differences were observed in 17-year-olds' reports about reading for fun.

These mixed results may seem troubling, since the publication of *Becoming a Nation of Readers: The Report of the Commission on Reading* in 1985 identified independent reading as a necessary aspect of literacy development. The recommendation of that report was that "children should spend more time in independent reading."[52] Unfortunately, since 1984 little progress in this area has been observed.

In 1994 students at age 9 were more likely to report reading on a daily basis than were students at ages 13 and 17. In 1994, daily reading habits were reported by more than one-half of 9-year-olds, but less than one-third of 13- and 17-year-olds. Approximately one-fifth of 13- and 17-year-olds reported that they read for fun only yearly or never. In 1994 those students at all three ages who reported daily reading for fun had higher average reading scores than students who reported never reading for fun.

There are numerous habits and practices characteristic of students who develop into lifelong readers. For example, discussing and sharing books with others has been described as an important literacy activity.[53] Social interaction related to reading may help students to see themselves as participants in a larger literacy community that supports and fosters reading interests.[54] Students who borrow books from the library or who spend their own money on books demonstrate a commitment to literacy practices. Students who seek, select, and read books written by an author they prefer show a sophisticated strategy for choosing books.[55]

Because of the importance of fostering activities that are characteristic of lifelong readers, NAEP reading assessments since 1984 have asked students to identify whether or not they have ever participated in four specific reading-related activities: telling a friend about a good book, taking books out of the library, spending their own money on books, and reading more than one book by an author they liked. Students in the NAEP reading assessments since 1984 have reported their involvement in these literacy activities.

---

[52] Anderson, R. C.; Hiebert, E. H.; Scott, J. A.; & Wilkinson, I. A. G., *Becoming a Nation of Readers: The Report of the Commission on Reading.* (U.S. Department of Education: The National Institute of Education, 1985).

[53] Carlsen, G. R. & Sherrill, A., *Voices of Readers: How We Come to Love Books.* (Urbana, IL: National Council of Teachers of English, 1988).

[54] Martin, P., "Readers/Leaders: Exploring the Why," *Language Arts, 80*, 47-53 (1989).

[55] Hiebert, E. H.; Mervar, K. B.; & Person, D., "Research Directions: Children's Selection of Trade Books in Libraries and Classrooms," *Language Arts*, 67, 758-763 (1990).

Table 9.7 presents the percentages of students in 1984 and 1994 who reported ever engaging in any or all of these four activities and their average reading score scales. In all three age groups, students who reported participating in all four reading-related activities had higher average reading scores than students who reported being involved in fewer of these activities. However, the difference between average scores of 17-year-olds who reported four activities and 17-year-olds who reported three activities was not significant.

At age 9, a greater percentage of students in 1994 than in 1984 reported participating in all four activities. This result was accompanied by a drop in the percentage of nine-year-olds who reported participating in 0 to 1 activities. Some indication of decreased reading-related activities was observed at age 17. The percentage of 17-year-olds reporting involvement in three activities declined between 1984 and 1994. Changes at other levels for age 17 were not significant.

Table 9.7

Trends in Engagement in Reading-Related Activities, 1984 and 1994

		AGE 9		AGE 13		AGE 17	
Number of Activities	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
0-1	1994	6(0.7)	195(5.8)	11(1.5)	236(6.6)	19(1.7)	266(4.3)
	1984	10(0.5)*	205(2.5)	12(0.8)	242(2.1)	17(0.8)	271(1.7)
2	1994	14(0.8)	205(4.1)	16(1.6)	245(4.2)	15(1.2)	282(5.2)
	1984	16(0.8)	208(1.7)	14(0.8)	246(2.6)	14(0.6)	282(2.1)
3	1994	29(1.2)	209(3.0)	29(1.8)	256(3.2)	17(1.7)	286(5.2)
	1984	31(1.0)	211(1.8)	25(0.9)	255(1.5)	23(0.7)*	289(1.8)
4	1994	51(1.2)	220(2.6)	45(2.1)	270(2.8)	49(2.5)	298(2.8)
	1984	44(1.0)*	216(1.5)	49(1.1)	264(1.3)	47(1.3)	298(1.6)

*Statistically significant difference from 1994, where alpha equals.05 per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding error.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessments.

Summary

The results presented in this chapter provide information about students' personal experiences related to reading and about activities and factors in the home environment that are contextually related to reading. In addition, these experiences and contextual factors are linked with students' reading scale scores.

In comparison to their peers in 1984, 9-year-olds in 1994 reported some increases in the number of pages they read each day for school assignments. In 1994, a greater percentage of 9-year-olds reported engaging in more varied reading-related activities but a smaller percentage reported having exposure to poems and plays than in 1984. There was some indication that 9-year-olds were reading for fun more frequently in 1994 than in 1984.

In 1994 13-year-olds reported reading more pages for school and had more exposure to poems and biographies than in 1984. In the last 10 years, the reports of 13-year-olds regarding the amount of time spent on homework, the frequency of reading for fun, and engagement in reading-related activities demonstrated few changes.

In 1994, students at age 17 reported exposure to a wider variety of diverse reading materials than their peers in 1984, but their reports indicated no significant changes in the amount of reading, either for fun or for school assignments. Relative to their peers of 10 years ago, 17-year-olds in 1994 indicated some decline in the amount of reading-related activities.

Home environmental influences on literacy have shown mixed results since 1984. Trend data indicate that there were no statistically significant differences reported by students between 1984 and 1994 in the extent of reading in the home by adults. However, from 1984 to 1994, the percentage of students reporting having access to all four types of reading materials declined at all three ages assessed.

The relationships between factors related to students' reading behaviors and environment and students' overall reading scores have remained stable from 1984 to 1994. As in 1984, current data reveal that, at most ages, higher reading scores were associated with student engagement in more reading and reading-related activities outside the home, living in a home in which adults read regularly, and having a variety of reading materials available to read in the home.

Part IV

Trends in Writing Achievement from 1984 to 1994

Introduction

This section of the report is based on five national assessments of writing conducted during the school years ending in 1984, 1988, 1990, 1992, and 1994. In each of the assessments, nationally representative samples of students in grades 4, 8, and 11 responded to a series of writing tasks. To assess the informative, persuasive, and imaginative writing performance of the nation's students and to track changes in performance across time, the 1994 assessment included a set of 12 writing tasks that had been administered in 1984, 1988, 1990, and 1992.[56] Thus, the same tasks were given to nationally representative samples of students at five different points in time. Students also were asked to answer a brief questionnaire about their writing experiences and instruction.

The past 20 years have seen a dramatic shift in the focus of writing research and practice, away from the text and toward the writer. Our

[56] *Writing Objectives: 1988 Assessment* (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1987).

Writing Objectives: 1984 Assessment (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1984).

understanding of the processes involved in writing and of effective ways to teach writing have evolved during this time.[57] Writing is now seen as recursive process involving invention and brainstorming, drafting and composing, reflecting and revising, and evaluating and editing. Research has shown that students learn to write well by frequent practice and by developing an understanding of the dynamics of the composing process.[58] Research also indicates that students are more likely to write competently when they routinely write in all subject areas and write a variety of types of texts.[59]

Because competence in one type of writing does not necessarily go hand-in-hand with competence in another, the trend writing assessments were designed to examine students' abilities to engage in three types of writing: informative, persuasive, and imaginative.[60] For example, students were asked to complete informative descriptions, reports, and analyses; to write persuasive letters and arguments; and to invent their own stories. The resulting papers were evaluated on the basis of the students' success in achieving the specific purpose of each task (primary trait scoring), their fluency compared to that of other students (as measured by holistic scoring), and their mastery of the conventions of written English (as measured by their spelling, punctuation, and grammar).

Summary of Procedures Used in the NAEP Trend Writing Assessments

To examine trends in writing achievement from 1984 to 1994, one set of analyses used primary trait scoring to focus on the writer's effectiveness in accomplishing each task.[61] Primary trait scoring is designed to be sensitive to the writer's understanding of the audience as well as to the inclusion of specific features needed to accomplish the specific purpose of that task. The

---

[57] Emig, J., *The Composing Process of Twelfth Graders* (Urbana, IL: National Council of Teachers of English. Research Report No. 13, ERIC Document No. ED 058205, 1971).

[58] Britton, J., *Prospect and Retrospect: Selected Essays of James Britton,* M. Pradl Gordon, editor (Montclair, NJ: Boynton/Cook Publishers, Inc., 1982).

[59] Hillocks, Jr., G., *Research on Written Composition: New Directions for Teaching* (Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills, 1986).

[60] Pringle, I. & Freedman, A., *A Comparative Study of Writing Abilities in Two Modes at the Grade 5, 8, and 12 Levels* (Toronto, Ontario: The Minister of Education, Ontario, 1985).

[61] Lloyd-Jones, R., "Primary Trait Scoring," in *Evaluating Writing: Describing, Measuring, Judging,* C. R., Cooper and L. Odell, editors (Urbana, IL: National Council of Teachers of English, 1977).

primary trait scoring criteria, while specific to each writing prompt, defined five levels of task accomplishment: not rated, unsatisfactory, minimal, adequate, and elaborated. General definitions of these levels are provided below.

Levels of Task Accomplishment

Level 4 - Elaborated

Students providing elaborated responses went beyond the essential, reflecting a higher level of coherence and providing more detail to support the points made.

Level 3 - Adequate

Students providing adequate responses included the information and ideas necessary to accomplish the task and were considered likely to be effective in achieving the desired purpose.

Level 2 - Minimal

Students writing at the minimal level recognized some or all of the elements needed to complete the task but did not manage these elements well enough to assure that the purpose of the task would be achieved.

Level 1 - Unsatisfactory

Students who wrote papers judged as unsatisfactory provided very abbreviated, circular, or disjointed responses that did not even begin to address the writing task.

Level 0 - Not rated

A small percentage of the responses were blank, indecipherable, completely off task, or contained a statement to the effect that the student did not know how to do the task; these responses were not rated.

The writing samples generated by students in the assessments represent their ability to produce first-draft writing on demand in a relatively short time and under less than ideal conditions. The guidelines for evaluating task accomplishment are designed to reflect these constraints and do not require a finished performance. Because primary trait scoring is based on established criteria, it is theoretically possible for all papers to be rated at the highest level on a straightforward task. Overall, the primary trait scoring procedure provides the best assessment of students' ability to perform each task.

A second set of analyses, based on general impression or holistic scoring, focused on the writer's fluency in responding to each task relative to the performance of other students at that grade level.[62] In holistic scoring, readers do not make separate judgments about specific aspects of writing, but instead consider the overall effect, rating each paper on a six-point scale on the basis of the paper's general fluency. Unlike primary trait scores, the average score for a set of papers rated holistically will generally fall near the midpoint of this scale. Thus, while primary trait scoring permits year-to-year and grade-level to grade-level comparisons based on specific criteria, holistic scoring permits year-to-year comparisons of relative fluency at each grade.

The final set of analyses, applied to a subset of the papers, focused on the mechanics of students' writing. Students' mastery of the sentence-level and word-level conventions of English, as well as their use of correct punctuation, were examined.[63] (See the Procedural Appendix for the scoring scheme.)

To further analyze trends in students' writing performance, the primary trait results across tasks were aggregated using sophisticated item response theory (IRT) scaling techniques that account for the multi-level student responses to the individual writing tasks (see Procedural Appendix for details). The trend writing scale ranges from 0 to 500.

An additional analysis was performed to relate the writing scale to the levels of student performance defined in the scoring guides. A special mapping procedure was used to profile students' task-by-task performance in relation to percentiles on the scale. For each grade, students' levels of performance on each individual writing task are mapped in relation to their overall writing achievement as summarized by the NAEP trend writing scale (see Chapter 11).

The NAEP writing scale discussed in this report should not be confused with the NAEP writing scale used to report student performance on the 1992 assessment at grades 4, 8, and 12.[64] The 1992 assessment contained an entirely different set of writing tasks, and students were given almost twice as much time to respond. At the eighth and twelfth grades, several 50-minute tasks were administered. Students were given a planning page in order to make notes and sketch out their ideas. New scoring criteria were also developed to

---

[62] Cooper, C. R., "Holistic Evaluation of Writing," in *Evaluating Writing: Describing, Measuring, Judging,* C. R. Cooper and L. Odell, editors (Urbana, IL: National Council of Teachers of English, 1977).

[63] Shaughnessy, M. P., *Errors and Expectations: A Guide for the Teacher of Basic Writing* (New York: Oxford University Press, 1977).

[64] Applebee, A. N.; Langer, J. A.; Mullis, I. V. S.; Latham, A. S.; & Gentile, C. A., *NAEP 1992 Writing Report Card* (Washington, DC: National Center for Education Statistics, 1994).

meet the expanded and more demanding nature of the tasks. Six levels of task accomplishment were defined and employed to classify and evaluate students' responses. Further the 1992 NAEP writing assessment was administered in the spring to grades 4, 8, and 12, while the trend assessment is administered in the fall at grade 8, the winter at grade 4, and the spring at grade 11. Thus, differences between the trend assessment and the 1992 assessment preclude direct comparisons between the two assessments.

NAEP reports the performance of groups of students, not individuals. The results in this trend report include measures of average performance on the trend writing scale for groups of students, the percentages of students attaining various levels on the trend scale, and the percentages of students responding to each of the assessment tasks at different rating levels. Because these averages and the percentages are based on samples, they are necessarily estimates. Like all estimates based on surveys, they are subject to sampling error. NAEP uses a complex procedure to compute standard errors that estimate the sampling error and other random error associated with observed assessment results. In the tables and figures presented in this report, each average scale score or percentage is presented with a standard error — an estimate of the sampling error and measurement errors. Statistically significant differences between previous assessments and 1994 are denoted with an asterisk, and statistically significant differences between 1984 and subsequent assessments are denoted with a dagger.

In addition to point-by-point multiple comparisons between assessment years, a second type of significance test was conducted to detect statistically significant linear and quadratic trends across the assessments. (See the Procedural Appendix for a discussion of the procedure.) This type of analysis makes it possible to discuss statistically significant patterns that may be missed by point-to-point comparisons. For example, from assessment to assessment, students' average scale scores may consistently increase by a small amount. Although these small increases between years may not be statistically significant under pairwise multiple comparisons, the overall increasing trend in average scores may be statistically significant and noteworthy. All of the differences and trend patterns discussed in this report are statistically significant at the .05 level.

The first chapter in Part IV, Chapter 10, presents trends in average writing scale scores for the nation and various demographic subpopulations, offering a global view of the assessment results. Chapter 11 presents trends in performance at each of the five performance levels along the trend writing scale.

Chapter 12 describes student performance on the informative, persuasive, and imaginative writing tasks included in the writing trend assessments, based on the results of the primary trait and holistic analyses. Trends in students' grammar, punctuation, and spelling are also discussed in Chapter 12. Lastly, Chapter 13 relates various background factors to students' performance, such as their instructional experiences and characteristics of their home environment.

10

National Trends in Writing Scale Scores from 1984 to 1994

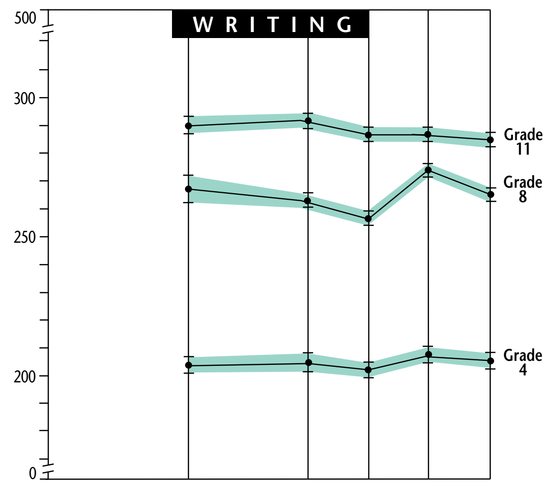
NAEP conducted trend writing assessments in 1984, 1988, 1990, 1992, and 1994. Each assessment involved nationally representative samples of fourth, eighth, and eleventh graders. In each trend assessment year, the same set of writing tasks or prompts were presented to students, with some tasks administered at overlapping grades.

As explained in the preceding introduction, student responses to the writing prompts were evaluated according to a 4-level description of task accomplishment: Unsatisfactory, Minimal, Adequate, and Elaborated. To further analyze trends in students' writing ability, the results across tasks were aggregated using item response theory (IRT) scaling techniques that account for the multiple-level rating criteria used to evaluate student responses (see the Procedural Appendix for details). The NAEP writing scale resulting from these analyses ranges from 0 to 500. This chapter presents overall trends in writing for the nation and for demographic subgroups based on students' average scale scores.

Figure 10.1 presents trend results in writing for the nation from 1984 to 1994.

Figure 10.1

Trends in Average Writing Scale Scores for the Nation, 1984 to 1994



	1984	1988	1990	1992	1994
Grade 11	290(1.6)*	291(1.3)*	287(1.0)	287(1.4)	285(1.2)†
Grade 8	267(2.0)	264(1.3)	257(1.2)*†	274(1.3)*†	265(1.3)
Grade 4	204(1.5)	206(1.6)	202(1.5)	207(1.5)	205(1.6)

95-percent confidence interval

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Fourth Grade. Across the five writing trend assessments since 1984, no significant changes in fourth graders' performance were observed.

Eighth Grade. The average writing score of eighth graders decreased from 1984 to 1990. In 1992, the average score increased and was higher than the average in 1984. Although there was a decrease between 1992 and 1994, the average score remained higher than the 1990 low point. The 1994 average score, however, was not significantly different from the 1984 average.

Eleventh Grade. The average score of eleventh graders showed an overall pattern of decrease across the assessment years. In 1994, the average score was lower than the average in 1984.

Trends in Writing Scale Scores from 1984 to 1994 by Quartiles

Table 10.1 shows trends in writing scale scores for students at each grade who were in the upper quartile (upper 25 percent), the middle two quartiles (middle 50 percent), and the lower quartile (lower 25 percent) of student performance in each assessment. The results presented by quartiles mostly parallel results observed for the nation.

Despite some fluctuation in the average score for fourth graders in the upper quartile, the 1994 average score for these students did not differ significantly from the 1984 average. For fourth graders in the middle two and lower quartiles, no significant changes in performance since 1984 were observed.

Eighth graders in the upper quartile displayed an overall pattern of increased performance across the assessment years. Although the 1994 average score for these students decreased from the high point reached in 1992, it remained at a level not significantly different than the 1984 average. The pattern of performance for eighth graders in the middle two and lower quartiles was similar to that observed for the nation. Average scores decreased in 1990 and then increased in 1992 to a level higher than that in 1984. However, a decline in performance since 1992 resulted in a 1994 average score that was not significantly different from the 1984 average.

At the eleventh grade, students in the upper quartile attained an average score in 1994 that was lower than that attained in 1990, but not significantly different from the average score in 1984. Eleventh graders in the middle two and lower quartiles demonstrated an overall pattern of decreased performance. The 1994 average score for eleventh graders in the middle two quartiles was lower than the 1984 average. While the pattern of decline was evident in the lowest quartile as well, there was no significant difference between 1984 and 1994 average scores for this group.

Table 10.1

Trends in Average Writing Scale Scores by Quartiles, 1984 to 1994

		Average Scale Score		
Quartile	Year	Grade 4	Grade 8	Grade 11
Upper Quartile	1994	242(1.3)	301(1.4)	319(1.3)
	1992	246(1.3)	311(1.0)*†	319(2.0)
	1990	242(1.6)	294(1.1)*	325(1.6)*
	1988	247(1.7)	296(1.3)*	320(1.0)
	1984	241(2.3)	297(1.4)	322(1.9)
Middle Two Quartiles	1994	206(1.2)	267(1.3)	285(1.1)†
	1992	208(1.7)	275(1.4)*	289(1.1)
	1990	203(1.5)	257(1.0)*†	287(0.8)†
	1988	207(1.5)	265(1.0)	293(1.2)*
	1984	204(1.4)	268(2.4)	291(1.2)*
Lower Quartile	1994	164(2.7)	227(1.8)	249(1.8)
	1992	167(1.4)	237(2.1)*	253(1.5)
	1990	158(2.4)	219(1.5)*†	250(2.1)
	1988	162(1.8)	231(1.8)	260(2.4)*
	1984	166(2.6)	233(3.2)	255(2.0)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Scale Scores from 1984 to 1994 by Race/Ethnicity and Gender

Race/Ethnicity. Trends in average writing scale scores for White, Black, and Hispanic students are presented in Figure 10.2.[65] Among White fourth-grade students, no significant changes across the assessments were observed. At the eighth grade, the performance of White students decreased from 1984 to 1990. In 1992 performance improved, reaching an average score that was higher than the 1984 average. However, in 1994 the average score for these students returned to a level that was not significantly different from that in 1984. An overall decrease in performance was observed for White eleventh graders; scores in 1994 were lower than they had been in 1984.

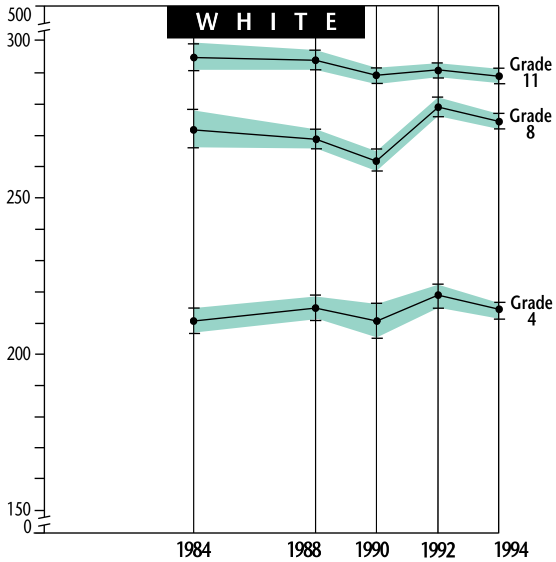
Similar to their White peers, fourth grade Black students displayed little or no change across the assessment years. The average score of Black eighth graders decreased from 1992 to 1994, but no significant difference was observed between the first and most recent assessments. At the eleventh grade, the performance of Black students displayed no significant changes across the assessment years.

Among Hispanic students, the performance of fourth graders remained relatively stable since 1984. For eighth-grade Hispanic students, the 1992 average score was higher than the 1984 average. However, similar to other groups of eighth graders, their average score declined in 1994 to a level not significantly different from that in 1984. The performance of Hispanic eleventh graders over the assessment years has followed a pattern somewhat different from other ethnic groups. The 1990 average was higher than the 1984 average. Since 1990, scores have seemed to follow a pattern of slight decline. The 1994 average was not different to a statistically significant degree from either the 1984 average or the 1990 high point.

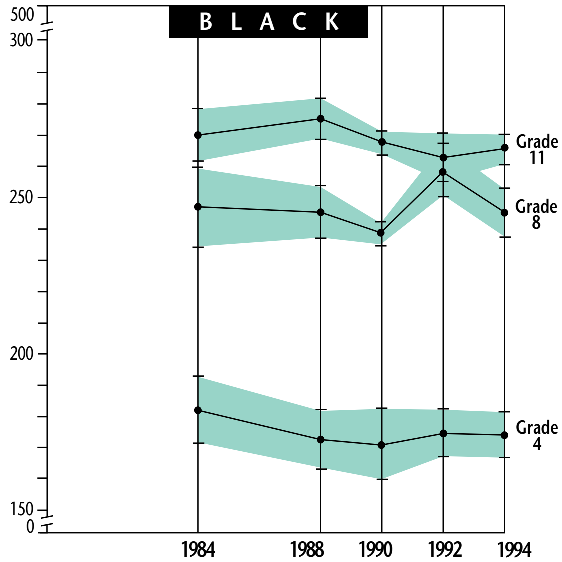
[65] For Asian/Pacific Islander or American Indian students, the sample sizes were insufficient to permit reliable trend estimates.

Figure 10.2

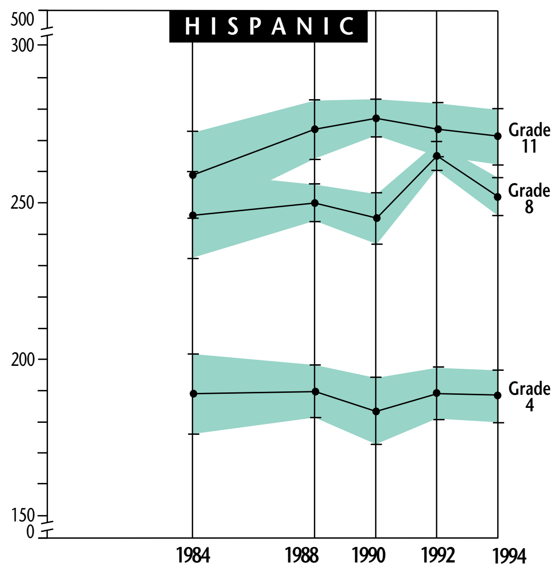
Trends in Average Writing Scale Scores by Race/Ethnicity, 1984 to 1994



	1984	1988	1990	1992	1994
Grade 11	297(1.8)* 75(1.1)*	296(1.3)* 74(0.1)*	293(1.2) 71(0.2)*†	294(1.2) 71(0.2)*†	291(1.4)† 68(0.3)†
Grade 8	272(2.1) 76(0.9)*	269(1.3) 71(0.2)*†	262(1.6)*† 70(0.2)†	279(1.3)*† 70(0.3)†	272(1.4) 70(0.2)†
Grade 4	211(1.9) 71(1.0)	215(1.9) 70(0.3)*	211(2.0) 70(0.3)*	217(1.7) 71(0.3)*	214(1.5) 69(0.4)



	1984	1988	1990	1992	1994
Grade 11	270(3.6) 15(1.0)	275(2.9) 15(0.1)	268(2.3) 16(0.2)*	263(3.2) 14(0.2)*	267(2.2) 15(0.2)
Grade 8	247(5.7) 12(0.6)*	246(3.5) 15(0.2)†	239(2.3) 15(0.2)†	258(4.0)* 16(0.1)†	245(3.4) 15(0.2)†
Grade 4	182(5.0) 15(0.9)	173(4.7) 16(0.2)	171(5.4) 15(0.3)	175(3.8) 16(0.2)	173(3.2) 15(0.3)



	1984	1988	1990	1992	1994
Grade 11	259(6.6) 8(0.6)*	274(4.4) 8(0.1)*	277(2.6)† 9(0.1)*	274(3.8) 11(0.2)*†	271(4.0) 12(0.2)†
Grade 8	247(6.4) 8(0.7)*	250(2.5) 10(0.1)†	246(2.8) 10(0.2)†	265(2.2)*† 10(0.2)†	252(3.3) 11(0.2)†
Grade 4	189(5.8) 11(0.7)	190(3.5) 11(0.3)*	184(4.1) 11(0.2)	189(3.6) 10(0.2)*	189(3.1) 12(0.3)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

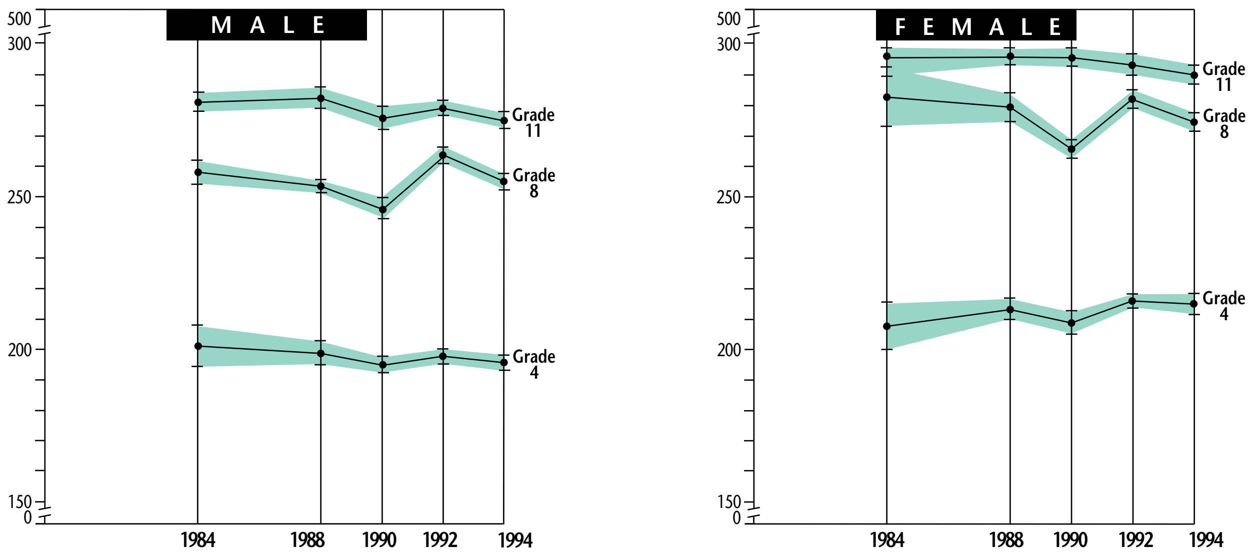
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Gender. Figure 10.3 presents trends in average writing scale scores for male and female students. Among male students, fourth graders displayed no significant changes in performance across the assessment years. At the eighth grade, the average score of males decreased between 1984 and 1990. Although an increase was observed in 1992, the average score returned to a level in 1994 that was not significantly different from that in 1984. The 1994 scores did remain above the 1990 lowpoint. Eleventh-grade male students displayed an overall pattern of decreased performance. Although their 1994 average score was not significantly different from the 1984 average, it was lower than that observed in 1988.

Similar to their male peers, female fourth graders showed no significant change in performance since 1984. Female eighth graders had lower performance in 1990 than in 1984. In 1992, they attained an average score that was significantly higher than that in 1984. However, a decline since 1992 resulted in an average score in 1994 that was not significantly different from that observed in the first assessment. At the eleventh grade, the 1994 average score was lower than 1988 and 1990 average scores, although not significantly different from the average in the first assessment.

Figure 10.3

Trends in Average Writing Scale Scores by Gender, 1984 to 1994



	Male 1984	Male 1988	Male 1990	Male 1992	Male 1994	Female 1984	Female 1988	Female 1990	Female 1992	Female 1994
Grade 11	281(1.4)	282(2.0)*	276(1.6)	279(1.2)	276(1.5)	299(2.5)	299(1.2)*	298(1.5)*	296(2.0)	293(1.5)
	51(1.6)	47(1.5)	51(0.8)	53(1.1)	49(1.7)	49(1.6)	53(1.5)	49(0.8)	47(1.1)	52(1.7)
Grade 8	258(2.3)	254(1.5)	246(1.5)*†	264(1.9)*	254(1.8)	276(2.4)	274(1.7)	268(1.3)*†	285(1.3)*†	278(1.4)
	51(1.2)	49(1.1)	51(0.9)	50(0.8)	52(0.9)	49(1.2)	51(1.1)	50(0.9)	51(0.8)	48(0.9)
Grade 4	201(2.8)	199(2.3)	195(1.9)	198(1.7)	196(1.7)	208(3.1)	213(2.0)	209(2.2)	216(1.7)	214(2.2)
	53(1.7)	50(1.0)	51(1.0)	50(0.6)	50(0.9)	47(1.7)	50(1.0)	49(1.0)	50(0.6)	50(0.9)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

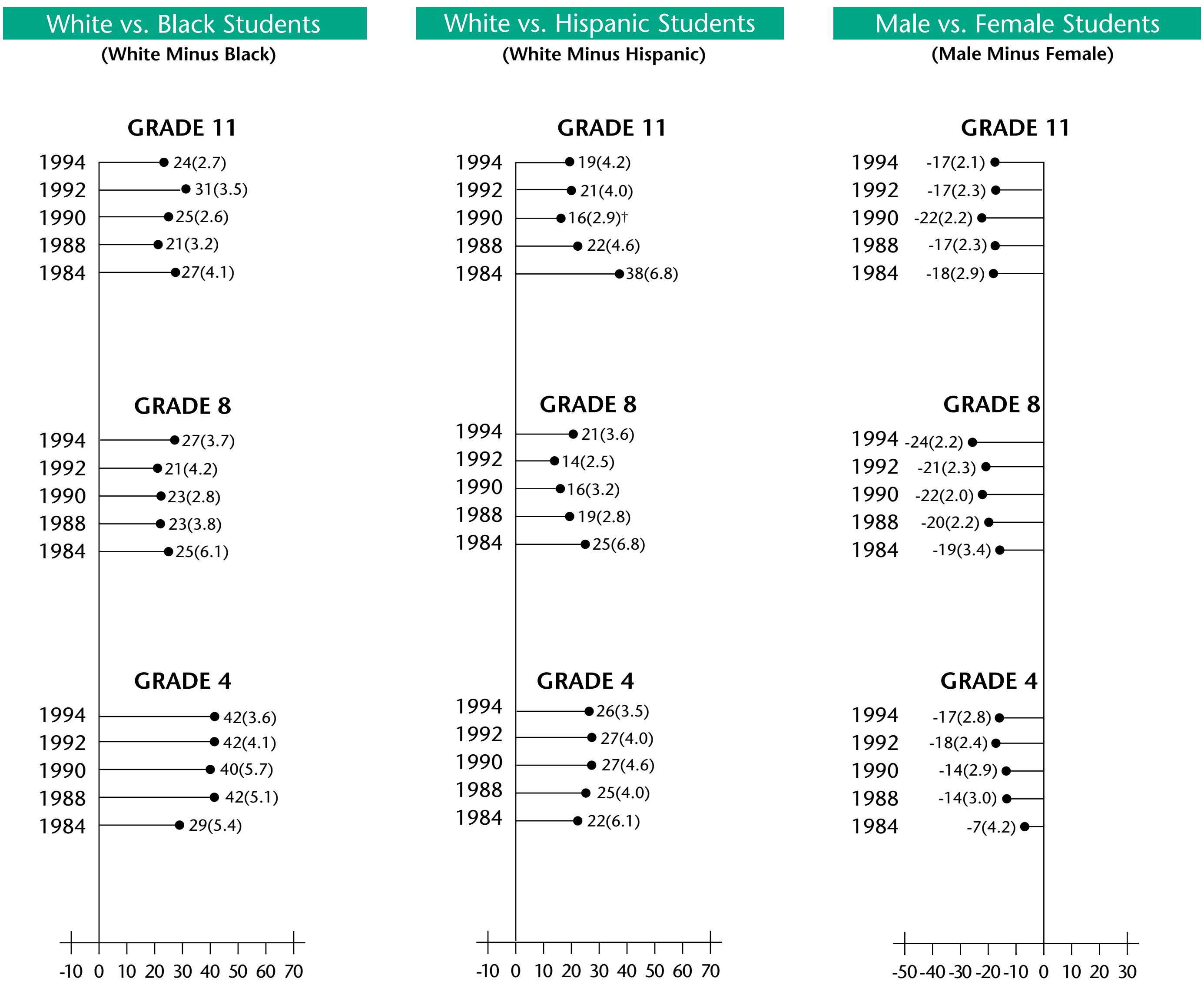
Trends in Differences in Average Writing Scale Scores

The previous section discussed the trends in writing achievement for White, Black, and Hispanic students, and for male and female students. As with past NAEP assessments, significant differences were observed between racial/ethnic subgroups and between males and females. Figure 10.4 presents trends in differences between the average scale scores for selected subgroups of students across the assessment years.

A comparison of performance in 1994 across the three racial/ethnic groups showed that, at all three grades, White students outperformed their Black and Hispanic peers. Also at grade 4, Hispanic students outperformed Black students. Across the assessment years, there has been no statistically significant change in the gap between White and Black students at all three grades. Also, the gap between White and Hispanic students has remained relatively stable at the fourth and eighth grades. Among eleventh graders, however, there was evidence that the gap between White and Hispanic students decreased across the assessments, although the magnitude of the gap in 1994 was not significantly different from that observed in 1984.

At each grade in 1994, female students outperformed male students on the writing assessment. At the fourth grade, there was some indication that the gap between males and females increased across the assessment years, although there was no significant difference between 1984 and 1994 in the magnitude of the gap. No significant changes in the gap between males and females at grades 8 and 11 were observed.

Figure 10.4
Trends in Differences in Average Writing Scale Scores



*Statistically significant difference from 1994, at a 5 percent combined significance level per set of comparisons.

† Statistically significant difference from 1969-70 (for scale scores for White vs. Black student differences and Male vs. Female student differences) or from 1977 (for White vs. Hispanic student differences), at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated scale score differences appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Scale Scores from 1984 to 1994 by Region

Figure 10.5 presents trends in average writing scale scores for students in the Northeast, Southeast, Central, and West regions of the country. In the Northeast, both fourth and eleventh graders showed no significant changes between 1984 and 1994. At the eighth grade, the overall pattern was one of improved performance. The 1994 average score of these students was higher than the average scores in 1988 and 1990, although it did not differ significantly from the average score in 1984.

In the Southeast, although a lower average score was observed in 1992 compared to 1984, the performance of fourth graders in 1994 was not significantly different from that in the first assessment. The average score of eighth graders declined between 1988 and 1990. However, the 1994 average did not differ significantly from that in 1984. The overall pattern among eleventh graders was one of decreased performance. The average score of these students in 1994 was lower than the average in 1988, but it did not differ significantly from the average in 1984.

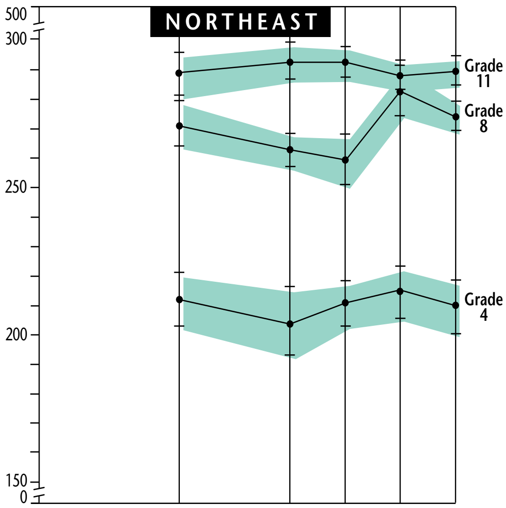
In the Central region, both fourth and eighth graders showed an overall pattern of slight improvement across the assessment years. For both groups, the 1992 average score was higher than the 1984 average score. However, in 1994 their performance was not significantly different from that observed in 1984. The 1994 average score of eleventh graders was not significantly different from the 1984 average.

In the West region, fourth and eleventh graders had relatively stable performance across the assessment years. At the eighth grade, a decline was observed between 1992 and 1994, but there was no significant difference between the first and most recent assessment years.

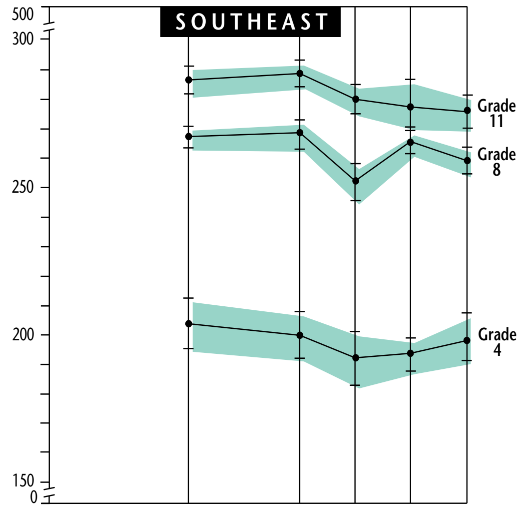
Comparisons of 1994 average scores across the four regions revealed three instances of differences. Eighth graders in the Northeast outperformed their counterparts in both the Southeast and West regions. Also, the average scores of eleventh graders in the Northeast were higher than that of eleventh graders in the Southeast. There were no statistically significant differences in regional performance at fourth grade.

Figure 10.5

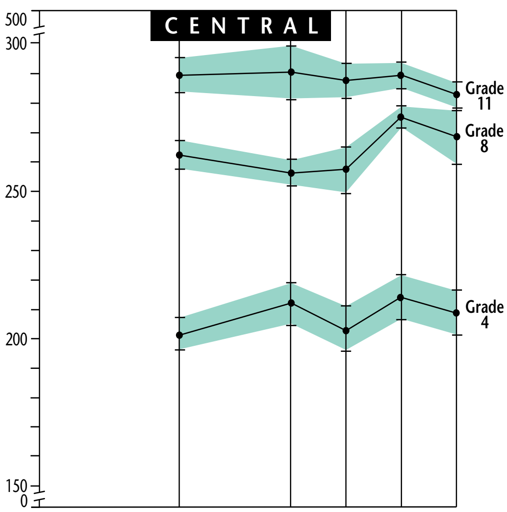
Trends in Average Writing Scale Scores by Region, 1984 to 1994



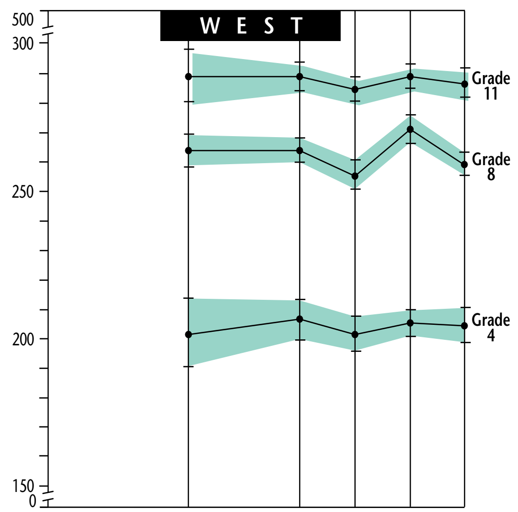
	1984	1988	1990	1992	1994
Grade 11	291(3.0) 26(0.8)*	295(2.8) 23(0.8)†	295(2.5) 22(1.0)†	290(2.3) 22(0.9)†	291(2.4) 22(0.8)†
Grade 8	273(3.6) 23(1.0)	265(2.7)* 23(2.2)	261(3.3)* 23(1.0)	285(3.3) 21(1.0)	277(2.2) 21(0.7)
Grade 4	212(4.0) 20(0.7)	204(4.9) 23(1.7)	211(3.6) 22(1.0)	216(4.0) 21(1.3)	210(4.0) 21(1.1)



	1984	1988	1990	1992	1994
Grade 11	287(4.9) 22(1.7)	289(2.2)* 22(1.0)	280(2.3) 23(0.9)	278(3.3) 24(1.4)	277(2.5) 23(1.3)
Grade 8	267(3.6) 23(1.7)	268(2.3)* 24(1.3)	252(2.8)† 25(1.1)	266(2.2) 25(1.5)	259(2.1) 23(1.8)
Grade 4	204(3.3) 25(1.5)	200(3.2) 26(1.7)	192(4.0) 24(1.2)	193(2.4)† 24(1.3)	198(3.4) 25(1.8)



	1984	1988	1990	1992	1994
Grade 11	291(2.7) 27(1.6)	292(4.0) 27(1.7)	289(2.7) 27(1.1)	291(2.2) 26(1.1)	284(2.3) 26(1.4)
Grade 8	264(2.3) 27(1.6)	258(2.2) 26(2.2)	259(3.9) 23(0.7)	277(2.0)† 25(1.0)	270(4.1) 26(1.6)
Grade 4	201(2.6) 28(1.5)	212(3.0)† 23(0.5)†	203(3.1) 25(0.7)	214(3.1)† 28(0.8)	209(3.3) 25(1.2)



	1984	1988	1990	1992	1994
Grade 11	289(3.7) 25(0.8)*	289(2.3) 28(1.1)†	285(2.1) 28(0.9)	289(2.1) 29(1.0)†	287(2.7) 29(1.0)†
Grade 8	264(3.0) 26(0.8)*	264(2.1) 28(1.1)	255(2.6) 29(1.0)	271(2.3)* 29(1.2)	259(1.6) 30(1.2)†
Grade 4	201(4.9) 27(0.9)	207(3.4) 28(1.1)	201(2.7) 29(0.8)	206(2.2) 28(0.7)	203(2.7) 29(1.0)

Below each average scale score, the corresponding percentage of students is presented.

95-percent confidence interval

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Scale Scores from 1984 to 1994 by Parents' Highest Level of Education

Table 10.2 presents trends in average writing scale scores by students' reports about their parents' highest level of education. It should be noted that there was an increase from 1984 to 1994 in the percentage of students at each grade who reported that at least one parent graduated from college. Correspondingly, the percentage of students who reported high school graduation as their parents' highest level of education decreased during the same time period. As was observed in the other subject areas, students who reported higher levels of parental education generally had higher average scale scores.

Among fourth graders who reported that the highest level of parental education was high school graduation, an overall pattern of moderately increased performance was observed. However, the 1994 average score of these students did not differ significantly from the 1984 average. At all other levels of parental education, the performance of fourth graders was relatively stable across the assessment years, with 1994 scores not significantly different from 1984 scores.

Eighth graders who reported the highest level of parental education (college graduation) displayed a decline in 1990, followed by increased performance in 1992. In 1994 scores had again declined to a point not significantly different from the 1984 average. However, 1994 performance remained better than 1990 performance. The average score of eighth graders who reported that high school graduation was their parents' highest level of education declined from 1984 to 1990. In 1992, the average score increased to a level higher than in 1984. However, a decline in 1994 returned the average score to a level not significantly different from that in the first assessment. At the other levels of parental education, no significant changes across the assessment years were observed in eighth graders' average scores.

At the two highest levels of parental education, eleventh graders displayed an overall pattern of decreased performance, resulting in 1994 average scores that were lower than those of their counterparts in 1984. No significant changes were observed for eleventh graders at the two lower levels of parental education.

Table 10.2

Trends in Average Writing Scale Scores by Parents' Highest Level of Education, 1984 to 1994

Level of Education	Year	GRADE 4 Percent of Students	GRADE 4 Average Scale Score	GRADE 8 Percent of Students	GRADE 8 Average Scale Score	GRADE 11 Percent of Students	GRADE 11 Average Scale Score
Graduated From College	1994	43(1.4)†	212(2.1)	46(1.6)†	275(1.3)	44(1.9)†	293(1.5)†
	1992	42(1.0)†	214(1.4)	44(1.8)†	284(1.9)*	43(1.4)†	296(1.4)
	1990	40(1.6)†	209(1.6)†	38(1.5)*	265(1.8)*†	41(1.4)	298(2.0)
	1988	41(1.5)†	212(2.2)	41(1.5)	271(1.8)†	41(1.8)	299(2.0)*
	1984	33(1.4)*	218(3.0)	36(1.5)*	278(1.8)	36(1.7)*	300(2.4)*
Some Education After High School	1994	5(0.5)	212(4.0)	12(0.7)	270(3.1)	20(1.0)†	286(1.7)†
	1992	6(0.4)	201(4.5)	12(0.7)†	280(2.2)	20(0.8)†	292(2.0)
	1990	5(0.4)	214(4.0)	12(0.7)	267(3.0)	19(0.6)†	292(2.7)
	1988	5(0.5)	211(6.3)	11(0.6)	275(3.3)	18(0.8)	296(2.6)*
	1984	5(0.4)	208(6.5)	10(0.8)	271(3.9)	15(0.9)*	298(2.5)*
Graduated From High School	1994	16(0.9)†	202(2.3)	27(1.4)†	259(2.2)	26(1.1)†	279(1.7)
	1992	17(0.7)†	202(3.2)	29(1.1)†	268(1.6)*†	27(0.9)†	278(2.2)
	1990	18(0.9)	197(3.0)	33(1.1)*	253(1.4)†	30(1.1)	278(1.9)
	1988	18(1.1)	199(3.0)	32(1.2)	258(2.1)	30(1.2)*	285(2.2)
	1984	20(1.1)*	192(3.4)	35(1.4)*	261(1.6)	35(2.1)*	284(3.0)
Less Than High School	1994	4(0.4)†	188(7.8)	7(0.4)†	250(4.1)	8(0.7)	269(4.7)
	1992	5(0.4)†	191(3.2)	7(0.8)†	258(5.3)	8(0.8)	271(3.7)
	1990	6(0.5)	186(3.9)	8(0.6)	246(3.7)	9(0.5)	268(4.0)
	1988	5(0.7)	194(5.4)	9(0.7)	254(3.9)	8(0.8)	276(3.5)
	1984	7(0.6)*	179(4.6)	10(0.8)*	258(4.8)	11(1.2)	274(5.2)
I Don't Know	1994	31(1.2)	197(1.9)	8(0.6)	240(2.8)	3(0.2)	251(3.2)
	1992	31(1.0)	205(2.3)	8(0.6)	250(3.1)	3(0.3)	259(6.6)
	1990	32(1.1)	197(2.2)	10(0.6)	235(2.8)	3(0.3)	259(4.1)
	1988	30(1.4)	202(2.7)	8(0.5)	248(2.3)	3(0.3)	268(4.7)*
	1984	35(1.3)	203(2.5)	9(0.8)	249(6.7)	3(0.5)	***(***)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Scale Scores from 1984 to 1994 by Type of School

Students' average writing scale scores by the type of school they attended are shown in Table 10.3. The performance of fourth graders in both public and nonpublic schools has remained relatively stable across the five trend assessments. Among public school eighth graders, the average score in 1990 was lower than that in 1984. The average score then increased in 1992 to a level higher than in 1984. In 1994, scores declined. While the resulting 1994 average was not significantly different from that in 1984, it remained above the 1990 low point. No significant changes across the assessment years were observed for eighth graders in nonpublic schools. The average scores of eleventh-grade students in both public and nonpublic schools decreased across the assessment years.

Comparing the 1994 average scores of public and nonpublic school students revealed no significant differences at either the fourth or eleventh grades. At the eighth grade, however, nonpublic school students had a higher average writing score in 1994 than did their public school peers.

Table 10.3

Trends in Writing Scale Scores by Type of School, 1984 to 1994

		GRADE 4		GRADE 8		GRADE 11	
Type of School	Year	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score
Public	1994	90(2.0)	204(1.8)	89(1.6)	264(1.6)	88(2.2)	284(1.4)
	1992	88(1.7)	205(1.6)	87(2.0)	272(1.3)*†	91(2.2)	287(1.6)
	1990	92(1.8)	200(1.4)	87(1.9)	254(1.2)*†	92(1.7)	286(1.1)
	1988	88(3.0)	204(2.0)	88(2.7)	262(1.5)	86(3.8)	290(1.2)*
	1984	86(1.9)	202(1.8)	87(1.6)	264(2.0)	89(1.5)	288(1.6)
Nonpublic	1994	10(2.0)	213(4.3)	11(1.6)	279(3.8)	12(2.2)	291(3.8)†
	1992	12(1.7)	222(3.3)	13(2.0)	288(3.2)	9(2.2)	295(4.4)
	1990	8(1.8)	216(5.7)	13(1.9)	277(4.4)	8(1.7)	306(5.2)
	1988	13(3.0)	216(4.1)	12(2.7)	276(3.0)	14(3.8)	300(3.6)
	1984	14(1.9)	215(4.6)	13(1.6)	282(5.5)	11(1.5)	305(3.7)*

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Scale Scores from 1984 to 1994 by Modal Age

Students' average writing scores by modal age are shown in Table 10.4. The modal age is the typical age of students attending a particular grade. The modal ages are: age 9 for grade 4, age 13 for grade 8, and age 17 for grade 11. Students who are less than modal age are younger than the expected age of students in their grade. Conversely, students who are greater than modal age are older than the expected age of students in their grade.

Fourth graders who were at modal age demonstrated stable performance across the assessment years. Fourth-grade students who were greater than modal age displayed a pattern of slightly increased performance, although there were no significant differences between the 1994 and 1984 average scores.

Eighth graders who were at modal age had a lower average score in 1990 than in 1984. The 1992 average score then increased to a level higher than that in 1984. However, a decline in 1994 returned the average score to a level not significantly different from that in 1984, but still higher than in 1990. The average score of eighth graders greater than modal age was lower in 1990 than in 1984. Scores increased in 1992, and declined again in 1994, although they continued to be higher than in 1990. However, the 1994 and 1984 average scores did not differ significantly.

Eleventh graders who were less than or greater than modal age had relatively stable performance across the assessment years. For eleventh graders at modal age, there was a decrease in the average score from 1984 to 1994.

Table 10.4

Trends in Average Writing Scale Scores by Modal Age

Modal Age	Year	GRADE 4 Percent of Students	GRADE 4 Average Scale Score	GRADE 8 Percent of Students	GRADE 8 Average Scale Score	GRADE 11 Percent of Students	GRADE 11 Average Scale Score
Less Than Modal Age	1994	1(0.2)	***(***)	1(0.2)	***(***)	11(0.9)	291(3.2)
	1992	0(0.1)	***(***)	1(0.2)	***(***)	10(0.8)	295(4.0)
	1990	0(0.1)	***(***)	1(0.2)	***(***)	11(0.7)	295(4.3)
	1988	1(0.2)	***(***)	1(0.2)	***(***)	11(0.8)	299(2.1)
	1984	1(0.3)	***(***)	1(0.3)	***(***)	13(1.3)	295(3.8)
At Modal Age	1994	60(1.0)	208(1.6)	57(1.0)†	272(1.2)	64(0.3)†	290(0.9)†
	1992	56(1.1)*†	212(1.5)	58(1.2)†	282(1.3)*†	64(0.2)	293(1.7)
	1990	59(1.1)	205(2.7)	59(0.3)*†	262(1.5)*†	64(0.2)†	292(1.1)
	1988	61(0.5)	210(2.5)	59(0.2)†	271(1.2)	68(0.2)*	297(1.2)*
	1984	63(1.0)	212(2.5)	64(1.4)*	272(1.8)	67(1.2)*	296(1.4)*
Greater Than Modal Age	1994	39(1.0)	200(3.0)	43(1.0)†	256(2.4)	26(0.9)†	269(3.4)
	1992	44(1.1)*†	201(1.9)	41(1.2)†	263(1.8)*	26(0.9)†	269(2.0)
	1990	40(1.1)†	196(1.7)	40(0.3)†	248(1.7)*†	25(0.7)†	270(2.0)
	1988	38(0.6)	199(2.0)	40(0.3)†	254(1.9)	21(0.7)*	270(4.5)
	1984	36(1.0)	190(4.5)	35(1.4)*	258(3.2)	20(1.5)*	267(3.1)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated scale scores and percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The overall picture of students' performance on the writing trend assessments is one of mixed results. Fourth graders had relatively stable performance across the assessment years. At the eighth grade, a decline observed in 1990 was reversed in 1992. However, the average score decreased since 1992, resulting in a 1994 average score that was not significantly different from that in 1984. The average score of eleventh graders decreased from 1984 to 1994.

Little or no change was observed among fourth graders in the upper, middle two, or lower quartiles. For eighth graders across the performance distribution, a pattern similar to that for the nation was observed: decreases and increases across the assessment years resulted in a 1994 average score that did not differ significantly from the 1984 average. Little or no change was observed for eleventh graders in the upper quartile. Eleventh graders in the middle two and lower quartiles displayed an overall decline in performance. For eleventh graders in the middle two quartiles, this resulted in a lower average score in 1994 than in 1984. In the lowest quartile, while the pattern suggested decline, the difference between 1984 and 1994 was not statistically significant.

White, Black, and Hispanic fourth graders had relatively stable performance in writing across the assessment years. The average score for White eighth graders declined in 1990, increased in 1992, and returned to a level in 1994 that did not differ significantly from that observed in 1984. The average score for eleventh grade White students decreased between 1984 and 1994. Among Black students, the average score of eighth graders declined between 1992 and 1994; however, there was no significant difference between 1994 and 1984 average scores. The performance of eleventh-grade Black students showed no significant changes across the assessment years. At the eighth grade, Hispanic students attained a higher average score in 1992 compared to 1984, followed by a decline in 1994 that resulted in an average score not significantly different from that in 1984. For eleventh-grade Hispanic students, performance increased from 1984 to 1990, but subsequently decreased to a level in 1994 that was not significantly different from that in 1984.

At all three grades in 1994, White students outperformed their Black and Hispanic peers. At grade 4, Black students outperformed Hispanic students. No significant change was observed since 1984 in the gap between White and Black students' writing scores at any grade. This was also observed for the gap between White and Hispanic students in grades 4 and 8. Although there was some indication that the gap between White and Hispanic eleventh graders decreased across the assessments, the magnitude of the gap in 1994 was not significantly different from that in 1984.

Male and female fourth graders' average scores remained relatively stable across the trend assessments. Eighth-grade male and female students had a lower score in 1990 than in 1984; however, their average scores returned in 1994 to a level not significantly different from that in 1984, but still higher than in 1990. An overall decline in performance was observed among male eleventh graders, although the 1994 and 1984 average scores were not significantly different. The average score of female eleventh graders was lower in 1994 than in 1988 and 1990, but there was no statistically significant difference between the first and last assessments.

In 1994, female students continued to outperform their male peers at all grades. Although there was some indication that this gap increased at grade 4, it was not significantly different in 1994 than in 1984. At grades 8 and 11, the gap between males and females remained relatively stable.

Fourth graders in each region of the country displayed few changes in writing performance from 1984 to 1994. The performance of eleventh graders was also relatively stable, except for students in the Southeast, where an overall decline was observed. However, there was no statistically significant difference between 1994 and 1984 average scores. A pattern of overall increased performance was observed for fourth graders from the Central region and for eighth graders from the Northeast and Central regions, although 1984 and 1994 average scores for these students did not differ significantly. Despite some fluctuations, there was no overall increase or decrease in the performance of eighth graders from the Southeast and the West regions.

At the lowest level of parental education, no significant change across the writing assessments was observed for students in each age group. Among fourth graders who reported high school graduation as their parents' highest level of education, an overall increase in performance was observed, despite the lack of a significant difference between 1994 and 1984 average scores. Eighth graders who reported that their parents' highest level of education was high school graduation had an average score in 1992 that was higher than in 1984. A decrease in 1994 resulted in an average score that did not differ significantly from 1984. Eleventh graders who reported the same level of parental education showed no significant change since 1984. At the two highest levels of parental education, fourth graders' performance was relatively stable. The average scores of eleventh graders at the same two levels decreased from 1984 to 1994. For eighth graders at the highest level of education, the average score decreased from 1984 to 1990, and then increased to a level in 1994 that did not differ significantly from that in 1984.

Stable writing performance was evident among fourth graders attending public or nonpublic schools. At the eighth grade, students in public schools had an average score in 1990 that was lower than that in 1984. The average score increased in 1992, but decreased again in 1994 to a level not significantly different from that in 1984. No significant change was observed in the performance of eighth graders attending nonpublic schools. Both public and nonpublic school eleventh graders displayed overall decreases in performance. For eleventh graders in nonpublic schools, the 1994 average score was lower than the 1984 average; in public schools the difference was not statistically significant.

Fourth graders at modal age had no significant changes across the assessment. Those fourth-grade students who were greater than modal age had an overall pattern of increased performance, although 1994 and 1984 average scores were not significantly different. Eighth graders who were at modal age and those who were greater than modal age, displayed some decreases and increases in performance across assessments. For both groups of students, there were no significant differences between 1984 and 1994 average scores. Eleventh graders less than or greater than modal age had relatively stable performance. The average score of eleventh graders at modal age decreased from 1984 to 1994.

11

National Trends in Levels of Writing Performance from 1984 to 1994

To provide a context for interpreting the overall writing trend assessment results presented in Chapter 10, the tasks used in the assessments were *mapped* onto the writing scale.[66] To map an assessment item is to identify the point on the scale at which individuals with that level of performance had a high probability of responding correctly to the item. Because the NAEP writing assessment is a direct measure of students' writing abilities, it does not contain items that can be scored as correct or incorrect. Instead, students' responses to the writing tasks are rated as being at one of four levels of task performance described in a scoring guide. Thus, instead of mapping an individual item at a point on the scale at which students were likely to provide a "correct" answer, the item mapping technique was modified to map the rating levels for each item onto the scale.

[66] This technique is an adaptation of a method developed to report the results of NAEP's 1985 literacy assessment of young adults. For more information see Irwin S. Kirsch and Ann Jungeblut, *Literacy: Profiles of America's Young Adults* (Princeton, NJ: Educational Testing Service, 1986). See the Procedural Appendix for a description of the mapping procedure.

For each of the four possible ratings on individual writing tasks (Unsatisfactory, Minimal, Adequate, or Elaborated), the point on the scale was identified at which individuals were likely to have received the rating or higher. Individuals with scale scores that were at least as high as the one identified had a 65 percent probability of writing a response that received the rating or higher. For example, a fourth-grade persuasive task (Spaceship task) mapped at Level 200 for responses that were rated as Minimal or better. This means that there was a 65 percent probability that students at Level 200 would write Minimal or better responses to the task.

Item mapping is a replacement for the anchoring procedures conducted for the science, mathematics, and reading trend assessments. Because it requires sets of items to define each anchor level, the anchoring procedure was not appropriate for the writing assessment due to its small number of tasks. More detailed information about item mapping and scaling procedures can be found in the Procedural Appendix.

The results of the item mapping for the writing assessment are displayed in Figure 11.1. In this figure, the grade 4, 8, and 11 results are presented separately along the 500-point writing scale. The labels for the various tasks are followed by the term "informative" or "persuasive," indicating the nature of the task. (See the figure on page 159 in the introduction to this section for a description of the ratings for each task, and see Chapter 12 for a discussion of each task.)

Figure 11.1

Difficulty Values Along the Writing Scale for the Different Levels of Performance on the Informative and Persuasive Tasks, Grades 4, 8 and 11, 1994

GRADE 4

- 329 **Appleby House — Informative** Adequate
- 308 **Radio Station — Persuasive** Adequate
- 268 **Spaceship — Persuasive** Adequate
- 257 **XYZ Company — Informative** Adequate
- 250 **Plants — Informative** Adequate
- 241 **XYZ Company – Informative** Minimal
- 233 **Radio Station — Persuasive** Minimal
- 200 **Spaceship — Persuasive** Minimal
- 180 **Appleby House — Informative** Minimal
- 161 **Plants — Informative** Minimal

GRADE 8

- 356 **Recreational Opportunities — Persuasive** Adequate
- 349 **Dissecting Frogs — Persuasive** Adequate
- 343 **Radio Station — Persuasive** Adequate
- 305 **Appleby House — Informative** Adequate
- 289 **Recreational Opportunities — Persuasive** Minimal
- 255 **Radio Station — Persuasive** Minimal
- 249 **XYZ Company — Informative** Adequate
- 233 **XYZ Company — Informative** Minimal
- 227 **Dissecting Frogs — Persuasive** Minimal
- 225 **Food on the Frontier — Informative** Minimal
- 183 **Appleby House — Informative** Minimal

GRADE 11

- 370 **Food on the Frontier — Informative** Adequate
- 364 **Recreational Opportunities — Persuasive** Adequate
- 353 **Space Program — Persuasive** Adequate
- 343 **Bike Lane — Persuasive** Adequate
- 322 **Appleby House — Informative** Adequate
- 285 **Bike Lane — Persuasive** Minimal
- 279 **Job Application — Informative** Adequate
- 275 **Recreational Opportunities — Persuasive** Minimal
- 249 **Space Program — Persuasive** Minimal
- 238 **Job Application — Informative** Minimal
- 232 **Food on the Frontier — Informative** Minimal
- 193 **Appleby House — Informative** Minimal

NOTE: In this graphic illustration, the locations of scale points are necessarily *approximate* for tasks clustered closely together.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 11.1 presents the percentages of students performing at or above five scale levels (150, 200, 250, 300, and 350). These percentages can be used in conjunction with the item mapping information to provide a portrait of the performance of students at each grade. As an example of how to interpret the information presented in Figure 11.1 and Table 11.1, examine the performance of fourth graders at or above Level 250. Across the assessment years, 10 to 15 percent of these students reached or surpassed Level 250. It is evident that fourth graders who reached Level 250 represented some of the best writers at that grade, since only one percent or fewer of their peers reached Level 300. To illustrate the types of writing abilities demonstrated by students who reached at least Level 250, refer to Figure 11.1. For those fourth-grade students estimated to be at 250 on the writing scale (see Figure 11.1, left-hand column), 65 percent provided responses to the Plants task (a persuasive task) that were rated Adequate or better.

The item mapping data reflect probabilities of success based on the performance of students who were at different points on the NAEP writing scale. Probabilities of success on individual writing tasks increase or decrease depending on students' scale scores. For example, 65 percent of the fourth graders with a scale score of 200 provided responses to the Spaceship task that were rated Minimal or better, while even greater percentages of students with higher scale scores did so. Conversely, smaller percentages of students with lower score scales received a similar rating on this task. Although the probability of success on this task was lower for students with lower scale scores, some students in the lower range of the scale did provide Minimal or better responses. Therefore, it is not appropriate to assume that only students with scores at or above 200 could provide Minimal or better responses to this task. It is also important to remember that although most students with a higher score would write responses rated Minimal or better, not all of them would do so.

Table 11.1

Trends in Percentages of Students At or Above Five Writing Performance Levels, 1984 to 1994

		ASSESSMENT YEARS				
Performance Levels	**Grade**	**1984**	**1988**	**1990**	**1992**	**1994**
Level 350	4	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Effective,	8	0(0.1)	0(0.1)	1(0.2)†	2(0.3)*†	1(0.2)
Coherent Writing	11	2(0.7)	1(0.4)*	4(0.7)	2(0.4)	3(0.3)
Level 300	4	1(0.4)	1(0.2)	1(0.1)	1(0.2)	0(0.2)
Complete,	8	13(1.8)	13(0.8)	12(0.8)*	25(1.5)*†	17(1.2)
Sufficient Writing	11	39(2.4)	39(1.7)*	37(1.1)	36(1.9)	33(1.5)
Level 250						
Beginning	4	10(1.0)	15(1.1)†	12(0.9)	13(1.1)	12(0.8)
Focused,	8	72(2.6)	67(1.7)	57(1.5)*†	75(1.4)*	67(1.3)
Clear Writing	11	89(1.0)*	93(1.5)*	84(1.3)†	87(1.3)	85(1.2)†
Level 200	4	54(2.0)	56(2.0)	53(1.7)	58(1.9)	56(2.0)
Incomplete, Vague	8	98(0.9)	97(0.6)	93(0.6)*†	98(0.4)	96(0.6)
Writing	11	100(0.3)	100(0.3)	99(0.3)	100(0.2)	99(0.2)
Level 150	4	93(1.3)	91(0.8)	89(1.1)	93(0.5)	92(0.9)
Disjointed,	8	100(0.0)	100(0.1)	100(0.1)†	100(0.1)	100(0.1)
Unclear Writing	11	100(0.0)	100(0.0)	100(0.1)	100(0.0)	100(0.1)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons.

†Statistically significant difference from 1984, at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Students Writing Task Accomplishment by Performance Levels

Level 350 — Effective, Coherent Writing: Students performing at this level tended to write responses that were rated as Adequate for even the more difficult persuasive tasks. The writing at this level represented clear and complete responses to the assigned task. It tended to contain supportive details and discussion that contributed to the effectiveness of the response. This writing was also characterized by an overall unity and coherence not found at lower levels.

Very few students achieved this level of performance. None of the fourth graders, almost none of the eighth graders, and only 3 percent of eleventh graders reached this level.

Level 300 — Complete, Sufficient Writing: Students performing at this level tended to write responses that were rated as Adequate for some of the tasks. At grade 8, students wrote Adequate responses to one-third of informative tasks (Appleby House) and two-thirds of the persuasive tasks (Dissecting Frogs and Radio Station). Grade 11 students at this level wrote adequate responses to one-third of the informative (Appleby House) and persuasive (Bike Lane) tasks. Responses at this level tended to be complete and to contain sufficient information to accomplish the task.

As shown in Table 11.1, almost none of the students at grade 4 reached this level across the assessments. At grade 8, less than one-fifth (17 percent) of the students reached or surpassed this level in 1994. This represented a decline since 1992, but an increase over the 1990 percentage. The 1994 percentage of eleventh graders at this level (33 percent) is not significantly different from the 1984 percentage, although it is smaller than the percentage in 1988.

Level 250 — Beginning Focused, Clear Writing: Fourth-grade students performing at this level tended to write Adequate responses to two-thirds of the informative tasks (Plants and XYZ Company) and one-half of the persuasive tasks (Spaceship). Eighth graders wrote Minimal responses to two-thirds of the persuasive tasks (Recreation Opportunities and Radio Station). Eleventh graders wrote minimal responses to two-thirds of the persuasive tasks (Bike Lane and Recreation Opportunities) and Adequate responses to one-third of the informative tasks (Job Application).

Adequate responses were more focused and clear than Minimal ones, containing enough development and detail likely to accomplish the assigned task successfully. However, at grades 8 and 11, if the task involved persuading an audience, students at Level 250 still did not provide an argument considered Adequate to convince the intended audience.

The percentage of fourth graders at or above the level increased in 1988, but subsequently decreased to a level in 1994 that did not differ significantly from the 1984 level. Among eighth graders, the percentage of students at or above this level decreased from 1984 to 1990, but returned to a level in 1994 that was not significantly different from that in 1984. Eleventh grade results at Level 250 revealed an overall pattern of decreasing percentages across the assessment years. A smaller percentage of eleventh graders attained at least Level 250 in 1994 than in 1984.

Level 200 — Incomplete, Vague Writing: Students performing at Level 200 tended to write responses that were rated as Minimal to two-thirds of the informative tasks at grades 8 (Food on Frontier and XYZ Company) and 11 (Food on Frontier and Job Application). At grade 4, Minimal responses were likely to be provided by students at this level to one-third of the informative tasks (XYZ Company) and to both persuasive tasks (Spaceship and Radio Station). Responses rated as Minimal, although clearer and more detailed than those rated as Unsatisfactory, still tended to be vague and incomplete. In 1994, nearly all eleventh graders (99 percent) and eighth graders (96 percent) performed at or above this level. For eighth graders, after a decline in 1990, the 1994 percentage returned to a level not significantly different from the 1984 percentage.

Level 150: The results presented in Figure 11.1 indicate that students performing at this level tended to write responses that received unsatisfactory ratings for most of the tasks, except the Appleby House and Plant tasks. Student writing that was rated as Unsatisfactory tended to be too brief and disjointed to be considered a response to the task or was so vague and unclear that it was difficult to understand. At all three grades, students at Level 150 were likely to provide Minimal responses to the Appleby House task. At grade 4, students at this level were likely to provide Minimal responses to the Plant task. No significant change from 1984 to 1994 was observed for any grade at this level of performance.

Trends in Writing Performance Levels by Race/Ethnicity between 1984 and 1994

Table 11.2 shows the percentages of fourth-, eighth-, and eleventh-grade White, Black, and Hispanic students performing at or above each of the performance levels in the first and most recent assessment years (1984 and 1994).

At grade 4, no significant change between 1984 and 1994 was observed in the percentage of White, Black, or Hispanic students at any level. In 1994, a greater percentage of White fourth graders attained at least Levels 150, 200, and 250, compared to their Black and Hispanic peers. Also in 1994, the percentages of Hispanic students reaching Levels 150 and 200 or higher were greater than the percentages of Black students doing so.

Among eighth graders, a small but statistically significant decline between 1984 and 1994 was observed in the percentage of White students attaining at least Level 200. Comparing the 1994 performance of students in the three racial/ethnic groups, a greater percentage of White students than Black or Hispanic students performed at or above Levels 200, 250, and 300.

At the eleventh grade, a smaller percentage of White students reached Level 250 or above in 1994 than in 1984. Among Hispanic eleventh graders, a larger percentage reached at least Level 300 in 1994 than in 1984. In 1994, higher percentages of White students than Black or Hispanic students attained Levels 250 and 300.

Table 11.2

Trends in Percentages of Students At or Above Five Writing Performance Levels by Race/Ethnicity, 1984 to 1994

		ASSESSMENT YEARS					
		1984			1994		
Performance Levels	Grade	White	Black	Hispanic	White	Black	Hispanic
Level 350	4	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Effective, Coherent	8	0(0.2)	0(0.0)	0(0.0)	1(0.3)	0(0.2)	0(0.4)
Writing	11	3(0.9)	0(0.4)	0(0.0)	3(0.5)	0(0.5)	1(1.0)
Level 300	4	1(0.5)	0(0.0)	1(0.9)	1(0.2)	0(0.0)	0(0.2)
Complete,	8	16(2.2)	3(1.7)	4(2.3)	21(1.5)	5(1.6)	9(2.5)
Sufficient Writing	11	46(2.9)	16(4.4)	8(4.3)*	39(2.4)	16(2.6)	20(4.2)
Level 250	4	13(1.1)	3(2.2)	5(2.3)	15(1.0)	2(0.7)	4(1.6)
Beginning Focused,	8	79(2.6)	48(9.2)	47(8.7)	75(1.5)	43(4.1)	52(3.9)
Clear Writing	11	95(1.3)*	76(5.2)	62(9.2)	89(1.0)	71(3.6)	74(4.1)
Level 200	4	62(2.1)	29(6.9)	37(5.8)	67(2.0)	22(3.9)	38(4.2)
Incomplete, Vague	8	99(0.4)*	95(3.4)	93(6.1)	98(0.5)	90(2.6)	92(2.0)
Writing	11	100(0.1)	99(1.0)	97(1.8)	100(0.2)	98(0.7)	97(1.3)
Level 150	4	96(1.1)	81(3.1)	84(6.3)	96(0.7)	75(2.7)	86(2.4)
Disjointed, Unclear	8	100(0.0)	100(0.0)	100(0.0)	100(0.1)	100(0.5)	100(0.3)
Writing	11	100(0.0)	100(0.2)	100(0.0)	100(0.0)	100(0.1)	100(0.0)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Writing Performance Levels by Gender between 1984 and 1994

Table 11.3 shows the percentages of males and females attaining each of the five performance levels in both the 1984 and 1994 assessments. Among fourth graders, no significant changes were observed between 1984 and 1994 in the percentages of male or female students at any of the performance levels. At the eighth grade, a smaller percentage of male students attained at least Level 200 in 1994 than in 1984. Among male and female eleventh graders, there was a decrease between the first and most recent assessments in the percentage who performed at Level 250 or above.

Table 11.3

Trends in Percentages of Students At or Above Five Writing Performance Levels, by Gender, 1984 to 1994

		Assessment Years			
		1984		1994	
Performance Levels	Grade	Male	Female	Male	Female
Level 350 Effective, Coherent Writing	4	0(0.0)	0(0.0)	0(0.0)	0(0.0)
	8	0(0.0)	0(0.3)	0(0.1)	1(0.4)
	11	1(0.4)	3(1.5)	1(0.2)	4(0.6)
Level 300 Complete, Sufficient Writing	4	0(0.7)	1(0.3)	0(0.1)	1(0.3)
	8	7(1.5)	20(3.6)	9(1.1)	25(1.7)
	11	28(1.9)	50(3.8)	24(2.1)	42(2.2)
Level 250 Beginning Focused, Clear Writing	4	9(1.3)	12(2.6)	7(1.1)	16(1.4)
	8	61(3.7)	84(2.6)	56(2.3)	80(1.6)
	11	84(1.6)*	95(1.1)*	79(1.6)	90(1.2)
Level 200 Incomplete, Vague Writing	4	50(4.1)	59(3.0)	47(2.5)	66(2.4)
	8	97(1.2)*	100(0.7)	93(1.1)	99(0.4)
	11	99(0.4)	100(0.2)	98(0.4)	100(0.3)
Level 150 Disjointed, Unclear Writing	4	92(1.8)	94(1.8)	89(1.2)	95(1.0)
	8	100(0.0)	100(0.0)	100(0.2)	100(0.1)
	11	100(0.0)	100(0.0)	100(0.1)	100(0.0)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

The percentages of students performing at the five performance levels identified on the NAEP writing scale have changed little since 1984. Trend analyses revealed overall patterns of declining percentages of eighth graders at or above Level 200 and of eleventh graders at or above Levels 250 and 300. In addition, the percentage of eleventh-grade students at Level 250 in 1994 was lower than the percentage in 1984.

Among White students, a decrease between 1984 and 1994 was observed in the percentage of eleventh graders attaining Level 250 or higher. Hispanic eleventh graders displayed an increase between 1984 and 1994 in the percentage of students reaching at least Level 300. Comparisons of the performance of students in different racial/ethnic groups in 1994 revealed higher percentages of White students than Black or Hispanic students at most performance levels across all three grades. Also, Hispanic fourth graders outperformed Black fourth graders at Levels 150 and 200 in 1994.

Few changes were observed among male and female students. There was a decrease between 1984 and 1994 in the percentage of male eighth graders reaching at least Level 200 on the writing scale. Among eleventh graders, both male and female students displayed a decrease in the percentage performing at or above Level 250. At all three grades, a larger percentage of females than males was at most of the performance levels.

12

Trends in Aspects of Students' Writing

This chapter explores trends in students' responses to the informative, persuasive, and narrative writing tasks they were given in the NAEP trend assessments. In addition to examining trends in students' ability to adhere to the conventions of written English, one task at each grade was selected for a detailed analysis of writing mechanics, including spelling, word choice, punctuation, and syntactic errors. The results of this study also are presented here.

National Trends in *Informative* Writing between 1984 and 1994

Informative writing is used to convey ideas — to inform others about facts, feelings, or procedures. It can involve simple retelling or reporting as well as more complex analyses or generalizations about experiences or knowledge.

We use informative writing when composing a letter, describing a trip we have taken, integrating lecture notes in a written report, and generalizing about lessons we have learned. Informative writing serves many purposes in meeting everyday as well as academic goals and can involve straightforward as well as highly complex thinking.[67]

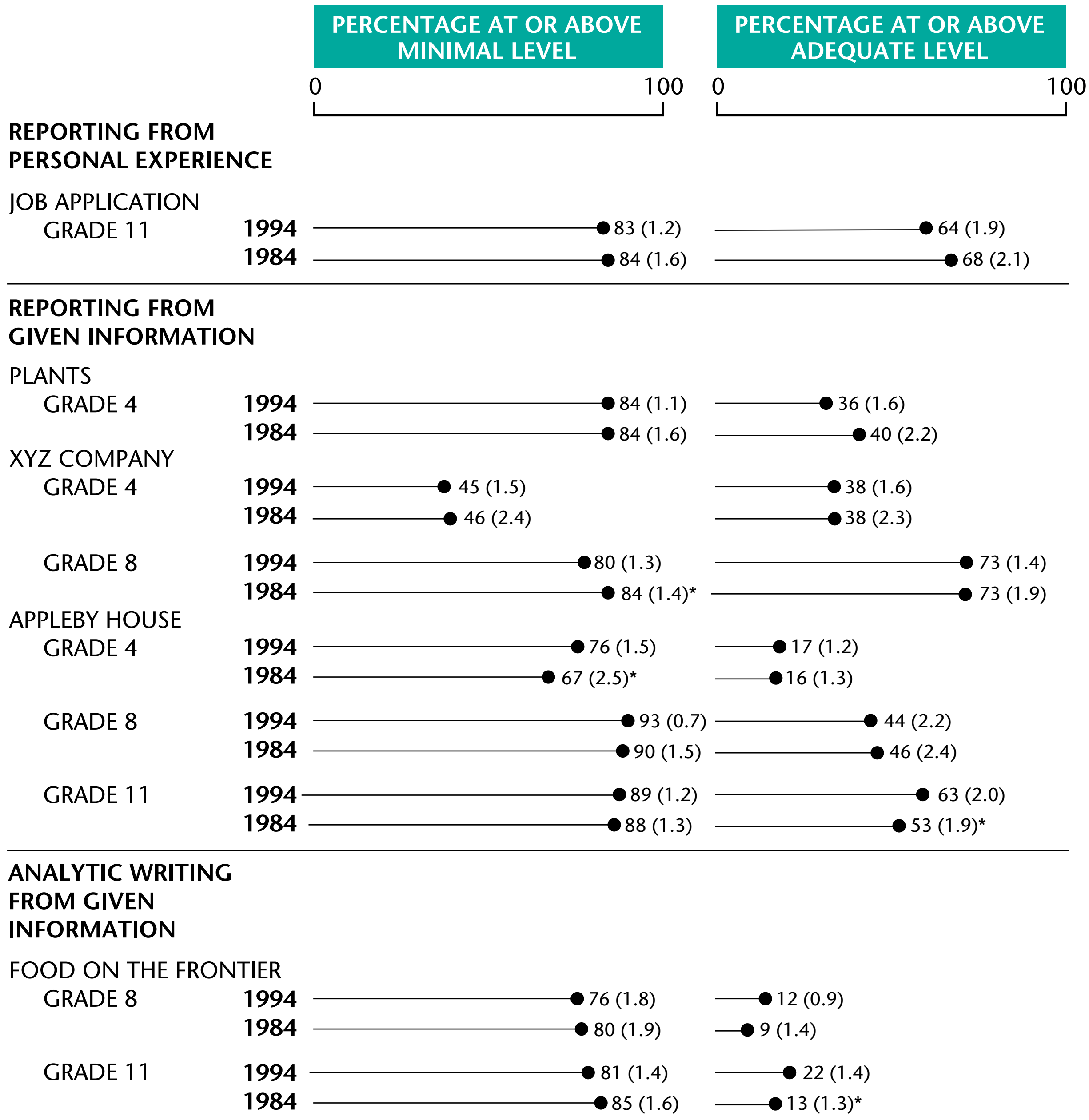
Of the five informative writing tasks included in the trend assessment, one required that students write reports based on their personal experience, three required that they write reports based on given information, and one required that they write a report based on an analysis of given information. Together, these tasks reflect some of the diversity of purposes for which informative writing is undertaken.

Figure 12.1 presents data on the percentages of students who performed at or above Minimal and Adequate levels of accomplishment for each informative task included in the 1984 and 1994 assessments. Discussion of performance at each level and sample papers are provided following this figure.

[67] Britton, J., *Prospect and Retrospect: Selected Essays of James Britton*, Pradl, G. M., editor (Montclair, NJ: Boynton/Cook Publishers, Inc., 1982).

Figure 12.1

Trends in Informative Writing at Grades 4, 8, and 11



*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Reporting from personal experience

"Job Application" required eleventh-grade students to provide a brief description of a desirable job and to summarize their previous experiences or qualifications for it.

There were no significant changes between 1984 and 1994 in the percentage of eleventh graders providing Minimal or better and Adequate or better responses to this task.

Responses that were rated as Minimal provided some details, but created no organizational framework for the reader to use in fitting the parts together. These responses mentioned the kind of job desired, but did not describe relevant hobbies, interests, or past employment. The following student writing sample is typical of the papers rated as Minimal.

I would like to have a part-time job.
Because I need a lot of time for school work
and mostly to do papers and other stuff.
School is my top priority, and the job would
be second. I would like to have most of my
weekend nights off to go out. I would not like
to sit at a desk all day and file papers, I
would like to have something that would keep
me busy all of the time at work so I can
have that time to get things done and I really
like to work. I wouldn't like to do hard work where
I would get bored, or something that is not for
my age level.

Responses judged as Adequate contained some information about the job desired and presented some relevant background information appropriate to the job. The following example is typical of the Adequate responses. In contrast, the most successful papers — rated as Elaborated — provided a full description within a cohesive framework. In these papers, students described the desired job as well as their qualifications and experience. They went beyond the basic elements required in an effort to "sell" themselves. However, in each of the trend writing assessments only 4 percent or fewer of the eleventh-grade responses to this task were rated as Elaborated (see Data Appendix for details).

I would like to work in a restuarant, or a store. I have worked in restuarants before and it was fun. I also think that it would be fun to be a salesperson, because I'm good with people. I want a fun job, because I'm the type of person that does well in a certain thing, when I like what I'm doing and I'd like to do well in my job.

Reporting from given information

"Plants" required fourth-grade students to summarize a science experiment depicted in a series of pictures showing different stages of a plant's growth.

"XYZ Company" required fourth- and eighth-grade students to complete a letter explaining that a previously ordered T-shirt had not been received and proposing a course of action.

"Appleby House" required fourth, eighth, and eleventh graders to write a newspaper article based on notes they were given about an unusual haunted house.

Between 1984 and 1994, there were a few variations in students' performance on the informative tasks that involved reporting from given information. On the "Plants" task, which was administered at only the fourth grade, no significant change in performance was observed between the first and most recent assessment years.

The "XYZ Company" task permitted a comparison across grades (4 and 8) as well as across time. For students in grade 4, performance on this task was relatively stable between 1984 and 1994. At the eighth grade, a smaller percentage of students in 1994 than in 1984 provided responses that were rated at least Minimal.

The "Appleby House" task was given at all three grades. At the fourth grade, an increase was observed between 1984 and 1994 in the percentage of students with responses rated as Minimal or better. However, the percentage of fourth graders with responses rated at least Adequate did not change significantly. At grade 8, the percentages remained relatively stable over time. Among eleventh graders, there was an increase between 1984 and 1994 in the percentage of students with Adequate or better responses.

In the "Appleby House" task, students were asked to reorganize the information provided and weave it into a report that would help the reader understand what the house was like. Minimal responses often simply enumerated the details in the sequence in which they were given without interrelating them. The following example is typical of responses that were rated as Minimal.

the house with no windows, this
is a house with deadend hallways,
36 rooms and stairs leading to the
cieling. doorways go nowhere and all
this to confuse ghosts.

Adequate responses tended to be brief, but presented information about the house in a report format, as illustrated by the following example. In contrast, the most successful reports, those rated as Elaborated, emulated a newspaper article and linked critical details within a cohesive thematic frame in ways that both interested and informed the reader. However, only about 1 percent of eighth-grade responses and 4 percent of eleventh-grade responses in the 1994 assessment were judged to be Elaborated.

Man builds stange house to scare ghosts. He says that he
did it to confuse the ghosts. But why we may ask would he want to
spend 10 years building a house. for instince there are staires
that go nowhere, doors that go nowhere and hallways that
go nowhere. This house has 36 rooms. If you ask me I think
it is kind of stange.

Analytic writing

"Food on the Frontier" required eighth- and eleventh-grade students to read a social studies passage about frontier life and then to explain why modern-day food differs from frontier food.

Analytic writing is qualitatively different from the other kinds of informative writing that students were asked to perform. Reporting from personal experience and from given information involves simple descriptions of what happened or what exists, while analytic writing calls for an explanation of why something happened as it did or how the parts fit together.

No significant change in students' performance on the "Food on the Frontier" task occurred at grade 8. Among eleventh graders, there was an increase between 1984 and 1994 in the percentage of students receiving at least an Adequate rating on this task.

In Minimal responses to this task, such as the following, students tended to present comparisons but did not provide explanations about cause and effect.

Today we have many types
of Foods. We have meat, vegetables
Fruits, + "Junkfood." Back in the
pioneer days They were limited
ON what they could eat. They
couldn't go to the store + buy
twinkies, pop, or chips.
through the centuries business's
methods, skill, + tools have
developed. Today we have so
much "artifical" thing in
our food. So through all this
we were able to expand

Responses judged as Adequate provided some explanation for their comparisons, but were either uneven or sparse in their presentation. The following is typical of such responses. In contrast, the few most successful papers, rated as Elaborated, went beyond the basic elements required, weaving analyses into an organized and elaborated whole. This rating was assigned to less than 1 percent of the responses at grade 8 and about 2 percent of the responses at grade 11.

The difference is that they didn't have as many
different kinds of food as we have today
because a lot of our stuff is imported from
other countries. We buy our food from the
supermarket. They either had to hunt for
their food or grow it. They could only use
the things in their environment.

Holistic Analysis

Eighth- and eleventh-grade responses to “Food on the Frontier” also were scored holistically, as a way of monitoring trends in writing fluency. As Table 12.1 shows, there was no overall improvement from 1984 to 1994 in the relative fluency of students’ responses at either grade 8 or grade 11.

Table 12.1

Trends in Fluency of Informative Writing: Holistic Ratings for “Food on the Frontier” Task, Grades 8 and 11

	PERCENTAGE OF STUDENTS			
	Grade 8		Grade 11	
Holistic Rating	1984	1994	1984	1994
0	6(1.1)	5(0.8)	6(0.8)	4(0.7)
1	3(0.7)	2(0.5)	2(0.6)	3(0.8)
2	21(1.5)	19(1.1)	16(1.9)	15(1.2)
3	36(1.9)	38(1.3)	35(1.9)	36(1.6)
4	23(1.9)	27(1.0)	32(2.2)	26(1.4)
5	9(1.2)	7(0.7)	9(1.2)	12(1.1)
6	2(0.6)	2(0.6)	0(0.2)*	3(0.5)
4, 5 or 6	34(2.2)	36(1.3)	41(2.4)	41(1.7)
Average Rating	3.0(0.1)	3.1(0.0)	3.1(0.1)	3.2(0.1)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

National Trends in *Persuasive* Writing between 1984 and 1994

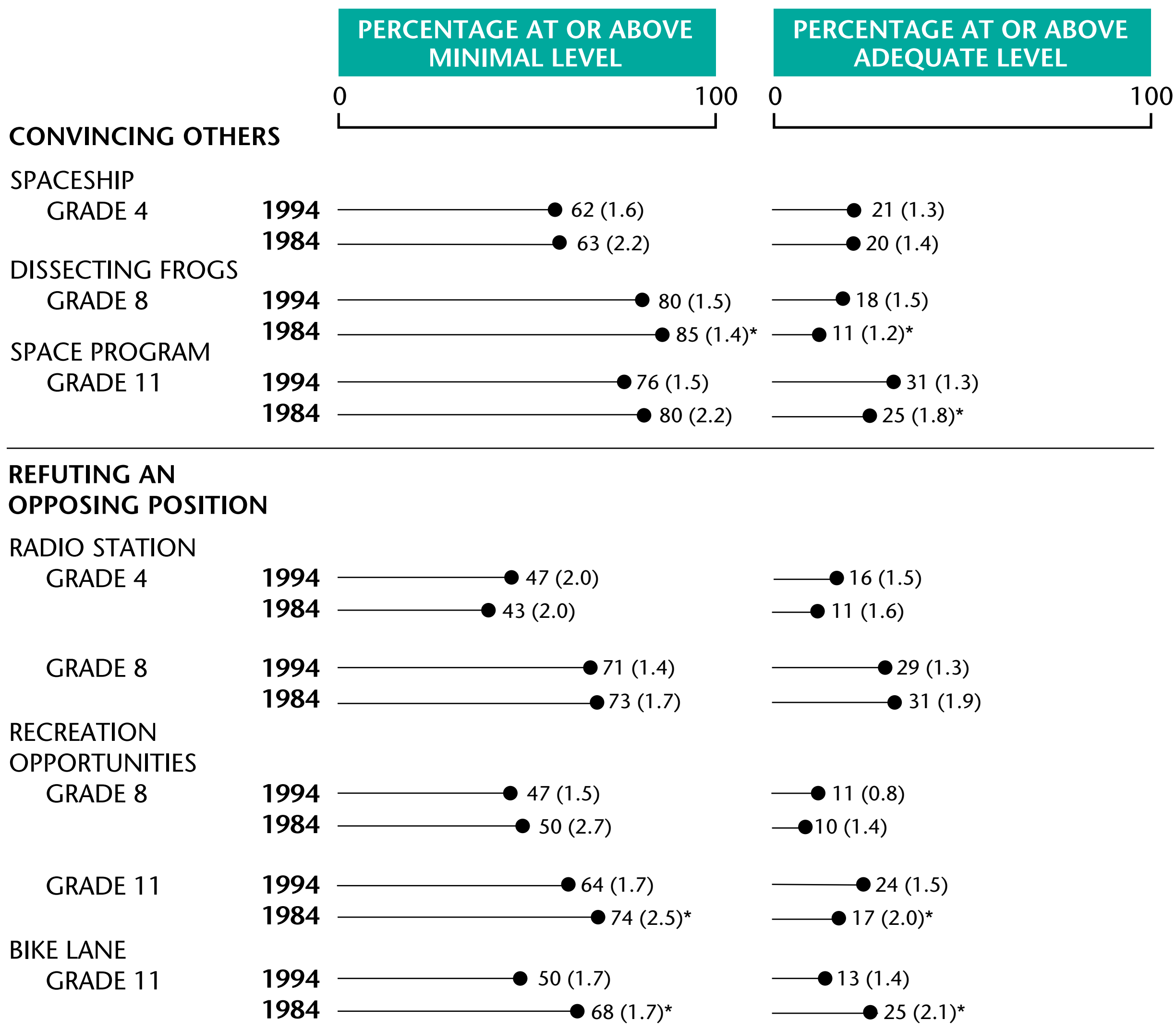
Persuasive writing is primarily intended to influence — to change ideas or actions. It is used to convince others of a point of view or a course of action, to refute their arguments, and to defend certain positions or behaviors. Persuasive writing necessitates awareness of the characteristics of one's audience and of ways to influence it.[68] We use persuasive writing in informal notes when we wish to convince a friend to go to one restaurant rather than another, as well as in formal critical essays when we present a tightly structured argument defending our preferred interpretation of a classical play. In all types of persuasive writing, both formal and informal, the writer must take a point of view and support or defend it.

Of the six persuasive tasks administered, three involved writing to convince others to adopt a particular point of view and the other three involved writing to refute an opposing position. Together these tasks reflect the kinds of writing intended to influence others and bring about change. Figure 12.2 presents information on trends in the percentages of students at each grade who performed at or above the Minimal and Adequate levels for each persuasive task. (It should be noted that students' responses to persuasive tasks are evaluated not on the specific opinion they contain but on their effectiveness in communicating this opinion and in supporting it with evidence or arguments.)

[68]Brewer, W. F., "Literary Theory, Rhetoric, and Stylistics: Implications for Psychology." In *Theoretical Issues in Reading Comprehension,* Spiro, R. J.; Bruce, B. C.; & Brewer, W. F., editors (Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1980).

Figure 12.2

Trends in Persuasive Writing at Grades 4, 8, and 11



*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Writing to convince others

"Spaceship" required fourth graders to form their own points of view about whether creatures from another planet should be allowed to return home or be detained for scientific study, and to support their points of view in ways that would convince others to agree with them.

"Dissecting Frogs" required eighth graders to take a stand on the dissection of frogs in science class, and to discuss and support their views.

"Space Program" required eleventh graders to adopt a point of view about whether or not funding for the space program should be reduced, and to write a letter to their senators explaining their position.

As reported in Figure 12.2, in both 1984 and 1994, less than one-third of the students in each grade provided responses to the "convincing" tasks that were rated Adequate or better. These results indicate that, although they appeared to understand the assignments and present their points of view, students were generally unable to support their ideas.

There were some changes between 1984 and 1994 in eighth and eleventh graders' performance on the "convincing" tasks, but not in fourth graders' performance. At the eighth grade, fewer students in 1994 provided responses to the "Dissecting Frogs" task that were rated as Minimal or better. However, more students wrote Adequate or better responses to this task. At grade 11, a greater percentage of students' responses to the "Space Program" task were rated Adequate or better in 1994 than in 1984. No significant change in eleventh graders' scores was observed at the Minimal or better level on this task.

The following examples of persuasive writing were written by eleventh graders in response to the "Space Program" task. Responses that were rated as Minimal took a point of view, but did not present reasons for the point of view, nor did they provide convincing evidence that would sway a senator's vote. The following is an example of such a paper.

Dear Senator:

I believe we have other problems on this planet which need to be solved first. I do believe money for this space program should be cut. Why do we need permanent colonies in space? It is only useful to those who are astronauts or one's involved with the space. And our money could go for something better to benefit everyone.

Adequate responses supported the point of view presented with some reasoning or examples, as exemplified below. The most successful responses, although rare, provided a well-organized argument with supporting evidence (see the Data Appendix for details).

Dear Senator:

I feel strongly against cuts in funds for the space program. The Space Program is an important part of our future. Space is one of our final frontiers. If money is needed for something, make a cut in the defence program. I believe it's more important to explore space than to be able to blow things away. If we fall behind in space exploration we might miss something vitally important. Lives have been lost in trying to explore space and those lives shouldn't be wasted. Seven people died on the space shuttle in an effort to explore space, and if the program ends, their deaths were for nothing. Please avoid the cut in the space program. Thank you.

Sincerely,
A concerned Citizen

Writing to refute an opposing position

"Radio Station" required fourth and eighth graders to provide reasons why their class should be permitted to visit a local radio station despite the manager's specified concerns.

"Recreation Opportunities" required eighth and eleventh graders to take a stand on whether their town should purchase a railroad track or a warehouse as a recreation center, to defend their choice, and to refute the alternative choice.

"Bike Lane" required eleventh graders to take a stand on whether or not a bike lane should be installed in their locality, and to refute the opposing view.

As shown in Figure 12.2, the patterns of student performance on the "refuting" tasks in 1984 and 1994 were similar to those observed on the "convincing" tasks. As might be expected, far more students wrote responses at or above the Minimal level than at or above the Adequate level — indicating that students were able to take a stand but did not provide sufficient support to refute others' views.

The trend data revealed no significant changes between 1984 and 1994 in fourth and eighth graders' ability to perform the "Radio Station" task. No significant change was seen at grade 8 in the performance of students on the "Recreation Opportunities" task either. At grade 11 on this task, however, fewer students in 1994 than in 1984 provided responses that were rated Minimal or better, and more students provided Adequate or better responses. An overall decrease in eleventh graders' performance on the "Bike Lane" task occurred between 1984 and 1994. There were declines in the percentage of students receiving Minimal or better ratings and in the percentage receiving Adequate or better ratings.

As shown in the following example, Minimal responses to the "Bike Lane" task reflected students' inability to appeal to their audiences. These papers tended to state students' views and sometimes provided elaboration, but did not construct a persuasive argument.

Dear Council Members:

I am writing to support the proposal for the bike lanes. It would be a benefit to society and our community to have these special lanes.

Riding bicycles will reduce the amount of pollution increasing the health and appearance of our city. Although it would decrease the area for parking it would not be a problem because more people would be riding bicycles and would not need a parking space.

Business would increase because the safety of bicycle riding would be guaranteed. Automobile accidents would be less likely to occur.

The majority of people would benefit from this propsal both from saving money on gasoline and getting exercise to improve their health. These are a few reasons I feel the proposal has to be passed.

As illustrated by the following example, responses judged as Adequate took a stand for or against the proposal and also briefly refuted some aspect of the opposing ideas. In comparison, the 2 percent of responses judged as Elaborated went beyond arguing for a particular point of view to present an interrelated set of reasons to support students' positions; they also responded to the explicit concerns of their opponents.

Dear Council Members:

I do not feel it is necessary to create bicycle lanes on major streets. People need places to park. Even though some could ride a bike, many could not. Many people do not have bikes, cannot ride bikes or are dissabled. Plus bike riding is seasonal. You can't ride a bike through the winter.

It may be safer and easier, but many bikers like to ride with the traffic. Those that don't can ride through the park where there is already a lane or on the sidewalks.

Instead of taking away parking altogether, maybe we should consider expanding our streets just a little to make room for bikers. These lanes need not be the size of the regular lanes but merely a couple feet wide.

Although the workers can ride their bikes, many customers cannot. These businesses are for the customers and we need to be aware of that. We cannot inconvenience 95% of the people who drive or have to drive for the 5% that ride their bikes. Rather we should make a little extra for this small population.

Sincerely,

Holistic Analyses

The responses of fourth graders to the "Spaceship" task and of eighth- and eleventh-grade students to the "Recreation Opportunities" task were analyzed holistically. As Table 12.2 indicates, there was a decline between 1984 and 1994 in the percentage of fourth-grade responses that received a fluency rating of 1, and an increase in the percentage that received a fluency rating of 4 or 5. These changes resulted in an overall increase in the percentage of fourth graders whose responses received a fluency rating of 4 or higher.

The fluency of eighth graders' responses also showed improvement between 1984 and 1994 on the "Recreation Opportunities" task. A higher percentage of responses received a rating of 4, and an overall increase was observed in the percentage of responses receiving a rating of 4 or higher. Among eleventh graders, a somewhat different trend was observed in the fluency ratings on this task. There was an increase in the percentage of eleventh-grade responses receiving a fluency rating of 3, and a decline in the percentage receiving a fluency rating of 4. There was also a decline in the percentage of eleventh graders receiving a 4 or higher rating.

Table 12.2

Trends in Fluency of Persuasive Writing: Holistic Ratings for "Spaceship" and "Recreation Opportunities" Tasks

	PERCENTAGE OF STUDENTS					
	Spaceship		Recreation Opportunities			
	Grade 4		Grade 8		Grade 11	
Holistic Rating	1984	1994	1984	1994	1984	1994
0	5(0.8)	7(0.8)	6(1.2)	5(0.9)	3(0.8)	5(0.8)
1	13(1.7)*	7(1.0)	8(0.9)	5(0.8)	2(0.6)	2(0.5)
2	26(1.9)	24(1.3)	14(1.4)	13(1.0)	10(1.4)	11(0.9)
3	45(1.7)	42(1.7)	48(3.0)	44(1.5)	32(1.8)*	41(1.6)
4	11(1.4)*	18(1.3)	21(2.0)*	28(1.1)	43(2.0)*	34(2.1)
5	1(0.4)*	2(0.5)	3(0.8)	5(0.7)	9(1.6)	6(0.9)
6	0(0.0)	0(0.1)	0(0.1)	1(0.2)	2(0.6)	2(0.4)
4, 5, or 6	11(1.4)*	20(1.4)	24(2.2)*	33(1.2)	54(2.0)*	41(2.0)
Average Rating	2.5(0.0)*	2.6(0.1)	2.8(0.1)*	3.0(0.0)	3.4(0.1)*	3.2(0.1)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

National Trends in *Narrative* Writing from 1984 to 1994

Narrative writing allows us to participate in literary experiences. Whether writing personal stories of pain and triumph or fictional tales of interplanetary visits, the goal is to create a momentary reality that is apart from the everyday.[69] As with the other types of writing, narrative writing can be more or less formal, academic, or complex.

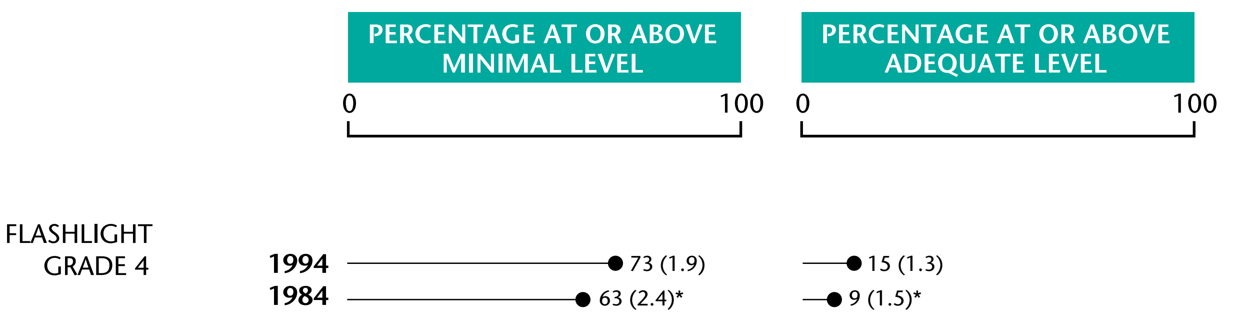
The following narrative writing task was presented at grade 4 in the 1984 and 1994 writing trend assessments.

> *"Flashlight" required fourth graders to write a story about their imagined adventures with a flashlight that has special powers.*

The percentages of students who wrote responses that were judged Minimal or better and Adequate or better in 1984 and 1994 are provided in Figure 12.3.

Figure 12.3

Trends in Narrative Writing at Grade 4



*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

[69] Moffett, J., *Teaching the Universe of Discourse* (Boston, MA: Houghton Mifflin, Co., 1968).

In both the 1984 and 1994 assessments, fourth graders found it difficult to write well-developed stories. In 1984, 63 percent of these students seemed to grasp the basic elements of storytelling, but only 9 percent were able to develop stories successfully. The 1994 data indicate some progress in story writing: 73 percent were able to write stories that were rated as Minimal or better and 14 percent wrote stories that were rated as Adequate or better.

Students providing responses at the Minimal level seemed to understand the narrative character of the "Flashlight" task, but were unable to carry it out. At this score level, students attempted a story, but provided only a bare outline with little detail. Sometimes they rambled or offered lists of details or events, with no point or structure. The following is an example of a response rated as Minimal.

One day I was in my room,
When my mom came in. There
was a praesant in her hand,
It was for me! Oboy, it
was heavey. I opened it,
It was a flashlight!
But It was powerful. I
held on to it, But all uv
a sudden it picked me
up me and my flashlight
went out the window. I
was 500 stories up in
the air. I was very, very
dizzy. We started to come
down and down.

The following is an example of an Adequate response. This type of response reflected the storyteller's obligation to develop a plot and expand on it with details including events, characters, and setting. However, the plots were not as explicit nor as clearly developed as in the stories that were judged to be Elaborated (less than 1 percent in 1994).

I took the Flash light and
waited till dawn then I asked
my mom if I could go camping
with my friends. We set out
that night and found a camp spot
We went exploring in the woods.
We heard somthing coming. It sound-
ed like something like a dog.
But is was a Wolf! I took the
flashligh I found to get a better
look but when saw it It was
turned to stone. So we had no
light to get back because if I
aicendly Shined it on one of my
friens they would turn to stone
to. So that the story about
"The
Magic
Flashlight"

Holistic Analyses

As shown in Table 12.3, the relative fluency of fourth graders' written responses to the "Flashlight" task remained essentially the same in 1994 as in 1984.

Table 12.3

Trends in Fluency of Narrative Writing: Holistic Ratings for "Flashlight" Task, Grade 4

	PERCENTAGE OF STUDENTS	
	Grade 4	
Holistic Rating	1984	1994
0	6(1.1)	5(0.9)
1	7(1.1)	7(1.2)
2	25(2.2)	23(1.8)
3	38(1.7)	43(2.5)
4	19(1.6)	19(1.5)
5	4(1.1)	4(0.8)
6	1(0.4)	1(0.3)
4, 5 or 6	24(2.1)	23(1.5)
Average Rating	2.7(0.1)	2.8(0.1)

*Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

National Trends in Grammar, Punctuation, and Spelling from 1984 to 1994

To examine students' abilities to adhere to the conventions of written English, one task at each grade was selected for further analysis. The tasks chosen were "Spaceship" at grade 4 and "Recreation Opportunities" at grades 8 and 11. Nationally representative subsamples of papers were drawn from the total national sample to permit a detailed analysis of writing mechanics. In addition to measures of overall quality, each paper was analyzed for a variety of aspects of spelling, word choice, punctuation, and syntax.

Trends in Overall Characteristics of Students' Writing

As students gain control of written English, they should be able to use a larger number of words in a growing number of sentences, with relatively greater ease and fewer errors.[70] Table 12.4 summarizes trends in the general characteristics of students' writing at each grade. Since the fourth-grade data are based on a different writing task, comparisons of the results for grade 4 to those for grades 8 and 11 are not appropriate. Table 12.4 also includes the average results for male and female, and Black and White students.

Please note that the sampling procedures for the mechanics analyses yielded an adequate number of cases to enable reporting data for Black as well as White students, and for males and females, but not for other subpopulations.

[70] Pringle, I. & Freedman, A., *A Comparative Study of Writing Abilities in Two Modes at the Grade 5, 8 and 12 Levels* (Toronto, Ontario: The Minister of Education, Ontario, 1985).

Table 12.4

Trends in Overall Characteristics of Papers for the Nation and Demographic Subpopulations, 1984 and 1994

	Grade	Year	Overall Average	RACE/ETHNICITY White	RACE/ETHNICITY Black	GENDER Male	GENDER Female
Number of Words	4	1994	32(0.7)	32(1.0)	29(1.3)	29(1.2)	35(1.2)
		1984	34(1.0)	34(1.2)	32(2.6)	30(1.2)	38(1.4)
	8	1994	71(1.8)	72(2.5)	71(2.9)	63(2.2)	80(2.7)
		1984	68(1.9)	70(2.1)	58(4.2)	60(2.3)	76(2.8)
	11	1994	106(3.0)	111(4.0)	90(3.8)	98(4.9)	113(3.4)
		1984	93(2.3)*	97(3.0)*	81(3.8)	81(2.5)*	106(3.6)
Word Length	4	1994	4(0.0)	4(0.0)	4(0.1)	4(0.0)	4(0.0)
		1984	4(0.0)	4(0.0)	4(0.1)	4(0.0)	4(0.0)
	8	1994	4(0.0)	4(0.0)	4(0.1)	4(0.0)	4(0.0)
		1984	4(0.0)	4(0.0)	4(0.0)	4(0.0)	4(0.0)
	11	1994	4(0.0)	4(0.0)	4(0.0)	4(0.0)	4(0.0)
		1984	4(0.0)	4(0.0)	4(0.0)	4(0.0)	4(0.0)
Number of Sentences	4	1994	2(0.1)	2(0.1)	2(0.1)	2(0.1)	3(0.1)
		1984	3(0.1)*	3(0.1)	2(0.3)	2(0.1)	3(0.2)
	8	1994	5(0.1)	5(0.2)	4(0.3)	4(0.2)	5(0.2)
		1984	4(0.1)	5(0.2)	4(0.3)	4(0.2)	5(0.2)
	11	1994	7(0.2)	7(0.2)	5(0.2)	6(0.3)	7(0.3)
		1984	6(0.2)	6(0.2)*	4(0.2)*	5(0.2)*	6(0.2)
Number of Words Per Sentence	4	1994	17(0.6)	16(0.8)	16(0.9)	17(0.8)	17(0.8)
		1984	15(0.4)	15(0.5)	16(0.6)	15(0.6)	15(0.5)
	8	1994	17(0.3)	17(0.4)	20(1.5)	18(0.6)	17(0.5)
		1984	17(0.4)	17(0.3)	19(1.2)	18(0.7)	16(0.4)
	11	1994	18(0.5)	17(0.5)	19(0.8)	19(1.1)	17(0.3)
		1984	18(0.4)	18(0.5)	21(0.7)	19(0.8)	17(0.5)
Number of Errors	4	1994	5(0.2)	5(0.2)	6(0.6)	5(0.3)	5(0.3)
		1984	5(0.2)	4(0.2)	6(0.5)	4(0.3)	5(0.3)
	8	1994	7(0.3)	7(0.3)	9(0.5)	7(0.3)	8(0.4)
		1984	6(0.2)*	6(0.2)*	6(0.5)*	5(0.2)*	6(0.3)*
	11	1994	8(0.3)	8(0.4)	9(0.7)	8(0.4)	8(0.4)
		1984	6(0.2)*	6(0.2)*	6(0.5)*	6(0.2)*	6(0.3)*
Number of Errors Per 100 Words	4	1994	18(0.6)	17(0.8)	21(1.5)	20(1.1)	17(0.8)
		1984	16(0.6)*	14(0.7)*	20(1.2)	17(1.0)	14(0.8)
	8	1994	11(0.3)	10(0.4)	14(0.5)	12(0.5)	10(0.4)
		1984	9(0.3)*	8(0.4)*	13(1.0)	10(0.5)	8(0.4)*
	11	1994	8(0.3)	8(0.3)	11(0.6)	9(0.5)	8(0.3)
		1984	7(0.2)*	6(0.2)*	8(0.4)*	8(0.3)*	6(0.3)*

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated values appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

At grades 4 and 8, students' written responses to the assessment tasks were about the same length (number of words) as those written in 1984. At grade 11, the 1994 responses were longer than those of 1984. The average length of words used by students (an index of vocabulary) and the number of words per sentence (an index of sentence complexity) showed no significant changes between 1984 and 1994 at any of the three grades. The average number of sentences declined slightly among fourth graders and rose slightly among eleventh graders. There was no statistically significant change observed among eighth and eleventh graders in number of sentences used. The average number of errors in students' responses increased between 1984 and 1994 at grades 8 and 11, but did not change at grade 4. The average number of errors per 100 words (error rate) increased at all three grades. Because students' responses represent first-draft writing, it is reasonable to expect some errors in their responses. It may be that the error rates displayed in Table 12.4 would be lower if students were given more time to look for and correct their errors.

An examination of trends in overall characteristics of students' responses by gender and by race/ethnicity revealed that between 1984 and 1994 there was an increase in the average number of words used by White eleventh graders and by male eleventh graders. Also, there was an increase in the number of sentences in papers written by White, Black, and male eleventh graders. At grades 8 and 11, the total number of errors increased between 1984 and 1994 for White and Black students, and for both males and females. The error rate (number of errors per 100 words) was higher in 1994 than in 1984 for White students at all three grades, for females in grades 8 and 11, and for males and Black students in grade 11.

Trends in Control of Sentence Structure

Students' control of syntax is reflected in the types of errors found in the sentences they create. To examine changes across time in students' command of sentence structure, four types of sentence errors — run-ons, fragments, awkward sentences, and sentences with agreement errors — were marked in the 1984 and 1994 student responses. It is important to remember that students' responses to the assessment tasks are first draft writing. There is little time within the assessment period for students to proofread their work. Therefore, a certain number of sentence-level errors, as well as other types of errors, can be expected.

Table 12.5 presents the average percentage of sentence-level errors per student for grades 4, 8, and 11. For all three grades, and for all racial/ethnic and gender groups, there were no significant changes across time in the percentage of run-on sentences in students' responses. The percentage of sentence fragments in students' responses increased between 1984 and 1994 at the fourth and eighth grades. This type of error also increased at the fourth grade for Black students and at the fourth and eighth grades for females. At all three grades, there was an increase between 1984 and 1994 in the percentage of sentences with agreement errors in students' responses. This type of error also increased for White students at all three grades, for Black students at grades 8 and 11, for males at grades 4 and 8, and for females at grades 8 and 11. The percentage of awkward sentences in students' responses decreased at the eleventh grade, however. This decrease was observed for White students, males, and females in the eleventh grade.

Table 12.5

Trends in Sentence-Level Errors for the Nation and Demographic Subpopulations, 1984 to 1994

	Grade	Year	Overall Average	RACE/ETHNICITY White	RACE/ETHNICITY Black	GENDER Male	GENDER Female
Percentage Run-on Sentences	4	1994	15(1.2)	14(1.8)	19(3.5)	16(1.9)	14(1.6)
		1984	15(1.5)	15(1.5)	11(2.5)	14(2.0)	17(2.2)
	8	1994	8(1.0)	7(1.1)	15(3.1)	11(1.9)	4(0.9)
		1984	7(0.9)	6(0.9)	8(2.4)	8(1.4)	6(1.0)
	11	1994	5(0.9)	4(0.9)	7(2.1)	7(1.9)	4(0.7)
		1984	5(0.7)	5(0.8)	5(1.6)	5(1.1)	4(0.9)
Percentage Sentence Fragments	4	1994	6(0.8)	6(0.9)	11(2.4)	6(1.2)	7(1.1)
		1984	3(0.5)*	3(0.6)	4(1.4)*	3(0.8)	3(0.6)*
	8	1994	6(0.7)	5(0.9)	8(1.1)	5(0.8)	7(1.0)
		1984	3(0.5)*	3(0.5)	5(1.4)	4(0.9)	3(0.5)*
	11	1994	4(0.6)	3(0.7)	8(1.7)	5(1.2)	4(0.5)
		1984	3(0.4)	2(0.4)	5(1.1)	4(0.8)	2(0.5)
Percentage Sentences with Agreement Errors	4	1994	10(1.3)	8(1.7)	11(2.0)	11(1.6)	9(1.6)
		1984	4(0.7)*	3(0.7)*	8(2.0)	3(0.7)*	4(1.1)
	8	1994	11(1.2)	8(1.2)	21(3.7)	12(1.7)	10(1.2)
		1984	3(0.6)*	3(0.7)*	3(1.3)*	3(0.8)*	3(0.9)*
	11	1994	8(0.9)	7(0.9)	10(2.0)	8(1.6)	8(1.0)
		1984	3(0.5)*	3(0.7)*	3(0.8)*	4(1.0)	1(0.3)*
Percentage Awkward Sentences	4	1994	26(2.2)	21(2.6)	34(4.1)	28(3.3)	24(3.1)
		1984	25(2.2)	20(2.1)	45(5.5)	26(2.6)	25(2.7)
	8	1994	28(1.2)	25(1.5)	37(3.3)	30(1.7)	26(1.7)
		1984	32(1.5)	28(1.7)	50(4.9)	34(2.5)	30(1.8)
	11	1994	20(0.8)	18(1.0)	28(2.1)	20(1.3)	20(1.5)
		1984	31(1.7)*	28(1.7)*	39(5.2)	35(2.4)*	27(2.0)*

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The average percentage of errors per student should be considered in light of the percentile distribution of errors presented in the Data Appendix. For example, the results by percentile in 1994 revealed that three-quarters of the responses at grade 4 contained virtually no run-on sentences, despite the fact that the average percentage of run-on sentences for grade 4 in 1994 was 15 percent. It is important to remember that if a student's response has only two or three sentences, one of which is a run-on sentence, the percentage of sentences with this type of error will appear high — although only one error has been made.

Trends in Control of Word-Level Conventions

Students' control of word-level conventions is reflected in their spelling, capitalization, and word choice errors, which are summarized in Table 12.6. At grades 4 and 8, there were no significant changes between 1984 and 1994 in the percentage of misspelled words in students' responses. At grade 11, there was a small increase in misspelled words. The percentages of word choice errors and capitalization errors in 1994 and 1984 were extremely low for students in each grade (2 percent or less).

Table 12.6

Trends in Word-Level Errors for the Nation and Demographic Subpopulations, 1984 to 1994

				RACE/ETHNICITY		GENDER	
		Year	Overall Average	White	Black	Male	Female
Percentage Misspelled Words	4	1994	9(0.4)	8(0.5)	9(1.1)	10(0.6)	7(0.6)
		1984	8(0.4)	8(0.6)	10(1.0)	9(0.7)	7(0.5)
	8	1994	4(0.2)	4(0.3)	4(0.4)	4(0.3)	4(0.3)
		1984	4(0.2)	4(0.2)	4(0.5)	4(0.4)	3(0.2)
	11	1994	3(0.1)	3(0.2)	3(0.2)	3(0.2)	2(0.2)
		1984	2(0.1)*	2(0.1)	2(0.2)	3(0.2)	2(0.2)
Percentage Word Choice Errors	4	1994	2(0.1)	1(0.2)	2(0.3)	2(0.2)	2(0.2)
		1984	1(0.1)*	1(0.1)*	2(0.4)	1(0.1)*	1(0.1)*
	8	1994	1(0.1)	1(0.1)	2(0.3)	1(0.1)	1(0.1)
		1984	1(0.1)	1(0.1)*	1(0.4)	1(0.1)	1(0.1)
	11	1994	1(0.1)	1(0.1)	1(0.2)	1(0.1)	1(0.1)
		1984	1(0.1)	1(0.1)	1(0.2)	1(0.1)	1(0.1)
Percentage Capitalization Errors	4	1994	2(0.3)	2(0.3)	2(0.3)	2(0.4)	2(0.3)
		1984	1(0.1)*	1(0.1)*	1(0.2)*	1(0.1)*	1(0.2)*
	8	1994	1(0.1)	1(0.1)	1(0.2)	1(0.2)	1(0.1)
		1984	0(0.1)*	0(0.1)*	1(0.2)*	0(0.1)*	0(0.1)*
	11	1994	1(0.1)	1(0.2)	2(0.3)	2(0.3)	1(0.1)
		1984	0(0.0)*	0(0.0)*	0(0.1)*	0(0.1)*	0(0.0)*

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Trends in Control of Punctuation

Punctuation was analyzed in terms of both the particular marks that students used correctly or incorrectly and the marks that should have been used when punctuation was omitted. Trends in punctuation uses and omissions are summarized in Table 12.7. The average number of total punctuation errors per 100 words found in 1994 responses was comparable to the number found in 1984 responses. In general, most students made few punctuation errors — either omitted punctuation or incorrect use of punctuation. In 1994 and 1984, students in each grade made 2 or fewer of either type of punctuation error in their responses.

Table 12.7

Trends in Punctuation Errors for the Nation and Demographic Subpopulations, 1984 to 1994

				RACE/ETHNICITY		GENDER	
	Grade	Year	Overall Average	White	Black	Male	Female
Total Punctuation Errors Per 100 Words	4	1994	2(0.2)	2(0.3)	2(0.3)	2(0.3)	2(0.2)
		1984	3(0.2)	2(0.3)	3(0.5)	3(0.3)	3(0.3)
	8	1994	2(0.1)	2(0.1)	2(0.3)	2(0.2)	2(0.1)
		1984	2(0.1)	2(0.1)	3(0.4)	2(0.2)	3(0.2)
	11	1994	2(0.1)	2(0.1)	2(0.2)	2(0.1)	2(0.1)
		1984	2(0.1)	2(0.2)	2(0.1)	2(0.2)	2(0.2)
Punctuation Omitted Per 100 Words	4	1994	2(0.2)	2(0.2)	2(0.3)	2(0.3)	2(0.2)
		1984	2(0.2)	2(0.2)	3(0.4)	2(0.3)	2(0.3)
	8	1994	2(0.1)	2(0.1)	2(0.3)	2(0.2)	2(0.1)
		1984	1(0.1)*	1(0.1)*	2(0.4)	1(0.1)*	1(0.1)*
	11	1994	1(0.1)	1(0.1)	2(0.2)	1(0.1)	2(0.1)
		1984	1(0.1)	1(0.1)	2(0.1)	1(0.2)	1(0.2)
Wrong Punctuation Per 100 Words	4	1994	0(0.1)	0(0.1)	0(0.1)	0(0.1)	0(0.1)
		1984	0(0.1)	0(0.1)	1(0.3)	0(0.1)	0(0.1)
	8	1994	0(0.0)	0(0.0)	0(0.1)	0(0.0)	0(0.1)
		1984	1(0.1)*	1(0.1)	0(0.1)	1(0.1)*	0(0.1)
	11	1994	0(0.0)	0(0.1)	0(0.1)	0(0.1)	0(0.1)
		1984	0(0.0)	0(0.0)	1(0.1)	0(0.1)	0(0.1)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

Trends in Informative Writing

Trends in students' responses to the five informative writing tasks reveal few changes across time. Between 1984 and 1994, there was no change in eleventh graders' performance on an informative task in which they were asked to report from personal experience. On one of the tasks that required students to report from given information, a larger percentage of fourth-grade responses in 1994 than in 1984 was rated as Minimal or above. A decrease in the percentage of eighth-grade responses rated as Minimal or better was observed on a similar task. The performance of eighth-grade students on the task requiring analytic writing did not change significantly between 1984 and 1994. For eleventh graders, there was an increase in the percentage of students with responses rated as Adequate or better.

A comparison between 1984 and 1994 of the relative fluency of students' analytic writing, based on holistic scoring procedures, revealed little difference between the two years.

Trends in Persuasive Writing

Mixed results were observed in the comparison of students' performance in 1984 and 1994 on the persuasive tasks that required students to convince others to adopt a point of view or to refute an opposing view. At grade 8, improvement was observed in students' performance on one task that involved convincing others. Compared to 1984, a larger percentage of eighth graders in 1994 wrote responses for this task that were rated at or above the Adequate level. At grade 11, there was an increase in the percentage of responses to a similar task that were rated as Adequate or better. An increase was also observed between 1984 and 1994 in the percentage of eleventh graders' responses at or above the Adequate level on a task in which they were asked to refute an opposing position. On another such task, however, the percentage of Adequate or better responses decreased.

Comparing the 1984 and 1994 relative fluency of students' responses to two of the persuasive tasks revealed some differences. On one task administered at grade 4, overall increases in students' writing fluency were observed between the first and most recent assessment years. On a second task administered at both grades 8 and 11, an increase in overall fluency was observed in the responses written by eighth graders, but a decline in overall fluency was observed at the eleventh grade.

Trends in Narrative Writing

On one narrative task administered at grade 4, overall improvement in students' performance was observed between 1984 and 1994. There was an increase in the percentage of students who wrote responses that were rated at or above the Minimal and Adequate levels. However, the comparison of relative fluency on this task between 1984 and 1994 revealed no significant change.

Trends in Grammar, Punctuation, and Spelling

Detailed analyses of the performance of fourth, eighth, and eleventh graders suggest that there have been some changes in students' mastery of English language conventions between 1984 and 1994. Although the number of words used in eleventh-grade students' responses has increased since 1984, there also have been increases in the rate of errors (number of errors per 100 words) at all three grades. Also, there were increases at all three grades in the percentage of sentences with agreement errors in students' responses. However, a decline in the percentage of awkward sentences in eleventh graders' responses was observed between the first and last assessment years. Generally, the errors that were most frequent for a particular group of students or at a particular grade were found in the responses written by only a small proportion of those students.

13

Trends in Attitudes, Writing Behaviors, and Instruction

In addition to responding to various writing tasks, students participating in the 1984 and 1994 writing assessments were asked to answer a series of questions related to their attitudes toward writing, their abilities to manage the writing process, and their instructional environments. This chapter summarizes trends in students' responses to these questions.

Learning to Value Writing

One of the key goals of writing instruction is to encourage students to see writing as a useful tool for their lives both in and out of school, and to think of themselves as writers.[71] As a way of measuring trends in students' attitudes towards writing, several sets of questions in the 1984 and 1994

[71] Odell, L. & Goswami, D., "Writing in a Nonacademic Setting." In R. Beach and L. S. Bridwell, *New Directions in Composition Research* (New York, NY: The Guilford Press, 1984).

assessments asked students about the value they placed on writing and the kinds of writing they encountered in their home environments.

The set of questions focusing on the value placed on writing was different at grade 4 from that used at grades 8 and 11. Fourth graders were asked to what extent they agreed with statements such as "Writing helps me share my ideas" and "Writing helps me show that I know things." Their responses are presented in Table 13.1.

Compared to 1984, there was an increase in 1994 in the percentage of fourth graders who thought the following statements were true more than half the time: "Writing helps me share my ideas" and "Writing helps me get a good job."

Table 13.1

Trends in the Value Placed on Writing, Grade 4, 1984 to 1994

		HOW OFTEN FOURTH GRADERS FIND STATEMENTS TO BE TRUE							
		More Than Half the Time		About Half the Time		Once in a While		Never or Hardly Ever	
Writing helps me . . .	Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Share my ideas	1994	67(2.3)	206(3.0)	13(1.5)	199(6.3)	13(1.5)	202(6.7)	8(0.9)	***(***)
	1984	53(4.2)*	203(6.0)	20(2.7)*	***(***)	16(3.1)	***(***)	12(2.2)	***(***)
Show I know things	1994	68(2.4)	206(2.4)	15(1.6)	201(6.3)	10(1.2)	198(7.4)	7(1.1)	***(***)
	1984	63(3.9)	203(3.5)	14(2.9)	***(***)	14(3.2)	***(***)	10(2.6)	***(***)
Keep in touch with friends	1994	62(1.9)	206(3.2)	10(0.9)	206(6.1)	14(1.2)	203(6.2)	14(1.2)	194(4.7)
	1984	65(3.3)	204(5.0)	11(1.8)	***(***)	10(1.9)	***(***)	14(3.2)	***(***)
Get a good job	1994	56(2.3)	204(2.9)	13(1.4)	203(6.6)	21(2.1)	202(5.6)	11(1.7)	205(7.3)
	1984	34(3.2)*	198(6.9)	20(2.9)*	***(***)	28(2.8)*	210(5.5)	18(3.2)*	***(***)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Eighth and eleventh graders were asked to what extent they agreed with statements such as "Writing helps me think more clearly," "Writing helps me tell others what I think," and "People who write well have a better chance of getting good jobs." Their responses are presented in Table 13.2.

There were few changes over time in students' responses to these questions. A decrease between 1984 and 1994 was observed in the percentage of eighth graders indicating that they found the statement, "Writing helps me tell others what I think," to be true less than half the time. For eleventh graders, there was an increase between 1984 and 1994 in the percentage of students indicating that they found the statement, "Writing helps me understand my own feelings," to be true more than half the time.

The results presented in Table 13.2 also indicate that students in 1994 who found many of the statements to be true more than half the time had higher average writing scores than students who never or hardly ever found the statements to be true. The following statements demonstrated this pattern at both grades 8 and 11: "Writing helps me think more clearly," "Writing helps me tell others what I think," and "Writing helps me understand my own feelings." The same pattern was also evident at grade 8 for the following statement: "Writing helps me tell others how I feel."

Table 13.2

Trends in the Value Placed on Writing, Grades 8 and 11, 1984 to 1994

			HOW OFTEN STUDENTS FIND STATEMENTS TO BE TRUE							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
	Grade	Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Writing helps me think more clearly	8	1994	42(1.1)	267(2.7)	29(1.5)	266(2.6)	16(1.1)	265(3.6)	13(1.0)	252(4.0)
		1984	44(2.0)	268(2.8)	27(2.0)	267(3.6)	17(1.9)	268(4.0)	12(1.7)	264(5.8)
	11	1994	54(1.2)	288(1.4)	25(1.2)	285(2.2)	14(0.9)	280(3.4)	7(0.8)	270(5.2)
		1984	52(2.5)	289(2.8)	26(1.9)	288(3.7)	13(1.1)	287(4.1)	9(1.3)	***(***)
Writing helps me tell others what I think	8	1994	54(1.3)	267(2.6)	22(1.0)	268(2.7)	13(1.0)	262(4.2)	11(0.8)	250(3.3)
		1984	52(2.1)	269(3.1)	20(1.4)	267(3.0)	17(1.5)*	269(4.8)	11(1.4)	***(***)
	11	1994	58(1.4)	287(1.6)	20(1.2)	281(3.2)	15(0.9)	285(4.0)	8(0.8)	274(4.3)
		1984	55(1.9)	289(2.6)	22(1.4)	290(4.0)	15(1.0)	285(3.9)	8(1.3)	***(***)
Writing helps me tell others how I feel	8	1994	52(1.5)	268(2.4)	22(1.1)	265(3.1)	14(1.2)	261(4.3)	12(0.9)	252(3.7)
		1984	50(2.0)	270(3.0)	21(2.1)	266(3.4)	14(1.6)	266(3.6)	15(1.4)	259(5.8)
	11	1994	60(1.4)	287(1.4)	19(1.1)	281(3.7)	14(1.1)	286(3.2)	8(0.7)	271(6.0)
		1984	55(2.3)	290(2.8)	23(2.1)	287(3.5)	14(1.5)	288(5.7)	8(1.3)	***(***)
Writing helps me understand my own feelings	8	1994	45(1.4)	267(2.8)	23(1.4)	266(2.6)	15(1.1)	265(3.8)	17(0.9)	255(3.4)
		1984	40(2.2)	270(2.6)	23(1.8)	266(3.7)	17(1.7)	268(3.9)	20(1.8)	263(3.9)
	11	1994	54(1.7)	287(1.4)	18(1.5)	284(3.3)	17(1.5)	285(2.9)	11(1.0)	273(5.1)
		1984	47(2.1)*	289(2.5)	22(1.8)	289(4.3)	16(1.4)	291(4.4)	15(2.1)	276(6.0)
People who write well have a better chance of getting good jobs	8	1994	51(1.6)	266(1.9)	29(1.8)	265(2.5)	13(1.1)	261(4.1)	8(0.7)	255(6.9)
		1984	47(2.0)	267(2.3)	31(1.7)	268(3.5)	14(1.6)	272(5.6)	8(1.3)	***(***)
	11	1994	58(1.6)	288(2.2)	28(1.5)	279(2.9)	9(0.9)	286(4.0)	5(0.7)	273(6.5)
		1984	54(2.1)	289(2.1)	30(1.9)	284(4.0)	11(1.3)	292(3.9)	5(1.1)	***(***)
People who write well are more influential	8	1994	51(1.5)	269(2.1)	27(1.3)	263(3.0)	12(1.1)	259(3.6)	9(0.8)	254(7.5)
		1984	49(1.9)	268(2.8)	28(2.2)	270(2.9)	12(1.2)	264(5.0)	11(1.2)	258(6.1)
	11	1994	57(1.5)	287(1.9)	26(1.1)	283(2.5)	12(1.2)	282(3.5)	5(0.8)	273(7.5)
		1984	54(2.4)	290(2.8)	24(2.0)	285(4.5)	15(1.7)	287(4.8)	7(1.1)	***(***)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages and scale scores appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Students' responses to questions about their attitudes toward writing are summarized in Table 13.3. Questions in this set asked students at all three grades to react to statements such as "I like to write" and "I write on my own outside of school."

Comparing responses in 1994 to those in 1984, a smaller percentage of eighth graders in the most recent assessment reported that the statement, "I like to write," was true about half the time. At both the eighth and eleventh grades, there was an increase in the percentage of students who indicated that the statement, "I am a good writer," was true more than half the time. Similarly, an increase was observed between 1984 and 1994 in the percentage of students at grades 8 and 11 who reported that the following statement was true more than half the time: "People like what I write."

There were mixed results in the comparison of 1984 and 1994 responses to questions about students' enjoyment of writing activities. Between the first and most recent assessment years, there was a decrease in the percentage of fourth graders who reported that the statement, "I don't like to write things that will be graded," was true more than half the time. However, at grades 8 and 11 a reversed pattern was observed: higher percentages of students indicated that this statement was true more than half the time.

Among fourth graders, a decrease was also observed in the percentage of students who indicated that the statement, "If I didn't have to write for school, I wouldn't write anything," was true more than half the time.

In 1994, eighth- and eleventh-grade students who indicated that they liked to write more than half the time had higher average writing scores than students who said they never or hardly ever liked to write. Similarly, eighth-and eleventh-graders who reported that the following two statements were true more than half the time had higher average writing scores than their peers who said the statements were never or hardly ever true: "I am a good writer," and "People like what I write." At the fourth and eighth grades, students who indicated that the statement, "If I didn't have to write for school, I wouldn't write anything," was never or hardly ever true had higher average writing scores than students who reported that the statement was true more than half the time.

Table 13.3

Trends in Attitudes Toward Writing, Grades 4, 8, and 11, 1984 to 1994

			HOW OFTEN STUDENTS FIND STATEMENTS TO BE TRUE							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
	Grade	Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
I like to write	4	1994	56(2.0)	205(2.3)	18(1.1)	212(4.5)	12(0.9)	207(3.2)	13(1.2)	195(3.9)
		1984	56(2.3)	204(3.0)	18(1.5)	208(4.6)	11(1.6)	203(5.3)	15(1.6)	199(4.4)
	8	1994	42(1.5)	270(1.6)	26(1.6)	265(3.3)	18(1.3)	263(2.2)	14(1.1)	251(4.2)
		1984	39(2.5)	270(3.0)	33(2.0)*	270(2.9)	17(1.6)	264(3.8)	11(1.5)	***(***)
	11	1994	42(1.8)	290(2.2)	28(1.7)	284(2.9)	19(1.2)	284(3.1)	11(1.0)	275(4.0)
		1984	40(2.5)	295(3.7)	32(2.1)	290(4.0)	18(1.5)	287(5.4)	10(1.3)	***(***)
I am a good writer	4	1994	64(1.4)	206(2.4)	18(1.2)	208(2.6)	8(0.8)	208(3.6)	10(1.0)	192(5.5)
		1984	60(2.1)	204(2.4)	20(1.7)	209(5.0)	11(1.7)	202(4.9)	9(1.2)	199(6.3)
	8	1994	49(1.9)	271(2.0)	27(1.4)	264(1.9)	13(1.2)	259(3.2)	11(0.8)	249(5.0)
		1984	42(1.8)*	271(3.3)	34(1.8)*	268(2.1)	13(1.3)	265(5.3)	12(1.3)	255(4.8)
	11	1994	46(1.8)	292(1.8)	30(1.4)	285(2.3)	14(1.3)	278(3.6)	9(0.8)	266(6.1)
		1984	39(1.9)*	298(3.5)	34(2.2)	291(3.7)	17(2.1)	282(3.5)	10(1.1)	***(***)
People like what I write	4	1994	58(1.6)	207(2.0)	21(1.1)	211(3.4)	10(0.8)	199(4.6)	11(1.1)	195(4.7)
		1984	53(2.1)	205(2.3)	21(1.6)	208(3.5)	12(1.6)	205(6.3)	14(1.1)	194(5.1)
	8	1994	44(1.8)	270(2.1)	32(1.3)	267(3.0)	13(1.2)	258(2.8)	10(0.8)	248(4.4)
		1984	38(2.4)*	268(3.3)	36(2.1)	272(2.8)	15(1.6)	262(4.8)	11(1.6)	254(4.7)
	11	1994	44(2.2)	290(1.9)	37(1.6)	286(2.3)	12(1.2)	279(5.2)	7(0.8)	269(5.6)
		1984	36(2.6)*	294(3.5)	39(2.6)	294(3.5)*	18(2.1)*	281(3.3)	8(1.6)	***(***)
I write on my own outside of school	4	1994	45(1.8)	207(2.2)	14(1.1)	204(4.1)	12(0.9)	212(4.0)	29(1.5)	199(3.0)
		1984	48(1.9)	204(2.6)	13(1.5)	205(5.5)	10(1.1)	206(4.2)	30(1.7)	203(3.1)
	8	1994	36(1.7)	268(2.3)	15(1.1)	268(2.8)	20(1.2)	267(2.7)	29(1.3)	259(2.7)
		1984	36(2.5)	270(2.8)	16(2.0)	270(4.1)	22(1.8)	266(3.5)	26(2.3)	262(3.7)
	11	1994	32(1.9)	289(2.8)	16(1.2)	284(3.4)	23(1.2)	287(2.5)	30(1.2)	282(2.3)
		1984	31(2.7)	293(3.6)	17(1.9)	285(4.5)	21(2.6)	291(4.4)	31(1.6)	290(4.4)
I don't like to write things that will be graded	4	1994	33(1.2)	201(2.7)	12(0.9)	211(3.4)	15(1.2)	212(2.3)	40(1.4)	204(3.4)
		1984	38(2.0)*	201(3.1)	14(1.8)	207(6.6)	12(1.3)	206(4.8)	35(2.1)	205(3.4)
	8	1994	38(1.4)	262(2.5)	20(1.1)	268(3.0)	18(1.1)	269(3.9)	24(1.2)	266(2.0)
		1984	31(2.2)*	268(2.8)	23(2.0)	272(3.0)	20(2.2)	270(3.5)	26(2.6)	260(4.2)
	11	1994	33(1.3)	285(2.5)	24(1.2)	289(2.3)	22(1.4)	286(3.2)	21(1.3)	281(2.5)
		1984	27(1.9)*	292(3.6)	26(2.1)	292(4.1)	24(2.0)	290(4.0)	23(2.1)	287(3.1)
If I didn't have to write for school, I wouldn't write anything	4	1994	27(1.3)	198(2.6)	12(0.8)	206(3.3)	13(1.2)	208(3.8)	48(1.3)	209(2.4)
		1984	33(1.7)*	198(3.1)	10(1.3)	204(6.3)	11(1.3)	211(5.6)	45(2.2)	207(2.1)
	8	1994	21(1.2)	256(3.4)	17(1.0)	262(3.4)	21(1.6)	267(2.7)	41(1.5)	270(1.9)
		1984	17(2.2)	262(4.3)	20(2.6)	268(3.3)	20(1.8)	270(4.0)	43(2.4)	268(2.8)
	11	1994	17(1.1)	280(4.0)	16(1.3)	283(3.4)	23(1.2)	284(2.7)	45(1.7)	289(1.6)
		1984	15(1.4)	290(5.4)	15(1.6)	285(5.9)	24(2.2)	296(3.9)*	46(2.4)	289(3.1)

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Another set of questions asked students about the uses of writing in their lives — including such personal uses as writing letters to friends or relatives, and more functional uses like writing notes and messages. Table 13.4 presents students' responses to questions asked at all three grades. Students in grades 8 and 11 were asked more detailed questions about their personal and social uses of writing; their responses are presented in Table 13.5.

Between 1984 and 1994, there were no statistically significant changes in fourth graders' reports about their engagement in various writing activities. At the eighth and eleventh grades, however, some changes were observed. At grade 8, there was an increase from 1984 to 1994 in the percentage of students who reported writing letters to friends or relatives at least once a week and a decrease in the percentage who reported doing so once or twice a month. At grade 11, there was a decline between the first and most recent assessment years in the percentage of students who reported never or hardly ever writing notes and messages. The reports of eighth and eleventh graders also indicated increased writing of stories or poems that are not for schoolwork between 1984 and 1994.

At grades 8 and 11 in 1994, students who reported writing notes and messages at least once a week had higher average writing scores than students who reported never or hardly ever doing this kind of writing. At grade 8 in 1994, students who reported writing letters to friends or relatives at least once a week had higher average writing scores than students who said they never or hardly ever did this.

Table 13.4

Trends in Personal and Social Uses of Writing, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEY ENGAGED IN VARIOUS KINDS OF WRITING ACTIVITIES					
			At Least Once a Week		Once or Twice a Month		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Write letters to friends or relatives	Grade 4	1994	35(1.1)	202(3.1)	33(1.2)	212(2.8)	32(1.3)	200(3.0)
		1984	33(2.1)	200(3.1)	37(2.1)	211(3.1)	31(1.6)	199(3.5)
	Grade 8	1994	45(1.9)	270(2.1)	29(1.6)	271(2.6)	26(1.5)	255(2.8)
		1984	37(2.0)*	270(2.4)	37(2.0)*	269(3.2)	26(1.5)	261(3.4)
	Grade 11	1994	38(2.0)	286(2.6)	35(1.8)	287(2.8)	27(1.5)	279(2.7)
		1984	36(1.8)	294(2.7)*	38(1.9)	292(2.8)	26(2.1)	289(2.7)*
Write notes and messages	Grade 4	1994	44(1.7)	209(2.4)	18(1.3)	203(4.4)	39(1.2)	201(2.7)
		1984	44(2.4)	205(3.4)	21(2.2)	202(4.2)	35(2.2)	203(3.1)
	Grade 8	1994	72(1.3)	271(1.8)	11(0.9)	261(3.4)	17(1.0)	249(3.7)
		1984	68(2.0)	271(2.5)	12(1.4)	263(4.9)	20(1.8)	256(4.4)
	Grade 11	1994	77(1.4)	288(2.0)	13(1.1)	275(3.2)	10(1.0)	270(3.9)
		1984	74(1.9)	295(2.0)*	11(1.2)	284(3.8)	15(1.4)*	284(4.0)*
Write stories or poems that are not schoolwork	Grade 4	1994	26(1.2)	199(2.7)	22(1.3)	213(2.7)	53(1.2)	204(2.6)
		1984	26(1.8)	197(4.5)	23(1.7)	208(5.1)	51(2.7)	206(2.6)
	Grade 8	1994	18(0.9)	267(2.8)	24(1.3)	271(3.0)	59(1.5)	264(1.9)
		1984	10(1.0)*	266(5.7)	18(1.2)*	271(3.8)	71(1.6)*	266(2.6)
	Grade 11	1994	16(1.3)	283(3.2)	21(1.1)	287(4.1)	63(1.4)	284(1.5)
		1984	12(1.1)*	290(3.8)	18(1.8)	293(3.6)	71(1.4)*	292(2.0)*

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

According to results presented in Table 13.5, there was a decrease between the first and most recent assessment years in the percentage of eighth-grade students who reported never or hardly ever making lists of things to do or buy. There were also indications that more eighth graders were copying recipes or directions in 1994 than in 1984. At grades 8 and 11, diary or journal writing was more frequent in 1994 than in 1984. At both grades, there was an increase in the percentage of students who reported keeping diaries or journals at least once a week. One other change in uses of writing was observed at grade 11, where a smaller percentage of students in 1994 than in 1984 reported doing crossword puzzles at least once a week.

Examining the 1994 average scores revealed three instances of a relationship between frequency of the writing activities presented in Table 13.5 and writing scale scores. At grade 11, students who reported making lists of things to buy or do at least once a week and students who reported keeping a diary or journal at least once a week had higher average writing scores than students who reported never or hardly ever doing these writing activities. At grade 8, students who reported copying recipes or directions monthly had higher average scores than students who reported never or hardly ever doing so.

Table 13.5

Trends in Personal and Social Uses of Writing, Grades 8 and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEY ENGAGED IN VARIOUS KINDS OF WRITING ACTIVITIES					
			At Least Once a Week		Once or Twice a Month		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Make lists of things to buy or do	Grade 8	1994	46(1.4)	267(2.5)	25(1.1)	267(2.7)	29(1.1)	259(3.7)
		1984	44(2.4)	267(2.5)	21(1.9)	274(3.7)	35(2.0)*	263(3.6)
	Grade 11	1994	48(1.3)	288(1.5)	25(1.3)	285(2.4)	27(1.4)	278(2.9)
		1984	46(2.3)	290(2.3)	24(2.1)	291(3.1)	31(2.6)	281(3.8)
Copy recipes or directions	Grade 8	1994	23(1.5)	263(5.2)	34(1.6)	271(1.7)	44(1.4)	261(2.3)
		1984	22(1.9)	266(3.5)	29(1.5)*	270(2.8)	50(2.6)*	266(2.9)
	Grade 11	1994	18(1.3)	284(2.8)	33(1.3)	288(2.0)	49(1.1)	283(1.8)
		1984	20(1.8)	286(3.8)	33(2.0)	292(3.5)	47(2.4)	284(3.4)
Keep a diary or journal	Grade 8	1994	33(1.4)	269(3.6)	12(0.9)	268(3.5)	55(1.2)	262(2.1)
		1984	26(1.8)*	270(3.0)	12(1.0)	270(3.7)	62(2.2)*	266(2.8)
	Grade 11	1994	27(1.2)	289(2.6)	13(0.8)	293(2.8)	60(1.4)	281(1.7)
		1984	19(1.4)*	291(4.3)	13(1.3)	287(3.9)	68(1.5)*	286(2.4)
Do a crossword puzzle	Grade 8	1994	39(1.7)	265(2.6)	32(1.1)	268(2.6)	28(1.6)	259(3.5)
		1984	38(2.6)	264(3.5)	35(2.1)	271(2.1)	27(2.3)	268(3.7)
	Grade 11	1994	24(1.4)	284(2.9)	38(1.6)	285(2.1)	39(1.2)	285(1.8)
		1984	30(1.9)*	289(3.5)	36(1.8)	288(3.3)	35(2.2)	285(4.2)

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

The final set of questions about uses of writing asked students to describe their families' engagement in various kinds of writing activities. Students' responses to these questions are presented in Tables 13.6a and 13.6b. The results are somewhat mixed, indicating that while some activities were occurring more frequently in 1994 than in 1984, the reverse was true for other activities.

Between 1984 and 1994, there was an increase in the percentage of eighth graders who reported that their families wrote notes or messages at least once a week. At grade 4, a larger percentage of students in 1994 than in 1984 reported that their families never or hardly ever made lists. A smaller percentage of fourth and eleventh graders in 1994 than in 1984 reported that copying recipes or directions occurred in their families at least once a week. There were mixed results regarding the working of crossword puzzles in students' families. At grade 4, students in 1994 were less likely than those in 1984 to say that their families never or hardly ever engaged in this activity. Eleventh graders were less likely to report that this activity occurred at least once a week.

Table 13.6a

Trends in Family Uses of Writing, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEIR FAMILIES ENGAGED IN VARIOUS KINDS OF WRITING ACTIVITIES							
			At Least Once a Week		Once or Twice a Month		Never or Hardly Ever		I Don't Know	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Write notes or messages	Grade 4	1994	45(1.4)	207(2.4)	17(1.1)	206(4.3)	23(0.9)	205(3.3)	15(1.1)	195(4.4)
		1984	42(2.3)	205(3.7)	18(2.2)	205(5.3)	26(2.0)	203(3.9)	15(1.6)	201(4.1)
	Grade 8	1994	68(1.4)	271(1.9)	10(0.8)	262(5.3)	16(0.9)	256(3.1)	6(0.7)	247(5.9)
		1984	59(2.2)*	273(2.2)	13(1.3)	266(4.6)	21(1.7)*	256(4.3)	7(0.9)	***(***)
	Grade 11	1994	75(1.3)	289(1.9)	10(1.0)	275(3.8)	11(1.0)	271(3.0)	4(0.5)	***(***)
		1984	74(1.9)	296(2.0)*	8(1.1)	***(***)	14(1.5)	280(3.4)*	5(1.0)	***(***)
Make lists of things to buy or do	Grade 4	1994	60(1.4)	207(2.3)	10(1.0)	209(4.6)	17(0.7)	204(2.6)	13(0.9)	194(4.8)
		1984	58(2.2)	205(2.4)	13(1.4)	206(6.3)	13(1.4)*	206(4.6)	17(1.8)*	197(5.1)
	Grade 8	1994	64(1.3)	269(2.2)	13(1.0)	267(2.6)	19(1.3)	258(3.3)	5(0.7)	***(***)
		1984	63(1.9)	271(2.1)	10(1.8)	***(***)	21(1.9)	262(3.9)	6(1.5)	***(***)
	Grade 11	1994	66(1.4)	289(2.2)	13(0.9)	284(3.9)	17(1.0)	276(3.0)	4(0.6)	270(4.0)
		1984	70(2.2)	293(2.5)	13(1.8)	284(5.5)	13(1.7)	284(4.6)	3(0.8)	***(***)
Copy recipes or directions	Grade 4	1994	31(1.1)	203(2.2)	20(1.2)	214(2.3)	31(1.0)	207(2.6)	19(1.1)	198(3.6)
		1984	36(2.0)*	204(3.1)	20(1.6)	208(3.7)	26(2.2)	205(3.1)	18(1.4)	197(5.3)
	Grade 8	1994	29(1.4)	266(3.4)	30(1.5)	273(1.8)	34(1.8)	261(2.1)	8(0.7)	252(4.7)
		1984	29(1.9)	267(2.9)	31(2.3)	274(3.1)	33(2.0)	263(3.8)	8(1.4)	***(***)
	Grade 11	1994	27(1.9)	285(3.0)	34(1.5)	290(2.4)	32(1.4)	281(2.1)	8(0.7)	281(5.5)
		1984	33(2.1)*	289(3.7)	34(2.3)	296(2.8)	26(2.1)*	286(4.9)	8(1.2)	***(***)
Work crossword puzzles	Grade 4	1994	57(1.6)	204(2.6)	15(0.9)	213(3.5)	21(1.3)	206(2.6)	7(0.8)	195(4.9)
		1984	51(2.6)	202(2.1)	15(1.8)	206(5.2)	27(1.9)*	207(3.3)	7(1.0)	***(***)
	Grade 8	1994	38(1.2)	265(2.1)	23(1.1)	269(2.4)	33(1.4)	265(2.8)	6(0.8)	255(6.4)
		1984	41(2.0)	265(2.7)	23(2.3)	269(2.8)*	30(2.4)	270(3.3)	6(1.3)	***(***)
	Grade 11	1994	31(1.6)	287(2.2)	23(1.4)	286(2.8)	39(1.6)	285(2.3)	7(0.9)	281(6.8)
		1984	37(2.3)*	291(3.4)	24(2.1)	292(4.0)	33(2.5)	290(4.3)	6(1.2)	***(***)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 13.6b

Trends in Family Uses of Writing, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEIR FAMILIES ENGAGED IN VARIOUS KINDS OF WRITING ACTIVITIES							
			At Least Once a Week		Once or Twice a Month		Never or Hardly Ever		I Don't Know	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Write letters to relatives or friends	Grade 4	1994	30(1.3)	201(2.7)	31(1.4)	212(2.8)	23(1.7)	206(3.2)	16(0.9)	194(4.8)
		1984	34(2.3)	202(3.3)	30(2.3)	210(3.5)	23(2.6)	201(4.8)	12(1.5)*	202(4.7)
	Grade 8	1994	33(1.3)	268(2.6)	38(1.5)	270(2.8)	23(1.2)	260(2.2)	6(0.8)	258(6.3)
		1984	35(2.1)	271(2.5)	39(2.4)	269(3.2)	20(2.1)	263(4.2)	7(1.0)	***(***)
	Grade 11	1994	29(1.5)	287(2.7)	38(1.3)	288(2.5)	27(1.4)	280(2.6)	6(0.7)	271(5.1)
		1984	36(2.4)*	295(3.0)*	39(1.9)	293(2.8)	20(1.6)*	287(2.9)	5(1.1)	***(***)
Write business letters	Grade 4	1994	33(1.1)	206(3.3)	19(1.1)	211(3.7)	25(1.3)	203(2.8)	24(1.6)	198(4.0)
		1984	33(2.1)	206(3.8)	16(1.6)	205(4.4)	29(2.2)	204(4.3)	22(2.0)	202(4.3)
	Grade 8	1994	41(1.8)	270(2.6)	21(1.1)	269(2.4)	24(1.3)	261(2.9)	14(1.1)	258(4.0)
		1984	35(1.3)*	271(3.0)	22(1.6)	271(3.2)	31(2.0)*	263(3.4)	12(1.1)	256(5.8)
	Grade 11	1994	39(1.5)	289(2.5)	24(1.5)	286(2.4)	27(1.8)	279(2.4)	11(1.0)	277(3.1)
		1984	38(2.1)	296(3.0)	27(1.6)	291(3.1)	26(2.2)	290(3.5)*	10(1.2)	***(***)
Write stories or poems	Grade 4	1994	18(1.1)	195(2.7)	14(0.9)	209(4.3)	49(1.7)	209(2.6)	20(1.4)	198(3.7)
		1984	18(1.5)	194(5.1)	16(2.0)	210(4.3)	46(2.1)	208(2.6)	21(1.7)	201(3.2)
	Grade 8	1994	11(0.7)	263(3.9)	19(1.0)	274(3.3)	56(1.3)	266(2.3)	14(1.0)	259(3.9)
		1984	8(1.2)*	***(***)	14(1.2)*	270(3.5)	66(2.3)*	268(3.0)	13(1.8)	258(5.1)
	Grade 11	1994	10(1.0)	285(4.0)	18(1.2)	285(3.8)	59(1.8)	286(1.9)	13(1.3)	276(3.4)
		1984	8(1.5)	***(***)	13(1.4)*	297(4.8)*	65(2.4)	292(2.1)*	14(1.8)	289(4.1)*
Keep diaries or journals	Grade 4	1994	34(1.6)	204(3.1)	8(0.7)	207(5.1)	35(1.7)	208(2.1)	24(1.2)	202(2.7)
		1984	26(1.9)*	203(2.5)	10(1.2)	***(***)	29(1.4)*	211(3.8)	36(1.9)*	201(3.1)
	Grade 8	1994	23(1.6)	267(2.4)	8(0.6)	273(4.9)	49(1.7)	264(1.9)	21(1.1)	262(3.2)
		1984	22(1.7)	269(3.7)	5(1.1)*	***(***)	50(1.8)	268(2.5)	23(1.8)	261(4.4)
	Grade 11	1994	22(1.5)	294(2.4)	10(1.0)	291(4.2)	48(1.7)	284(1.8)	20(1.4)	277(4.3)
		1984	18(2.5)	291(4.7)	7(1.4)	***(***)	56(3.1)*	290(2.8)	20(2.2)	289(3.4)*

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Personal letter writing (to relatives or friends) in eleventh-grade students' families decreased between 1984 and 1994. There was a decline in the percentage of students who reported that this activity occurred at least once a week. At the eighth grade, the percentage of eighth graders who reported that business letters were written in their families at least once a week was higher in 1994 than in 1984.

The families of eighth graders were more likely to engage in writing stories or poems on a weekly or monthly basis in 1994 than in 1984. Among eleventh graders, there was an increase in the percentage of students who reported that writing stories or poems occurred in their families once or twice a month. Diary or journal writing also showed some indication of being more frequent in students' families in 1994 than in 1984. In the most recent assessment, fourth graders were more likely to report that this type of writing occurred at least once a week in their families; eighth graders were more likely to report that it occurred once or twice a month, and eleventh graders were less likely to report that it never or hardly ever occurred.

At grades 8 and 11 in 1994, students who reported that their families wrote notes and messages at least once a week, or made lists of things to buy or do at least once a week, had higher average writing scores than their peers who reported that their families never or hardly ever engaged in these activities. At grade 11, students who indicated that business letters were written at least once a week in their families had higher average writing scores than students who said that this never or hardly ever happened in their families. Also at grade 11, students who reported that their families kept diaries or journal at least once a week had higher average writing scores than students who reported that their families never or hardly ever did this. Among fourth graders in 1994, students who reported that their families engaged in writing stories or poems at least once a week had lower average writing scores than students who reported that their families never or hardly ever did this type of writing.

Managing the Writing Process

In addition to building positive attitudes toward writing, teachers seek to help students develop effective strategies for managing writing processes, including strategies for planning and revising what they write. Previous studies and earlier national assessments of writing have shown that students who make use of a variety of strategies are more effective writers.[72]

One writing task given as part of the assessment provided an opportunity to observe explicit planning strategies. The "Recreation Opportunities" task was formatted so that the remainder of the page on which the writing prompt was printed was left blank and the students were told that they could use this space to make notes before writing. The subsequent pages were used for students' actual responses. In addition to rating the quality of the responses, raters noted whether the students had used the space provided to make notes.

Table 13.7 summarizes the evidence of overt planning for the eighth and eleventh graders who were given this writing task. There were no significant differences between 1984 and 1994 in the percentage of eighth and eleventh graders who used the space provided for planning their response. In both assessments, approximately 16 to 20 percent of the students used the planning space.

[72]Applebee, A. N.; Langer, J. A.; Jenkins, L. B.; Mullis, I. V. S.; & Foertsch, M. A., *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1990).

Table 13.7

Trends in Overt Planning on "Recreation Opportunities" Task, Grades 8 and 11, 1984 to 1994

	GRADE 8		GRADE 11	
	Used Planning Space	Did Not Use Planning Space	Used Planning Space	Did Not Use Planning Space
Year	Percent	Percent	Percent	Percent
1994	17(1.1)	83(1.1)	20(1.4)	80(1.4)
1984	16(2.4)	84(2.4)	18(2.0)	82(2.0)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. The standard errors of the estimated percentages appear in parentheses. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

A variety of questions asked students about the revising and editing strategies they used, including their attention to writing conventions (spelling, punctuation, and grammar) as well as to the structure and organization of the text as a whole. Their responses are summarized in Tables 13.8a and 13.8b.

Table 13.8a

Trends in the Use of Specific Revising and Editing Strategies, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEY USE STRATEGIES							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Correct spelling	Grade 4	1994	76(1.0)	207(1.8)	10(0.7)	206(3.8)	7(0.7)	203(5.3)	7(0.7)	194(6.4)
		1984	75(1.7)	206(2.1)	10(1.3)	203(7.0)	6(1.1)	***(***)	9(0.9)	***(***)
	Grade 8	1994	77(1.2)	268(1.6)	10(0.8)	258(4.2)	7(0.6)	258(4.3)	6(1.0)	252(5.6)
		1984	75(1.7)	267(2.2)	12(1.3)	273(5.3)*	6(1.1)	***(***)	7(1.2)	***(***)
	Grade 11	1994	76(1.5)	288(1.6)	11(0.9)	274(4.9)	8(1.0)	282(7.4)	6(0.7)	273(4.8)
		1984	76(2.4)	294(2.4)	10(1.3)	***(***)	9(1.5)	***(***)	5(1.4)	***(***)
Correct punctuation	Grade 4	1994	63(1.3)	208(1.8)	12(1.0)	205(3.5)	10(1.0)	201(5.1)	16(1.0)	199(4.3)
		1984	65(2.5)	206(2.7)	12(1.7)	206(5.9)	10(1.4)	200(3.4)	14(2.1)	196(4.8)
	Grade 8	1994	68(1.2)	269(1.8)	15(0.9)	258(2.9)	11(0.9)	260(3.8)	7(0.8)	252(3.3)
		1984	69(2.0)	268(2.1)	12(1.3)	270(5.7)	10(1.4)	***(***)	9(1.6)	***(***)
	Grade 11	1994	71(1.5)	289(1.7)	12(0.8)	278(3.5)	9(1.1)	280(6.7)	7(0.8)	272(4.6)
		1984	70(2.7)	294(2.4)	12(1.3)	281(6.5)	11(1.9)	***(***)	7(1.5)	***(***)
Correct grammar	Grade 4	1994	54(1.3)	209(2.6)	15(1.1)	205(3.1)	9(0.9)	206(3.7)	22(1.1)	198(3.6)
		1984	51(2.1)	205(2.6)	11(1.2)*	209(6.5)	9(0.9)	***(***)	29(2.1)*	201(3.9)
	Grade 8	1994	68(1.0)	268(1.7)	16(1.0)	262(3.4)	9(0.7)	259(3.9)	7(0.9)	250(4.1)
		1984	65(2.2)	269(2.4)	17(1.9)	268(3.7)	10(1.5)	***(***)	8(1.4)	***(***)
	Grade 11	1994	73(1.6)	289(1.5)	13(1.2)	274(4.9)	9(0.9)	284(5.0)	6(0.8)	274(6.4)
		1984	70(2.3)	294(2.5)*	13(1.3)	286(5.0)	11(1.7)	***(***)	6(1.0)	***(***)
Change words	Grade 4	1994	67(1.2)	206(1.9)	15(1.0)	211(3.0)	8(0.7)	200(5.2)	11(1.2)	197(5.9)
		1984	62(2.2)*	206(2.4)	15(1.4)	206(5.2)	9(1.2)	198(4.7)	14(1.2)	196(5.7)
	Grade 8	1994	72(1.3)	269(1.8)	15(1.0)	260(3.4)	8(0.9)	256(4.7)	5(0.8)	243(5.8)
		1984	65(2.3)*	271(2.7)	16(1.3)	266(3.8)	11(1.6)	***(***)	9(1.3)*	***(***)
	Grade 11	1994	74(1.4)	289(1.6)	15(1.0)	280(4.0)	7(0.8)	268(5.2)	4(0.7)	***(***)
		1984	71(2.5)	295(2.5)	16(2.0)	283(5.0)	8(1.3)	***(***)	4(1.1)	***(***)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 13.8b

Trends in the Use of Specific Revising and Editing Strategies, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN THEY USE STRATEGIES							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Add ideas or information	Grade 4	1994	65(1.5)	206(2.3)	16(1.3)	207(3.4)	8(0.8)	209(4.5)	10(1.0)	195(6.1)
		1984	61(2.0)	204(2.5)	17(1.5)	209(4.8)	10(1.5)	203(7.1)	12(1.3)	194(5.9)
	Grade 8	1994	65(1.3)	269(1.6)	20(1.4)	261(2.7)	8(0.7)	262(4.8)	6(0.8)	245(6.5)
		1984	60(2.8)	270(2.6)	22(1.9)	269(3.3)	12(1.9)	***(***)	7(1.3)	***(***)
	Grade 11	1994	72(1.2)	289(1.8)	18(1.1)	280(3.7)	6(0.7)	266(5.2)	3(0.5)	***(***)
		1984	70(2.0)	293(2.6)	19(1.3)	289(4.5)	8(1.4)	***(***)	4(0.9)	***(***)
Take out parts you don't like	Grade 4	1994	51(1.5)	207(1.9)	14(1.0)	207(4.7)	11(0.8)	210(3.8)	24(1.4)	198(4.0)
		1984	45(2.1)*	205(3.1)	13(1.4)	204(5.5)	11(1.2)	206(4.8)	32(2.0)*	202(3.3)
	Grade 8	1994	62(1.5)	269(1.9)	20(1.4)	264(2.8)	10(1.0)	257(3.5)	9(1.0)	245(5.0)
		1984	56(2.2)*	271(2.7)	20(1.8)	267(3.3)	13(1.8)	263(4.2)	11(1.5)	***(***)
	Grade 11	1994	69(1.9)	290(1.8)	18(1.2)	280(3.1)	8(1.1)	274(5.4)	5(0.8)	264(6.7)
		1984	58(3.0)*	295(2.5)	23(2.5)	285(5.2)	11(1.4)	285(5.7)	7(1.2)	***(***)
Move sentences or paragraphs	Grade 4	1994	38(1.4)	201(2.9)	17(1.2)	212(2.8)	17(1.4)	210(3.3)	28(1.4)	203(3.3)
		1984	44(2.0)*	202(3.4)	16(1.3)	205(4.9)	12(1.5)*	207(5.5)	27(1.8)	206(3.5)
	Grade 8	1994	40(1.7)	270(2.0)	25(1.4)	267(2.4)	19(1.1)	263(3.0)	17(1.3)	255(4.0)
		1984	30(2.0)*	272(3.0)	28(2.2)	268(3.5)	19(1.8)	268(3.4)	23(2.4)*	258(3.6)
	Grade 11	1994	46(1.6)	293(2.2)	25(1.5)	281(3.5)	17(1.0)	281(2.8)	12(1.0)	274(5.2)
		1984	46(2.6)	295(3.2)	24(1.9)	292(3.5)*	17(2.0)	287(5.7)	13(1.8)	276(4.6)
Rewrite most of the paper	Grade 4	1994	37(1.7)	203(2.2)	11(1.0)	206(4.3)	15(1.1)	211(3.1)	37(1.6)	205(3.5)
		1984	36(1.7)	201(2.7)	13(1.3)	207(5.2)	15(1.3)	202(4.2)	37(1.8)	207(3.1)
	Grade 8	1994	44(1.6)	265(2.1)	16(0.8)	268(2.7)	21(1.1)	266(3.0)	19(1.4)	262(3.0)
		1984	40(1.8)	268(2.7)	15(1.8)	268(3.8)	22(1.7)	271(3.3)	23(2.0)	262(4.6)
	Grade 11	1994	36(1.7)	286(2.2)	18(1.3)	281(4.1)	24(1.1)	290(2.3)	22(1.4)	283(2.6)
		1984	44(2.6)*	292(2.0)*	18(1.4)	290(5.7)	21(2.3)	295(5.6)	17(1.6)*	282(4.8)
Throw out and start over	Grade 4	1994	26(1.3)	198(3.3)	12(1.0)	205(3.9)	15(0.9)	209(3.9)	47(1.4)	208(2.5)
		1984	34(2.2)*	198(3.4)	13(1.3)	206(4.6)	14(1.1)	205(5.2)	40(2.2)*	208(2.6)
	Grade 8	1994	27(1.3)	260(2.8)	14(1.1)	266(2.6)	23(0.9)	272(2.4)	36(1.5)	265(2.4)
		1984	33(2.0)*	266(3.6)	17(1.9)	266(2.9)	19(1.5)*	271(3.8)	31(2.5)	267(4.3)
	Grade 11	1994	20(1.7)	279(3.1)	15(1.0)	283(3.7)	26(1.4)	287(3.2)	38(2.1)	289(1.9)
		1984	25(1.8)	286(3.4)	14(1.9)	285(5.6)	29(2.0)	294(4.1)	32(2.1)*	293(4.3)

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

At the fourth grade, there was some indication that more students were using grammar correction as a revising strategy. The percentage of fourth graders in 1994 who reported never or hardly ever doing this was smaller than the percentage in 1984. Also, there was an increase in the percentage of fourth graders who reported using this strategy about half the time. At grades 4 and 8, a larger percentage of students in 1994 than in 1984 reported changing words as a strategy for revising and editing.

At all three grades, there was an increase between the first and most recent assessment years in the percentage of students who reported that more than half the time when they were revising and editing they took out parts of writing that they did not like. A decline was observed between 1984 and 1994 at grades 4 and 8 in the percentage of students who reported that they throw out and start over more than half the time as an editing and revising strategy. At the eleventh grade, a smaller percentage of students in 1994 than in 1984 reported rewriting most of their papers more than half the time as revising or editing strategy.

At grades 8 and 11 in 1994, students who reported using the following strategies for revising and editing more than half the time had higher average writing scores than students who reported never or hardly ever using these strategies: correct spelling, correct punctuation, take out parts you don't like, and move sentences or paragraphs. At grade 8, three additional strategies demonstrated the same pattern: correct grammar, change words, and add ideas or information. At grade 11, the "throw out and start over" strategy had a reversed relationship with writing performance. Eleventh graders who reported using this method more than half the time had lower average writing scores than students who reported never or hardly ever using it.

Bereiter, C.; & Scardamalia, M., *The Psychology of Written Composition* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1987).

The Instructional Context

To better understand the state of writing achievement, NAEP looked at various factors that influence students in school. Research on effective instructional practices in writing emphasizes that students should write frequently and for a wide range of purposes.[73] Also, researchers have found that an emphasis on mechanics and correctness in writing has little or no effect on the improvement of writing.[74] Instead, they call for teachers to respond to students' writing in ways that communicate high expectations for all students, that emphasize the students' authority over their writing, and that support students throughout the writing process.[75]

As a way of measuring trends in writing instruction, several questions in the 1984 and 1994 assessments focused on the kinds and amount of writing that students did in school and on the kinds of responses that students received from their teachers.

Tables 13.9a and 13.9b summarize students' responses to a question about the kinds of writing they did for English class the previous week. At all three grades, there was evidence of increased writing of essays, compositions, or themes between 1984 and 1994. A larger percentage of fourth, eighth, and eleventh graders reported writing one or two of these types of papers during the previous week. In addition, there was an increase in the percentage of eighth-grade students who reported writing three or more of these types of papers. "Other" types of reports were also being written more frequently in 1994 than in 1984. There was a decrease in the percentage of students at all three grades who reported completing none of this type of report during the previous week. Correspondingly, there were increases in the percentages of eighth and eleventh graders who reported writing one or two of this type of report and in the percentage of eighth graders who reported writing three or more of this type of report the previous week.

---

[73] Graves, D. H., *Writing: Teaching and Children at Work* (Portsmouth, NH: Heinemann Educational Books, 1983).

[74] Hillocks, G., Jr., *Research on Written Composition: New Directions for Teaching* (Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills, 1986).

[75] Freedman, S. W., *Response to Student Writing* (Urbana, IL: National Council of Teachers of English, 1987).

Table 13.9a

Trends in Types of Writing for English Class, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW MANY OF VARIOUS TYPES OF PAPERS THEY WROTE FOR ENGLISH CLASS LAST WEEK					
			Three or More		One or Two		None	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Essay, composition, or theme	Grade 4	1994	3(0.3)	195(6.3)	22(1.0)	202(3.1)	75(1.1)	207(1.4)
		1984	3(0.7)	***(***)	16(1.5)*	206(4.1)	81(1.7)*	206(1.9)
	Grade 8	1994	6(0.5)	264(3.3)	48(1.6)	268(1.4)	46(1.7)	265(2.1)
		1984	4(0.5)*	271(3.0)	37(1.8)*	271(1.8)	59(1.9)*	268(2.5)
	Grade 11	1994	9(0.7)	280(2.3)	60(1.2)	287(1.3)	31(1.4)	283(1.8)
		1984	8(0.7)	282(3.2)	52(1.4)*	292(1.7)*	40(1.3)*	290(2.4)*
Book report	Grade 4	1994	6(0.7)	192(4.1)	32(1.2)	201(3.0)	62(1.5)	209(1.5)
		1984	6(0.7)	193(7.0)	31(1.4)	206(2.7)	64(1.5)	206(2.0)
	Grade 8	1994	4(0.4)	259(4.3)	31(2.0)	265(2.0)	65(2.1)	268(1.5)
		1984	3(0.6)	***(***)	32(1.6)	267(2.2)	65(1.7)	271(2.5)
	Grade 11	1994	5(0.5)	272(3.6)	23(1.0)	281(1.5)	72(1.1)	287(1.4)
		1984	4(0.6)	266(4.9)	26(1.3)	282(2.5)	70(1.6)	295(1.7)*
Other reports	Grade 4	1994	5(0.5)	195(4.8)	29(1.1)	205(2.9)	67(1.1)	206(1.6)
		1984	3(0.6)	***(***)	25(1.7)	202(2.8)	72(1.8)*	207(2.0)
	Grade 8	1994	5(0.4)	261(3.4)	32(1.1)	266(1.5)	64(1.2)	267(1.7)
		1984	3(0.4)*	***(***)	24(1.3)*	268(2.5)	74(1.5)*	270(2.3)
	Grade 11	1994	6(0.5)	273(4.0)	37(1.0)	286(1.4)	57(1.0)	286(1.6)
		1984	5(0.7)	277(4.4)	33(1.0)*	287(2.2)	62(1.0)*	293(1.9)*

* Statistically significant difference from 1994 at a 5 percent combined significance level per set of comparisons. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. The standard errors of the estimated percentages and scale scores appear in parenthesis. It can be said with 95-percent certainty that, for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 13.9b

Trends in Types of Writing for English Class, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW MANY OF VARIOUS TYPES OF PAPERS THEY WROTE FOR ENGLISH CLASS LAST WEEK					
			Three or More		One or Two		None	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Letter	Grade 4	1994	7(0.7)	191(4.9)	32(1.5)	202(1.9)	61(1.9)	208(1.7)
		1984	6(0.6)	191(5.7)	32(1.5)	204(3.1)	62(1.6)	208(1.8)
	Grade 8	1994	5(0.4)	257(4.2)	25(1.1)	261(1.7)	71(1.2)	269(1.7)
		1984	4(0.7)	265(4.5)	17(1.1)*	264(2.1)	79(1.3)*	271(2.2)
	Grade 11	1994	4(0.4)	271(4.2)	18(1.0)	278(2.1)	78(1.2)	288(1.5)
		1984	3(0.5)	272(4.7)	13(1.1)*	276(2.2)	84(1.1)*	293(1.7)*
Story	Grade 4	1994	6(0.6)	189(3.6)	40(1.8)	205(1.9)	54(1.7)	207(1.7)
		1984	6(0.8)	197(5.8)	32(1.9)*	204(3.2)	63(2.1)*	207(1.8)
	Grade 8	1994	6(0.5)	260(3.3)	46(1.7)	264(1.5)	48(1.6)	269(1.8)
		1984	4(0.5)*	263(3.8)	37(1.7)*	268(2.4)	59(1.6)*	271(2.5)
	Grade 11	1994	7(0.6)	273(2.7)	35(1.0)	281(1.5)	59(1.2)	289(1.7)
		1984	6(0.5)	276(4.0)	34(1.6)	283(1.8)	60(1.6)	296(2.0)*
Poem	Grade 4	1994	4(0.7)	197(6.8)	22(1.6)	201(2.4)	73(1.9)	207(1.5)
		1984	4(0.6)	***(***)	22(1.6)	200(2.8)	74(1.7)	207(1.8)
	Grade 8	1994	3(0.6)	263(5.0)	17(1.4)	266(2.5)	80(1.6)	267(1.6)
		1984	3(0.8)	***(***)	12(0.9)*	263(3.5)	85(1.3)*	270(2.2)
	Grade 11	1994	6(0.6)	277(3.5)	22(1.1)	283(1.9)	73(1.4)	287(1.3)
		1984	4(0.5)*	286(4.0)	14(1.0)*	279(2.4)	82(0.9)*	292(1.8)*
Play	Grade 4	1994	2(0.3)	182(7.7)	12(1.0)	194(2.8)	86(1.1)	207(1.5)
		1984	3(0.6)	***(***)	11(1.1)	196(3.5)	86(1.5)	207(1.7)
	Grade 8	1994	1(0.2)	***(***)	12(0.8)	262(2.8)	86(0.9)	267(1.5)
		1984	1(0.3)	***(***)	9(0.9)*	263(4.3)	90(1.0)	270(1.9)
	Grade 11	1994	1(0.2)	***(***)	12(0.9)	279(3.1)	87(1.0)	287(1.3)
		1984	1(0.3)	***(***)	11(0.9)	281(2.6)	87(0.9)	292(1.8)*

* Statistically significant difference from 1994 where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

At various grades, there was evidence that more letters, stories, poems, and plays were also being written in English classes in 1994 than in 1984. The percentage of eighth and eleventh graders who reported writing one or two letters during the previous week increased from 1984 to 1994, and the percentage who reported writing none decreased. The percentage of fourth and eighth graders who reported writing one or two stories during the previous week increased, and the percentage who reported writing no stories decreased. There was also an increase over time at grade 8 in the percentage who reported writing three or more stories. At grades 8 and 11, students in 1994 were more likely than those in 1984 to report writing one or two poems during the previous week and less likely to report not writing any poems. Also, at grade 11, students in 1994 were more likely to report writing three or more poems. At grade 8, there was an increase from 1984 to 1994 in the percentage of students who reported writing one or two plays in English class during the previous week.

In 1994, frequency of writing activities appears to be, in many cases, *negatively* related to writing scores. Fourth and eleventh graders who had not written a book report in the previous week had higher average writing scores than their peers who reported writing three or more book reports during the same time. At all three grades, students who reported that they had not written a letter or a story in the previous week for English class had higher average scores than their peers who reported that they had written at least three letters or stories. This same pattern was observed for eleventh graders' reports about "other reports" and poems, and for fourth and eighth graders' reports about writing plays.

This association of lower writing scores with more frequent writing activities might be explained, in part, by the difficulty students have in evaluating the types of assignments and class work in which they engage. At the fourth grade, in particular, students may not be able to distinguish between work done for language arts and writing done for other subject areas. Also, it may reflect that the lower performing students are assigned more activities to complete. In considering this association, it is important to keep in mind that for any given writing activity listed in Tables 13.9a and 13.9b, few students (9 percent or less) reported engaging in the activity three or more times in the previous week.

Another set of questions asked students about the ways in which their teachers responded to their writing assignments. The responses of students in grades 8 and 11 are summarized in Tables 13.10a and 13.10b.

Table 13.10a

Trends in Teachers' Comments on Completed Papers, Grades 8 and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN TEACHERS COMMENT ON ASPECTS OF THEIR PAPERS							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Ideas in paper	Grade 8	1994	48(1.6)	266(2.5)	22(1.1)	264(2.5)	17(1.2)	264(3.9)	13(0.9)	264(5.0)
		1984	41(1.8)*	267(3.6)	23(1.3)	268(3.2)	17(1.4)	269(3.4)	19(1.4)*	267(3.8)
	Grade 11	1994	49(1.5)	286(1.7)	24(1.0)	281(3.3)	16(1.4)	289(2.9)	11(1.0)	280(4.0)
		1984	40(1.8)*	289(2.6)	23(2.0)	283(3.3)	24(2.3)*	288(4.4)	14(1.6)	288(5.9)
Way ideas explained	Grade 8	1994	44(1.8)	265(2.7)	26(1.2)	264(2.5)	16(1.1)	265(3.6)	15(1.2)	266(3.4)
		1984	37(1.7)*	266(3.5)	25(1.9)	267(3.5)	21(1.9)*	267(3.3)	17(1.9)	271(5.0)
	Grade 11	1994	47(1.8)	286(2.0)	23(1.0)	283(2.6)	20(1.2)	285(2.6)	11(1.0)	283(4.7)
		1984	40(2.2)*	287(2.6)	25(2.0)	282(4.1)	23(2.1)	292(3.1)	13(1.5)	290(5.1)
Way feelings expressed	Grade 8	1994	42(1.6)	265(2.6)	20(1.1)	264(2.8)	18(1.1)	265(2.4)	20(1.4)	266(3.7)
		1984	33(2.2)*	265(3.4)	21(2.3)	268(4.0)	21(2.1)	271(3.8)	26(1.8)*	266(3.1)
	Grade 11	1994	42(1.8)	283(1.9)	20(1.3)	286(3.3)	20(1.4)	286(3.4)	19(1.2)	285(3.3)
		1984	31(2.3)*	285(3.1)	25(1.8)*	285(4.3)	25(1.7)*	289(3.3)	19(1.6)	292(3.9)
Organization	Grade 8	1994	47(1.8)	264(2.6)	21(1.2)	265(3.5)	16(0.9)	267(3.9)	16(1.3)	266(3.7)
		1984	43(2.0)	267(3.0)	22(1.3)	266(3.4)	16(1.2)	270(3.2)	19(2.0)	267(3.7)
	Grade 11	1994	47(1.5)	285(1.9)	20(1.1)	282(2.7)	18(1.3)	287(3.4)	15(1.3)	285(3.3)
		1984	40(2.1)*	288(3.0)	21(1.8)	287(4.0)	22(1.6)	286(3.3)	17(1.4)	289(6.3)

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Table 13.10b

Trends in Teachers' Comments on Completed Papers, Grades 8 and 11, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN TEACHERS COMMENT ON ASPECTS OF THEIR PAPERS							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Followed directions	Grade 8	1994	44(1.5)	260(2.2)	18(1.0)	261(4.7)	18(1.1)	269(2.6)	20(1.3)	274(3.6)
		1984	42(1.8)	262(3.0)	18(1.8)	270(3.5)	19(1.8)	268(3.5)	22(1.5)	274(3.3)
	Grade 11	1994	33(1.5)	278(2.1)	16(1.4)	283(3.1)	24(1.6)	287(2.4)	28(1.5)	291(2.3)
		1984	30(2.4)	277(3.2)	16(1.5)	284(5.5)	22(1.9)	290(3.5)	31(1.8)	297(3.0)
Wrote enough	Grade 8	1994	39(1.3)	261(2.5)	24(1.5)	261(3.1)	18(1.1)	269(3.1)	20(1.2)	273(3.5)
		1984	33(1.9)*	263(3.5)	23(1.9)	264(3.6)	24(1.5)*	270(3.8)	21(1.5)	274(2.9)
	Grade 11	1994	30(1.6)	281(2.7)	19(0.9)	280(2.8)	25(1.4)	286(1.6)	26(1.3)	291(2.5)
		1984	26(2.0)	281(3.1)	19(1.7)	284(4.4)	27(2.3)	289(3.2)	28(2.1)	293(3.5)
Words	Grade 8	1994	43(1.6)	264(2.6)	22(1.5)	262(2.7)	18(1.2)	271(4.0)	18(1.0)	263(4.7)
		1984	38(2.3)	264(3.0)	23(1.5)	267(3.4)	21(1.7)	271(4.1)	19(1.6)	270(4.4)
	Grade 11	1994	36(1.4)	283(2.3)	21(0.9)	283(2.7)	26(1.2)	288(3.3)	18(1.0)	287(3.4)
		1984	32(1.6)*	284(3.2)	24(1.7)	285(2.9)	22(1.5)	291(4.4)	23(1.7)*	292(3.6)
Spelling, punctuation, and grammar	Grade 8	1994	54(1.3)	265(2.4)	18(1.3)	262(2.5)	16(1.0)	267(4.3)	13(1.0)	264(4.9)
		1984	51(1.9)	267(2.4)	20(1.9)	264(4.5)	15(1.3)	271(4.4)	14(1.5)	268(6.0)
	Grade 11	1994	45(1.5)	282(2.1)	20(1.1)	284(3.3)	21(1.3)	290(2.5)	15(0.9)	287(3.8)
		1984	45(2.3)	285(2.7)	18(1.5)	286(4.3)	20(1.9)	289(3.6)	17(1.7)	294(3.0)
Neatness and handwriting	Grade 8	1994	46(1.5)	262(2.5)	15(1.0)	264(3.6)	15(1.0)	264(3.8)	24(1.3)	271(3.3)
		1984	48(2.2)	265(2.6)	14(1.6)	270(4.0)	15(1.5)	268(5.2)	22(1.9)	270(3.3)
	Grade 11	1994	31(1.3)	279(2.1)	13(1.0)	281(3.8)	19(1.4)	288(3.1)	38(1.4)	289(1.9)
		1984	31(2.7)	277(3.2)	10(1.0)*	278(4.9)	16(1.3)	287(5.1)	44(2.5)*	296(2.3)*

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

At grades 8 and 11, a larger percentage of students in 1994 than in 1984 indicated that their teachers commented on the following aspects of their completed papers more than half the time: the ideas in their paper, the way their ideas are explained, and the way their feelings are expressed. At grade 8, there was an increase between 1984 and 1994 in the percentage of students who said that their teachers commented more than half the time on whether they had written enough. At grade 11, there was an increase in the percentage who reported that their teachers commented more than half the time on the organization of their papers and the words they used in their papers.

At both grades 8 and 11 in 1994, students who reported that their teachers never or hardly ever commented on whether they had followed directions or on whether they had written enough had higher average writing scores than students who said their teachers comment on these aspects more than half the time. Similarly, eleventh graders who reported that their teachers never or hardly ever commented on neatness and handwriting had higher average scores than their peers with teachers who did so more than half the time.

Table 13.11 summarizes fourth and eighth graders' responses to a related series of questions asking about the types of feedback (oral or written) that they received from teachers on their writing. Between 1984 and 1994, there was a decline in the percentage of fourth- and eighth-grade students who reported that their teachers marked mistakes more than half the time as a part of their feedback on students' writing. At grade 4, there was also a decline in the percentage of students who said that their teachers pointed out what is not done well more than half the time in feedback on students' writing. At grade 8, there was an increase in the percentage of students who reported that their teachers more than half the time wrote notes, pointed out what is well done, and made suggestions for next time.

Table 13.11

Trends in Teachers' Feedback on Writing, Grades 4 and 8, 1984 to 1994

			STUDENTS' REPORTS ON HOW OFTEN TEACHERS PROVIDE TYPES OF FEEDBACK WHEN THEY WRITE							
			More Than Half the Time		About Half the Time		Less Than Half the Time		Never or Hardly Ever	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score	Percent	Average Score
Mark mistakes	Grade 4	1994	51(1.4)	206(2.7)	12(1.0)	203(4.1)	19(1.3)	208(2.9)	17(1.4)	203(2.9)
		1984	60(2.5)*	204(2.3)	15(1.6)	206(5.3)	12(1.6)*	200(5.5)	12(1.6)*	203(4.8)
	Grade 8	1994	63(1.4)	270(2.0)	14(1.0)	259(3.7)	13(1.0)	264(3.7)	11(1.0)	259(4.2)
		1984	69(1.7)*	268(2.3)	11(1.1)	260(5.8)	11(1.2)	257(6.7)	9(1.0)	***(***)
Write notes	Grade 4	1994	30(1.5)	203(2.9)	14(1.0)	206(4.0)	17(1.1)	209(2.7)	39(1.8)	206(2.5)
		1984	31(2.2)	203(3.1)	13(1.4)	204(4.3)	13(1.4)*	212(5.3)	43(2.3)	201(2.3)
	Grade 8	1994	46(1.6)	269(1.7)	19(1.3)	268(2.8)	17(1.4)	265(3.7)	17(1.4)	260(3.7)
		1984	38(1.9)*	267(3.2)	19(1.4)	266(4.0)	19(1.9)	265(3.6)	24(1.6)*	260(4.4)
Point out what is well done	Grade 4	1994	53(1.9)	206(2.1)	16(0.9)	205(3.0)	12(1.0)	206(4.5)	19(1.4)	202(3.8)
		1984	54(2.4)	203(2.4)	18(1.3)	205(4.6)	9(1.2)	202(6.0)	19(1.7)	204(4.0)
	Grade 8	1994	50(1.7)	269(2.0)	23(1.4)	266(2.4)	13(1.1)	265(4.4)	14(1.1)	261(3.4)
		1984	39(2.5)*	268(2.5)	23(1.7)	266(3.8)	18(1.3)*	265(4.1)	19(1.4)*	259(5.6)
Point out what is not well done	Grade 4	1994	43(1.6)	206(2.1)	18(1.0)	206(3.8)	18(1.1)	207(3.4)	22(1.1)	204(3.5)
		1984	50(1.9)*	204(2.2)	17(1.8)	208(4.9)	13(1.5)*	198(4.6)	20(1.6)	202(4.7)
	Grade 8	1994	57(1.6)	267(2.0)	19(0.9)	265(2.7)	13(0.9)	270(3.2)	12(0.9)	264(4.2)
		1984	54(2.9)	268(2.1)	19(1.9)	263(4.6)	14(1.6)	263(5.5)	14(1.5)	260(6.3)
Make suggestions for next time	Grade 4	1994	48(1.4)	206(2.4)	17(1.1)	209(3.4)	14(1.0)	207(3.4)	21(1.3)	202(3.1)
		1984	50(2.3)	202(2.7)	17(1.7)	203(3.9)	15(1.4)	210(6.0)	18(1.9)	203(4.2)
	Grade 8	1994	55(1.5)	266(1.9)	17(1.2)	266(3.8)	14(1.0)	270(3.7)	14(1.1)	264(3.9)
		1984	49(2.0)*	263(2.9)	21(1.5)*	271(4.2)	16(1.5)	266(4.2)	15(1.4)	264(5.0)
Show an interest in what you write	Grade 4	1994	59(1.7)	207(2.5)	17(1.1)	207(2.7)	11(0.9)	206(4.6)	13(0.9)	196(3.6)
		1984	61(2.2)	203(2.4)	17(1.9)	208(5.8)	11(1.3)	201(6.6)	11(1.4)	204(4.5)
	Grade 8	1994	56(2.1)	268(2.1)	19(1.4)	269(2.9)	13(0.9)	266(2.6)	13(1.0)	255(3.8)
		1984	49(2.7)	266(3.1)	20(1.6)	265(4.6)	15(1.5)	264(4.1)	16(1.8)	263(5.2)

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

In 1994, the average writing scores of eighth graders who indicated that, about half the time or more, their teachers showed an interest in what they wrote had higher average writing scores than those who reported that their teachers never or hardly ever showed an interest.

Trends in Computer Use

As educators explore the benefits of word processing and computer-delivered writing instruction, access to and use of computers becomes a key issue in America's schools.[76] Tables 13.12 and 13.13 summarize student responses to questions about the availability and use of computers at home and in school.

[76] Self, C.L.; Rodrigues, D.; and Oates, W.R., *Computers in English and the Language Arts* (Urbana, IL: National Council of Teachers of English, 1989).

McCurry, N.; and McCurry, A., "Writing Assessment for the 21st Century," *Computing Teacher,* 19(7), 35-35, 1992.

Table 13.12

Trends in Availability and Use of Computers, Grades 4, 8, and 11, 1984 to 1994

			STUDENTS' REPORTS ON COMPUTER USE					
			Grade 4		Grade 8		Grade 11	
		Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Use computer at home	Yes	1994	50(2.0)	209(3.1)	50(1.9)	274(2.2)	51(2.0)	291(2.6)
		1984	45(3.7)	203(4.5)	37(4.6)*	***(***)	30(2.9)*	295(3.6)
	No	1994	50(2.0)	202(4.2)	50(1.9)	261(2.6)	49(2.0)	276(2.9)
		1984	55(3.7)	202(3.9)	63(4.6)*	262(6.2)	70(2.9)*	289(3.0)*
Use computer at library	Yes	1994	48(3.0)	208(3.6)	57(3.3)	268(2.4)	61(2.6)	285(2.6)
		1984	25(3.8)*	***(***)	21(3.9)*	***(***)	22(2.9)*	***(***)
	No	1994	52(3.0)	203(4.6)	43(3.3)	264(3.0)	39(2.6)	281(3.7)
		1984	75(3.8)*	203(3.1)	80(3.9)*	264(6.2)	78(2.9)*	291(2.9)*
Use computer at friend's house	Yes	1994	43(2.6)	210(5.4)	43(2.0)	271(2.2)	39(2.5)	285(4.2)
		1984	39(4.5)	205(5.3)	43(5.1)	***(***)	32(2.9)	290(4.1)
	No	1994	58(2.6)	203(4.3)	57(2.0)	263(2.5)	61(2.5)	284(2.3)
		1984	61(4.5)	202(3.6)	57(5.1)	264(6.7)	68(2.9)	290(3.2)
Use computer to learn things	Yes	1994	82(1.8)	207(3.4)	76(2.0)	269(1.9)	71(2.2)	283(2.2)
		1984	68(3.1)*	204(3.0)	58(4.5)*	268(4.7)	55(3.0)*	292(2.9)*
	No	1994	18(1.8)	203(4.6)	24(2.0)	262(3.6)	29(2.2)	285(3.7)
		1984	32(3.1)*	200(5.3)	42(4.5)*	***(***)	45(3.0)*	290(4.2)
Use computer to play games	Yes	1994	87(1.8)	207(3.0)	87(1.4)	267(1.8)	77(2.0)	284(2.4)
		1984	72(3.7)*	204(3.4)	84(3.7)	264(4.2)	76(2.5)	292(2.7)*
	No	1994	13(1.8)	198(7.7)	13(1.4)	264(4.8)	23(2.0)	283(4.2)
		1984	28(3.7)*	202(5.1)	16(3.7)	***(***)	24(2.5)	288(4.6)
Use computer to write stories or papers	Yes	1994	68(2.6)	210(3.6)	82(1.7)	270(1.9)	87(2.0)	286(2.7)
		1984	23(3.7)*	***(***)	15(3.5)*	***(***)	19(2.2)*	***(***)
	No	1994	32(2.6)	198(4.3)	18(1.7)	254(5.0)	13(2.0)	272(4.1)
		1984	77(3.7)*	206(2.3)	85(3.5)*	261(5.7)	81(2.2)*	291(3.0)*

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Between 1984 and 1994, computer use increased at all grades and in various situations. At grades 4, 8, and 11, there was a rise in the percentage of students reporting that they used computers at the library, to learn things, and to write stories or papers. At grade 4, there was an increase in the percentage of students who reported using computers to play games. At grades 8 and 11, there was also an increase in the percentage of students who reported using a computer at home.

In 1994, fourth, eighth, and eleventh graders who reported using computers to write stories or papers had higher average writing scores than their peers who reported not using a computer for this purpose. Eighth and eleventh graders who reported using a computer at home had higher average scores than their peers who did not. Eighth graders who reported using a computer at a friend's house also had higher average writing scores than eighth graders who did not report doing so.

As indicated in Table 13.13, there has been a dramatic increase since 1984 in the use of computers in schools. In 1984, the majority of students reported never using computers at school — 61 percent at grade 4, 67 percent at grade 8, and 55 percent at grade 11. In 1994, these percentages had dropped to only 14 percent at grade 4, 28 percent at grade 8, and 26 percent at grade 11.

Table 13.13

Trends in Computer Use in School, Grades 4, 8, and 11, 1984 to 1994

		STUDENTS' REPORTS ON HOW OFTEN THEY USE A COMPUTER AT SCHOOL					
		Grade 4		Grade 8		Grade 11	
	Year	Percent	Average Score	Percent	Average Score	Percent	Average Score
Every day	1994	8(1.4)	***(***)	15(1.8)	266(4.7)	23(1.9)	285(3.4)
	1984	3(1.3)*	***(***)	4(1.5)*	***(***)	12(2.0)*	***(***)
Two or three times a week	1994	23(3.0)	206(4.6)	15(1.3)	266(4.6)	12(1.6)	280(5.5)
	1984	8(1.7)*	***(***)	5(1.9)*	***(***)	6(1.3)*	***(***)
Once a week	1994	40(3.0)	207(3.9)	16(1.8)	267(3.6)	8(1.4)	***(***)
	1984	16(3.1)*	***(***)	8(2.8)*	***(***)	6(1.5)	***(***)
Less than once a week	1994	16(1.4)	209(4.9)	27(2.1)	270(2.9)	31(2.0)	287(3.6)
	1984	13(1.8)	***(***)	17(3.6)*	***(***)	21(2.6)*	***(***)
Never	1994	14(2.0)	198(4.9)	28(2.7)	265(4.6)	26(2.4)	283(3.8)
	1984	61(3.5)*	199(4.3)	67(4.2)*	260(6.2)	55(2.9)*	289(3.6)

* Statistically significant difference from 1994, where alpha equals .05 per set of comparisons. The standard errors of the estimated percentages and proficiencies appear in parentheses. Significance tests for extreme percentages (either >90 or <10 percent) should be interpreted with caution. It can be said with 95-percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 due to rounding or the selection of "I don't know" by a small percentage of students.

*** Sample size insufficient to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Long-Term Trend Assessment.

Summary

Overall, students' reports on the value they place on writing and their attitudes toward writing reflected moderate increases from 1984 to 1994. For example, increased percentages were observed in the following responses: fourth graders who indicated they believe writing helps them to share their ideas or to get a job, eleventh graders who indicated that writing helps them to understand their own feelings, and eighth and eleventh graders who viewed themselves as good writers or thought that other people liked their writing.

Students' personal and social uses for writing have changed to some degree since 1984. In the most recent assessment year, eighth and eleventh graders were more frequently keeping diaries or journals and writing stories or poems. Eighth graders were also more frequently writing letters to friends or relatives.

Students' responses to questions about revising and editing strategies indicated a trend toward the use of more complex strategies. At all three grades, students were more frequently taking out parts of their writing that they did not like, but less frequently throwing out their writing and starting over. At grades 4 and 8, students indicated more frequent use of moving sentences or paragraphs as a revising strategy. At grade 8, changing words and adding ideas or information were strategies being used more frequently in 1994 than in 1984. And eleventh graders in 1994 were less likely to rewrite most of the paper as they were revising and editing than were their counterparts in 1984.

Changes were also observed over time in the frequency with which students were asked to complete different types of writing activities as a part of their English classes. For example, essays, compositions, or themes appeared to be more common in 1994 than in 1984 in the English classes of students in each grade. Also, letters and poems appeared to be more frequent in eighth- and eleventh-grade classes; and stories were apparently more frequent in fourth- and eighth-grade classes.

Students' responses also revealed some changes in the comments that teachers made on students' papers. Compared to 1984, fourth and eighth graders in 1994 were less likely to report that their teachers often marked mistakes. At grades 8 and 11, students were more likely to indicate that their teachers often commented on the ideas in their paper, the way their ideas were explained, and the way their feelings were expressed. Fewer fourth graders indicated that their teachers often pointed out what was not done well. More eighth graders reported that their teachers often commented on the amount of writing, provided written notes, and pointed out what was done well. According to eleventh graders, teachers often commented on organization and word use.

The most dramatic changes between 1984 and 1994 appeared to take place in the use of computers inside and outside of school. At all three grades, there was an increase in students' use of computers in school. Students in all three grades were also more likely in 1994 to use computers in a library, to learn things, and to write stories or papers. At grades 8 and 11, there was an increase in the percentage of students using computers at home. At grade 4, there was an increase in the percentage of students using computers to play games.

During the past decade, educators have advocated changes in students' attitudes towards writing, in the kinds of writing they do in and out of school, in the processes they engage in as they write, and in the ways teachers should respond to students' writing. NAEP data indicate that some positive changes have occurred in the value that some students place on writing, on the variety of ways in which some students use writing, in the strategies that students use for revising and editing, in the nature of some teachers' responses to student writing, and in students' access to and use of computers.

Procedural Appendix

Overview of Procedures Used in the 1994 Science, Mathematics, Reading, and Writing Trend Assessments

This appendix provides more detailed information about the methods and procedures used in NAEP's 1994 trend assessments. The *NAEP 1994 Technical Report* contains even more extensive information about these procedures.

This NAEP trend report is based on eight science assessments, seven mathematics assessments, eight reading assessments, and five writing assessments, with the most recent assessment in each of the four curriculum areas having been conducted during the 1993-94 school year. NAEP also conducted various cross-sectional assessments and state assessments in 1994. The composition of each of the trend assessments is described below.

Science

NAEP conducted trend assessments of the science achievement of in-school 9-, 13-, and 17-year-olds during the school years ending in 1970, 1973, 1977, 1982, 1986, 1990, 1992, and 1994. However, in the first assessment, 17-year-olds were assessed during the spring of the school year ending in 1969.

The science trend assessments measured student achievement based on objectives developed by nationally representative panels of scientists, science educators, and concerned citizens. The objectives for each successive assessment were based on the framework used for the previous assessment with some revisions that reflected changes in content and trends in school science. Although changes were made from assessment to assessment, some tasks were retained from one assessment to the next in order to measure trends in achievement across time. The objectives which formed the basis for the 1986 trend assessment (see 1985-1986 Science Objectives) replicated the objectives used in previous assessments, and these also governed the 1990, 1992, and 1994 assessments. The results of the 1994 trend assessment are described in this report. Results from an even more recent science assessment, conducted in 1996, are not presented herein. That assessment was based on a comprehensive new framework developed by the National Assessment Governing Board under a contract to the Council of Chief State School Officers.[77]

The 1994 science trend assessment contained 63 multiple-choice questions at age 9, 83 multiple-choice questions at age 13, and 82 multiple-choice questions at age 17. The assessment covered a range of science content areas, including topics from the life sciences, physical sciences, and earth and space sciences. Questions assessed students' abilities to understand basic scientific facts and principles, solve problems in scientific contexts, design experiments, interpret data and read tables and graphs, and understand the nature of science.

---

[77] National Assessment Governing Board. *Science Framework for the 1996 National Assessment of Educational Progress*. (Washington, DC: National Assessment Governing Board, 1994).

Mathematics

NAEP has assessed the mathematics achievement of in-school 9-, 13-, and 17-year-olds seven times: in the school years ending in 1973, 1978, 1982, 1986, 1990, 1992, and 1994. The long-term trend assessment, which forms the basis of the results detailed in this report, uses procedures established in 1973. Results from an even more recent mathematics assessment, conducted in 1996, are not presented herein. The 1996 assessment continues a short-term trend mathematics assessment which was also conducted in 1990 and 1992. The framework for the 1996 mathematics assessment was developed by the National Assessment Governing Board under a contract to the College Board.[78]

The long-term trend assessments contained a range of constructed-response and multiple-choice questions measuring performance on sets of objectives developed by nationally representative panels of mathematics specialists, educators, measurement experts, and other interested parties.[79] A set of the 1986 questions was readministered in 1990, 1992, and 1994. This set of questions contains 127 multiple-choice and 34 constructed-response questions at age 9, 158 multiple-choice and 56 constructed-response questions at age 13, and 231 multiple-choice and 56 constructed-response questions at age 17.

The questions covered a range of content, including numbers and operations, measurement, geometry, and algebra. The process areas included knowledge, skills, application, and problem solving.

---

[78] National Assessment Governing Board. Mathematics Framework for the 1996 National Assessment of Educational Progress. (Washington, DC: National Assessment Governing Board, 1994).

[79] *Math Objectives: 1985-86 Assessment* (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1986).

Reading

NAEP has assessed students' reading performance at ages 9, 13, and 17 in eight national reading assessments conducted during the school years ending in 1971, 1975, 1980, 1984, 1988, 1990, 1992, and 1994.[80]

The reading tasks included in the trend assessment asked students to read and answer questions based on a variety of materials, including informational passages, literary text, and documents. Although some tasks required students to provide written responses, most questions were multiple choice. The assessment was designed to evaluate students' ability to locate specific information, make inferences based on information in two or more parts of a passage, or identify the main idea in a passage. For the most part, these questions measured students' ability to read either for specific information or for general understanding. Although the reading assessments conducted through the 1970s underwent some changes from test administration to administration, the set of reading passages and questions included in the trend assessments has been kept essentially constant since 1984, and most closely reflects the objectives developed for that assessment.[81] The reading trend assessment administered at age 9 included 54 passages and 118 questions, including five that required students to construct written responses. At age 13, the assessment included 40 passages and 94 questions, eight of them requiring constructed responses. At age 17, the assessment contained 34 passages and 112 questions, nine of them requiring constructed responses.

---

[80] NAEP also conducted a reading assessment in 1986, but the resulting estimates of students' reading performance appeared anomalous, and the trend results were not disseminated to the general public. Concern about these apparently anomalous results prompted a thorough investigation of the NAEP technology by ETS/NAEP staff, which was reported in *Disentangling the NAEP 1985-86 Reading Anomaly: A Technical Report*. An independent technical review panel was convened by NCES, and its findings were summarized in *Report of the NAEP Technical Review Panel on the 1986 Reading Anomaly, the Accuracy of NAEP Trends, and Issues Raised by State-Level NAEP Comparisons.* As part of the 1988 assessment, NAEP conducted a study to provide further information about the 1986 reading anomaly. Analyses of data collected in the study revealed some, but not all, of the reasons for the unusual assessment results in 1986. Further information on this issue is available in *Disentangling the NAEP 1985-86 Reading Anomaly: A Technical Report* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, 1989).

[81] *Reading Objectives: 1983-84 Assessment* (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1984).

Results from a short-term trend reading assessment, conducted in 1992 and 1994, are not presented in this report. The 1992 and 1994 short-term trend assessments were based on a different framework and procedures. The framework for these assessments was developed by the National Assessment Governing Board under a contract to the Council of Chief State School Officers.[82]

Writing

NAEP has assessed the writing performance of students in grades 4, 8, and 11 in the school years ending 1984, 1988, 1990, 1992, and 1994. In all five assessments, the same tasks were administered in the same manner to comparable samples of students. The writing tasks and background questions were designed to measure aspects of writing performance and related factors that were designated as important by a nationally representative panel of writing specialists, educators, and concerned citizens. The primary objective of the trend assessment was to measure students' ability to write for various purposes; related objectives were to evaluate the extent to which students managed the writing process, controlled the forms of written language, and valued writing.[83] At each grade, six different writing tasks were administered.

---

[82]National Assessment Governing Board. *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

[83]*Writing Objectives: 1988 Assessment* (Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service, 1987).

The Design of the Science and Mathematics Trend Assessments

At each of the three ages assessed, both the science and mathematics trend assessments consisted of three different 15-minute segments or "blocks" of content questions, each also containing a small set of background questions that pertained to students' experiences and instruction with the particular subject area being assessed (i.e., either science or mathematics).

The blocks were assembled three to a booklet, together with a background questionnaire that was common to all booklets. This questionnaire included questions about demographic information as well as home environment.

At ages 9 and 13, the blocks were placed in three booklets, each containing one block of mathematics questions, one block of science questions, and one block of reading questions. The reading block in these booklets is not used in the reading trend assessment, but is included in order to preserve the context effects of previous assessments. To replicate past procedures, at age 17, two booklets were used. One contained two mathematics blocks and one science block, while the other contained two science blocks and one mathematics block.

At all three ages, to replicate past procedures, the science and mathematics questions were administered using a paced audiotape. The tape recording that accompanied the booklets standardized timing and was intended to help students with any difficulty they might have in reading the questions. Thus, in an administration session, all students were being paced through the same booklet.

The Design of the Reading and Writing Trend Assessments

The reading trend assessment consisted of ten 15-minute blocks of reading passages and questions at each of three age/grade levels, while the writing trend assessment included five 15-minute blocks. Each writing block contained one prompt, except one block which contained two short prompts, for a total of six prompts. In addition, each content block contained a short set of background questions. The background questions in the reading blocks pertained to students' reading habits and experiences, while those in the writing blocks asked about students' writing practices, instruction, and attitudes.

In keeping with procedures used in previous reading and writing trend assessments, the reading and writing blocks were assembled into six booklets at each age/grade assessed, with each pair of reading or writing blocks appearing in one of the booklets. Each student participating in the reading and writing assessments received a booklet containing three content blocks as well as a six-minute section of background questions about demographic information and the students' home environment.

Sampling and Data Collection

Sampling and data collection activities for the 1994 trend assessments were conducted by Westat, Inc. Based on procedures used since the inception of NAEP, the data collection schedule was 13-year-olds/eighth graders in the fall (October to December, 1993), 9-year-olds/fourth graders in the winter (January to mid-March, 1994), and 17-year-olds/eleventh graders in the spring (mid-March to May, 1994). Although only 9-, 13-, and 17-year-olds were assessed in science and mathematics, both age- and grade-eligible students were assessed in reading and writing. Age eligibility was defined by calendar year for 9- and 13-year olds, while the birth date range for 17-year-olds was from October, 1976 through September 30, 1977.

As with all NAEP national assessments, students attending both public and nonpublic schools were selected for participation based on a stratified, three-stage sampling plan. The first and second stages included defining geographic primary sampling units (PSUs), which are typically groups of contiguous counties, but sometimes a single county; classifying the PSUs into strata defined by region and community type; and randomly selecting schools, both public and nonpublic, within each PSU selected at the first stage. The third stage involved randomly selecting students within a school for participation. Some students selected (fewer than 8 percent) were excluded because of limited English proficiency or severe disability.

The student sample sizes for the trend assessments as well as the school and student participation rates are presented in the following tables. Because students within schools were randomly assigned to either mathematics/science or reading/writing assessment sessions subsequent to their selection for participation in the 1994 assessments, the school and student participation rates shown are for all four subject areas combined. Based on the sampling design, these rates are also the best estimates for each individual subject area. They are included in the individual tables for each subject area for convenience in comparing across assessment years. For assessments conducted prior to 1984, the school and student participation rates were obtained from the Public Use Data Tape User Guides. Figures for more recent assessments then were obtained from the reports on the NAEP field operation and data collection activities, prepared by Westat, Inc. Although sampled schools that refused to participate were replaced, school cooperation rates were computed based on the schools originally selected for participation in the assessments. The student participation rates represent the percentage of students assessed of those invited to be assessed, including in follow-up sessions when necessary.

TABLE A.1

Student Sample Sizes for the Science Trend Scaling

	1977	1982	1986	1990	1992	1994
Age 9	17,345	1,960	6,932	6,235	7,335	5,663
Age 13	25,653	7,873	6,200	6,649	5,909	6,052
Age 17 (in school)	31,436	7,974	3,868	4,411	4,359	3,813
Total	74,434	17,817	17,000	17,295	17,603	15,528

TABLE A.2

School and Student Participation Rates for the Science Trend Assessments

	Age	Weighted Percentage of Schools Participating	Weighted Percentage of Students Participating
1970	9	–	88.0
	13	–	85.6
	17	–	74.5
1973	9	93.9	91.0
	13	93.8	84.6
	17	92.4	73.6
1977	9	91.5	88.6
	13	91.3	86.2
	17	89.5	73.1
1982	9	88.3	90.5
	13	89.2	85.5
	17	86.5	74.2
1986	9	88.7	92.9
	13	88.1	89.2
	17	82.7	78.9
1990	9	87.0	92.5
	13	89.0	90.2
	17	79.0	82.1
1992	9	87.8	94.4
	13	85.6	90.9
	17	81.0	82.3
1994	9	87.1	94.4
	13	80.4	92.3
	17	79.5	84.8

TABLE A.3

Student Sample Sizes for the Mathematics Trend Scaling

	1978	1982	1986	1990	1992	1994
Age 9	14,752	12,038	6,932	6,235	7,335	5,663
Age 13	24,209	15,758	6,200	6,649	5,909	6,052
Age 17 (in school)	26,756	16,319	3,868	4,411	4,359	3,813
Total	65,717	44,115	17,000	17,295	17,603	15,528

TABLE A.4

School and Student Participation Rates for the Mathematics Trend Assessments

	Age	Weighted Percentage of Schools Participating	Weighted Percentage of Students Participating
1973	9	93.9	90.9
	13	93.8	84.2
	17	92.4	73.5
1978	9	91.5	87.2
	13	91.5	85.2
	17	89.5	73.2
1982	9	88.3	90.5
	13	89.2	85.5
	17	86.5	74.2
1986	9	88.7	92.9
	13	88.1	89.2
	17	82.7	78.9
1990	9	87.0	92.5
	13	89.0	90.2
	17	79.0	82.1
1992	9	87.8	94.4
	13	85.6	90.9
	17	81.0	82.3
1994	9	87.1	94.4
	13	80.4	92.3
	17	79.5	84.8

TABLE A.5

Student Sample Sizes for the Reading Trend Scaling

	1971	1975	1980	1984	1988	1990	1992	1994
Age 9	23,201	21,697	21,159	22,291	3,782	4,268	4,944	5,335
Age 13	25,545	21,393	22,330	22,693	4,005	4,609	3,965	5,547
Age 17 (in school)	23,661	19,624	18,103	25,193	3,652	4,383	4,447	4,840
Total	72,407	62,714	61,592	70,177	11,439	13,260	13,356	15,722

TABLE A.6

School and Student Participation Rates for the Reading Trend Assessments

	Age	Weighted Percentage of Schools Participating	Weighted Percentage of Students Participating
1971	9	92.5	90.9
	13	92.0	84.2
	17	90.5	73.5
1975	9	93.9	87.2
	13	92.8	85.2
	17	91.0	73.2
1980	9	94.5	90.5
	13	93.2	85.5
	17	90.5	74.2
1984	9	88.6	92.9
	13	90.3	89.2
	17	83.9	78.9
1988	9	87.2	92.5
	13	92.7	90.2
	17	78.1	82.1
1990	9	87.0	92.5
	13	89.0	90.2
	17	79.0	82.1
1992	9	87.0	93.8
	13	85.3	90.8
	17	80.9	83.3
1994	9	86.7	94.1
	13	79.7	91.8
	17	80.1	84.2

TABLE A.7

Sample Sizes for the Writing Trend Assessments by Task and Scoring Method

		1984 GRADE			1988 GRADE			1990 GRADE			1992 GRADE			1994 GRADE		
Writing Task	Scoring Method	4	8	11	4	8	11	4	8	11	4	8	11	4	8	11
Informative																
Plants	Primary	656	–	–	1285	–	–	1416	–	–	1677	–	–	1347	–	–
XYZ Company	Primary	544	616	–	1152	1334	–	1288	1489	–	1583	1333	–	1217	1313	–
Appleby House	Primary	530	588	599	925	1256	1041	1111	1396	1277	1337	1249	1264	1013	1255	1141
Food on the Frontier	Primary	–	603	629	–	1339	1212	–	1503	1401	–	1316	1447	–	1321	1305
Food on the Frontier	Holistic	–	2233	2373	–	1339	1209	–	1500	1399	–	1308	1436	–	1341	1269
Job Application	Primary	–	–	603	–	–	1169	–	–	1424	–	–	1403	–	–	1214
Persuasive																
Spaceship	Primary	611	–	–	1258	–	–	1367	–	–	1653	–	–	1313	–	–
Spaceship	Holistic	2025	–	–	1256	–	–	1359	–	–	1574	–	–	1312	–	–
Spaceship	Mechanics	–	–	–	481	–	–	–	–	–	678	–	–	521	–	–
Radio Station	Primary	585	612	–	1234	1364	–	1386	1512	–	1650	1362	–	1303	1359	–
Dissecting Frogs	Primary	–	641	–	–	1356	–	–	1518	–	–	1359	–	–	1331	–
Rec. Opportunities	Primary	–	494	521	–	1372	1242	–	1498	1415	–	1317	1416	–	1301	1272
Rec. Opportunities	Holistic	–	2228	2354	–	1364	1238	–	1496	1411	–	1309	1406	–	1319	1191
Rec. Opportunities	Mechanics	–	473	517	–	516	497	–	–	–	–	563	566	–	–	509
Space Program	Primary	–	–	632	–	–	1195	–	–	1451	–	–	1427	–	–	1276
Bike Lane	Primary	–	–	636	–	–	1178	–	–	1424	–	–	1425	–	–	1282
Imaginative																
Flashlight	Primary	609	–	–	614	–	–	702	–	–	850	–	–	656	–	–
Flashlight	Holistic	2015	–	–	611	–	–	697	–	–	840	–	–	671	–	–

TABLE A.8

School and Student Participation Rates for the Writing Trend Assessments

	GRADE	PERCENTAGE OF SCHOOLS PARTICIPATING	PERCENTAGE OF STUDENTS PARTICIPATING
1984	4	88.6	92.5
	8	90.3	90.3
	11	83.9	82.2
1988	4	87.2	92.3
	8	92.7	88.2
	11	78.1	77.4
1990	4	87.0	92.5
	8	89.0	90.2
	11	79.0	82.1
1992	4	87.0	93.8
	8	85.3	90.8
	11	80.9	83.3
1994	4	86.7	94.1
	8	79.7	91.8
	11	80.1	84.2

Scoring the Booklets

Materials from NAEP 1994 assessments, including the trend assessments, were shipped to National Computer Systems (NCS) in Iowa City, Iowa, for processing. Receipt and quality control were managed through a sophisticated bar-coding and tracking system. After all appropriate materials were received from a school, they were forwarded to the professional scoring area, where the responses to the constructed-response questions were evaluated by trained staff using guidelines prepared by NAEP. Each constructed-response question had a unique scoring guide that defined the criteria to be used in evaluating students' responses. Subsequent to the professional scoring, the booklets were scanned, and all information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. An overview of the professional scoring for mathematics, reading, and writing follows (no constructed-response questions were scored for science).

Scoring the Mathematics Constructed-Response Questions

Most of the constructed-response mathematics trend questions were scored on a correct/incorrect basis. The scoring guides identified the correct or acceptable answers for each question in each block. The scores for these questions included a 0 for no response, a 1 for a correct answer, or a 2 for an incorrect or "I don't know" response. Because of the straightforward nature of the scoring, lengthy training was not required. In an orientation period, the readers were trained to follow the procedures for scoring the mathematics questions and given an opportunity to become familiar with the scoring guides, which listed the correct answer for the questions in each of the blocks.

During the scoring, every tenth booklet in a session was scored by a second reader to provide a quality check. These quality checks were recorded on a separate sheet with the few discrepancies noted, and the scores were corrected. For the most part, this entailed providing a score because one had not been coded. In total, over 380,000 answers were read and classified, including approximately 130,000 responses at age 9, 100,000 at age 13, and 150,000 at age 17.

Scoring the Reading Constructed-Response Questions

The 1994 reading trend assessment included five questions at age 9 for which students were required to construct written responses, seven such questions at age 13, and eight such questions at age 17. Some of the questions were administered to more than one age group of students.

The scoring guides for the constructed-response reading questions focused on students' ability to perform various reading tasks — for example, identifying the author's message or mood and substantiating their interpretations, making predictions based on given details, supporting an interpretation, and comparing and contrasting information.

The scoring guides for the reading questions varied somewhat, but typically included the distribution of score points shown below.

Outline for Scoring of Constructed-Response Reading Trend Assessment Questions

Rating Category	
4	**ELABORATED REFERENCE OR INTERPRETATION.** These responses exceeded the requirements of the task by including illustrative examples or details and demonstrating a high level of cohesiveness.
3	**SATISFACTORY REFERENCE OR INTERPRETATION.** These responses identified at least two relevant examples or reasons to support a given interpretation.
2	**MINIMAL REFERENCE OR INTERPRETATION**. These responses did not provide evidence to support a stated interpretation.
1	**UNSATISFACTORY REFERENCE OR INTERPRETATION.** These responses did not provide an interpretation, but instead digressed or avoided the task.
0, 7, 8, 9	These responses were, respectively, blank, indecipherable, completely off-task, or included a statement to the effect that the student did not know how to do the task. (In the analysis, scores of 7, 8, and 9 were collapsed into the score point of 9).

Some of the scoring guides included secondary scores, which typically involved categorizing the kind of evidence or details the student used as support for an interpretation. The document literacy tasks, most of which required short answers, were scored on a correct/incorrect basis.

The training program for the reading trend assessment scoring was carried out on all assessment questions one at a time for each age group and covered the range of student responses. Because the purpose of the scoring was to measure trends from the 1984 assessment, preparation for training included rereading hundreds of 1984 responses and compiling training sets. In order to ensure continuity with the past scoring of the trend questions, at least half of the sample papers in the training sets were taken from the 1984 training sets, and previously scored 1984 booklets were masked to ensure that scoring for training and the subsequent trend reliability scoring would be done without knowledge of the previous scores given.

The actual training was conducted by ETS staff assisted by NCS's scoring director and team leaders. Training began with each reader receiving a photocopied packet of materials consisting of a scoring guide, a set of 15 to 20 scored samples, and an additional 20 to 40 response samples to be scored. The trainers reviewed the scoring guide, explained all the applicable score points, and elaborated on the rationale used to arrive at a particular score. The readers then reviewed the 15 to 20 scored samples, as the trainers clarified and elaborated on the scoring guide. After this explanation, the additional samples were scored and discussed until the readers were in agreement. If necessary, additional packets of 1984 responses were used for practice scoring.

As a further step to achieve reliability with 1984, a 25 percent sample of the 1984 responses was scored on separate scoring sheets following the formal training session. These sheets were key entered, and a computerized report was generated comparing the new scores with those assigned in 1984. After some further discussion, scoring of the 1994 responses began. Two reliability studies were conducted as part of this scoring. For the 1994 material, 25 percent of the constructed responses were scored by a second reader to produce interreader reliability statistics. In addition, a trend reliability study was conducted by rereading 20 percent of the 1984 responses. The reliability information is shown in Table A.9.

Table A.9

Percent Exact Agreement Between Readers: Reading Trend Assessment Scoring

	1984 RESPONSES RESCORED IN 1994		1994 RESPONSES SCORED TWICE	
Age	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
9	87.4	85.0-90.0	90.4	85.9-94.0
13	84.0	81.9-85.6	82.4	75.4-86.0
17	86.5	81.3-91.7	83.7	74.4-93.5

Note: The reading scoring was generally based on 5 scoring categories.

Scoring the Writing Tasks

Primary Trait Scoring. A primary trait scoring guide was developed for each writing task to focus raters' attention on how successfully students' responses accomplished the task set forth in the prompt. As illustrated in the introduction to Part IV of this report, the scoring guides typically defined five levels of task accomplishment — not rated, unsatisfactory, minimal, adequate, and elaborated — based on the rhetorical demands of the task. (A few of the scoring guides did not define an "elaborated" category as it was not appropriate to do so given the nature of the task.)

Because the results for the 1984 and 1988 writing trend assessments were based on a 1988 scoring of both 1984 and 1988 papers, the undertaking for writing trend scoring in 1990, 1992, and 1994 involved replicating the standards used in 1988. The procedure for training readers proceeded as outlined above for the reading trend assessment scoring, except that the writing scorers were trained using 1988 sample papers and practiced with a 25 percent sample of 1988 responses. As part of the scoring, two reliability studies were conducted. For the 1994 responses, 25 percent of the papers were scored by a second reader to produce interrater reliability statistics. In addition, a trend reliability study was conducted to ensure that the scoring procedures were consistent with those used in 1988. The results of these studies are presented in Table A.10.

Table A.10

Percent Exact Agreement Between Readers for Primary Trait Scoring: Writing Trend Assessment Scoring

	1988 PAPERS RESCORED IN 1994		1994 PAPERS SCORED TWICE	
Grade	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
4	90.4	86.2-95.0	89.0	81.0-94.9
8	87.7	82.6-94.1	88.1	79.2-95.6
12	91.2	88.1-93.2	83.1	77.0-87.4

Note: The primary trait scoring was based on 5 scoring categories.

Holistic Scoring. To offer another perspective on students' writing abilities, selected tasks included in the trend assessment were scored holistically for overall fluency (i.e., a global view of the ideas, language facility organization, mechanics, and syntax of each paper taken as a whole). As previously noted, these tasks were "Spaceship" and "Flashlight" at grade 4, and "Recreation Opportunities" and "Food on the Frontier" at grades 8 and 11. Trained readers evaluated the relative fluency of students' writing on a 6-point scale. A small percentage of papers — such as those that were blank or undecipherable — were not rated.

The holistic scale was anchored by a chief reader and assistant chief reader chosen for their expertise in holistic scoring. They, together with the table leaders and ETS staff members, studied the pool of 1994 student responses to select papers that represented each point on the holistic scale, and then used these sample papers to train the raters. In addition, for each task, a random sample of 50 papers from across the prior assessment years was drawn and evaluated by the group for use as practice papers in the training. Using the sample papers as a guide, the readers were asked to determine whether papers corresponded to the top half or the bottom half of the holistic scale and then to make finer distinctions between adjacent points on the scale. Because the emphasis of the holistic scoring was to detect trends across time at each of the three grade levels assessed, when a task was given at more than one grade level, responses were rated separately for each grade. A training session preceded the scoring of responses to each task at each grade level.

Because student papers are evaluated relative to one another in holistic scoring — rather than against specific criteria, as with primary trait scoring — the distribution of scores for the total sample of papers should be approximately normal, with scores evenly distributed around the center of the scale. To detect changes in writing fluency across time at each grade level, papers from the 1984, 1988, 1990, 1992, and 1994 assessments were randomly mixed prior to scoring. Thus, if more papers from one or another assessment were judged to be in the "top half" of the scale, the results would indicate changes across time in overall writing fluency. Twenty percent of the papers scored holistically were scored again by a second reader to provide information on interrater scoring agreement. These data are presented in Table A.11.

Table A.11

Writing Trend Assessment Percent Agreement for Adjacent Scores for Holistic Scoring of the 1984, 1988, 1990, and 1992 Papers Conducted in 1994

	1994 HOLISTIC SCORING		
	4	8	11
Spaceship	97.9	–	–
Flashlight	99.3	–	–
Recreation Opportunities	–	94.8	97.3
Food on the Frontier	–	97.6	97.5

Note: The holistic scoring was based on 7 scoring categories. Adjacent scores did not differ by more than one category.

Since certain writing tasks included in the writing trend assessments were submitted to both holistic and primary trait scoring, it is also possible to examine the relationship between the two sets of scores. As shown in Table A.12, the correlations range from .34 to .70. While the two scoring measures are clearly related, it is evident that they capture somewhat different aspects of writing performance. The primary trait score is closely tied to the features of specific writing tasks, providing a measure of students' success in accomplishing the assigned purpose of the writing. Alternatively, the holistic score provides a general measure of writing fluency, since the impression marks that raters give are affected by writers' attention to organization, adherence to the conventions of written English, word choice, handwriting, and quality of ideas.

Table A.12

Correlation Coefficients Between Primary Trait and Holistic Scores

	1984 PAPERS			1988 PAPERS			1990 PAPERS			1992 PAPERS			1994 PAPERS		
	4	8	11	4	8	11	4	8	11	4	8	11	4	8	11
Spaceship	.66	–	–	.70	–	–	.68	–	–	.59	–	–	.72	–	–
Flashlight	.64	–	–	.59	–	–	.66	–	–	.63	–	–	.63	–	–
Recreation Opportunities	–	.34	.48	–	.39	.44	–	.38	.44	–	.42	.46	–	.46	.54
Food on the Frontier	–	.43	.43	–	.46	.38	–	.49	.45	–	.52	.44	–	.55	.61

Mechanics Scoring. To provide for an examination of trends in students' control of the conventions of written English, NAEP evaluated a random subsample of the 1994 writing responses using the mechanics scoring criteria it used to evaluate writing responses from the 1984, 1988, 1990, and 1992 assessments.[84] One task at each grade level was selected for the mechanics scoring; these tasks were "Spaceship" at grade 4 and "Recreation Opportunities" at grades 8 and 11. A random probability sample of approximately 600 responses to each task at each grade level was selected for evaluation. To ensure that the comparisons between Black and White students were reasonably precise, responses of Black students were oversampled. Readers were trained by practicing on a 10 percent sample of the 1992 papers. Another 10 percent sample of essays previously scored for mechanics from the 1988, 1990, and 1992 assessments was rescored for reliability. A comparison of the 1994 data with the original scores indicated a between-year reliability ranging from .81 to .86 across the three grade levels.

In the mechanics scoring, each response was analyzed for a variety of aspects of spelling, punctuation, grammar, word choice, and syntax by English teachers who had been trained in the use of detailed criteria. The entire text of the scored papers, with the scoring marks, was then entered into a computer-readable database to provide for the subsequent analyses.

An outline of the features of writing mechanics included in the scoring and analysis is provided below.

I. Sentence Types

1. Simple — A sentence that contains a subject and a verb. It may also have an object subject complement, phrase, appositive, nominative absolute, or verbal. Also, a word group used in dialogue, for emphasis, or as an exclamation that is not an independent clause.
2. Compound — A sentence containing two or more simple sentences joined by something other than a comma.
3. Complex (and compound-complex) — A sentence that contains at least one independent clause and one dependent clause.

---

[84] Applebee, A. N.; Langer, J. A.; and Mullis I. V. S., *Grammar, Punctuation, and Spelling: Controlling the Conventions of Written English* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, 1987).

4. Run-On Sentence
   a. Fused — A sentence containing two or more independent clauses with no punctuation or conjunction separating them.
   b. On and on — A sentence consisting of four or more independent clauses strung together with conjunctions.
   c. Comma splice — A sentence containing two or more independent clauses separated by a comma instead of a semicolon or a coordinating conjunction.
5. Fragment — A word group, other than an independent clause, written and punctuated as a sentence.

II. Faulty Sentence Construction

(These scores are in addition to the sentence types.)

1. Agreement Error — A sentence in which at least one of the following occurs: subject/verb do not agree, pronoun/antecedent do not agree, noun/modifier do not agree, subject/object pronoun is misused, or verb tense shifts.
2. Awkward Sentence (The awkward categories are listed in order of category precedence, since only one score was given to a sentence.)
   a. Faulty parallelism — A parallel construction that is semantically or structurally dysfunctional.
   b. Unclear pronoun reference — A pronoun's antecedent is unclear.
   c. Illogical construction — Faulty modification or a dangling modifier or a functionally misarranged or misproportioned sentence.
   d. Other dysfunctions — A sentence containing an omitted or extra word or a split construction that definitely detracts from readability.

III. Punctuation Errors

Every error of commission and error of omission was coded for commas, dashes, quotation marks, semicolons, apostrophes, and end marks. The most informal rules of usage were used, with the writer receiving the benefit of any doubt.

IV. Word-Level Conventions

1. Word Choice — The writer needs a word that is different from the one written. This category also includes attempts at a verb, adjective, or adverb form that is nonexistent or unacceptable.
2. Spelling — In addition to misspellings, this category includes word-division errors at the end of a line, two words written as one, one word written as two, superfluous plurals, and groups of distinguishable letters that do not make a legitimate word.
3. Capitalization — A word is given a capitalization error score if the first word in a sentence is not capitalized, if a proper noun or adjective within a sentence is not capitalized, and if the pronoun "I" is not capitalized.

The mechanics scoring was designed to allow the writer as much flexibility as possible under existing conventions of correct writing; consequently, any time two authorities on mechanics disagreed, the more informal interpretation was used.

Because the papers were entered into a computer-readable database, the number of words per paper, number of words per sentence, and number of letters per word were tabulated by computer.

As discussed in Chapter 12, the agreement errors increased significantly in 1994. In checking that the statistics for agreement errors in 1994 were correct, the assessment coordinator read all of the papers of students whose responses contained agreement errors in more than 50 percent of the sentences. This was done to determine whether the correct number of agreement errors had been recorded into the database. In all cases, scorers had accurately identified agreement errors, albeit in a small number of papers. Thus, a real increase in total agreement errors stemmed from comparatively few student responses.

Data Analysis and IRT Scaling

After the assessment information had been compiled in the NAEP database, the data were weighted according to the population structure. The weighting for the samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse. Through poststratification, the weighting assured that the representation of certain subpopulations corresponded to figures from the U.S. Census and the Current Population Survey.

Analyses were then conducted to determine the percentage of students who gave various responses to each cognitive and background question. Item response theory (IRT) was used to estimate average proficiency for the nation and various subgroups of interest within the nation. IRT scaling was performed separately within each age/grade level for each of the four trend assessments (science, mathematics, reading, and writing). Each of the four assessments employs slightly different steps in data analysis and IRT scaling. The steps for each subject area are described in detail in the *NAEP 1994 Technical Report*. Because these descriptions are rather lengthy they are not repeated in this appendix.

IRT models the probability of answering a question correctly as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups, such as those defined by age, assessment year, or subpopulations (e.g., race/ethnicity or gender).

Students do not receive enough questions about a specific topic to permit reliable estimates of individual performance. Traditional test scores for individual students, even those based on IRT, would contribute to misleading estimates of population characteristics, such as subgroup averages and percentages of students at or above a certain proficiency level. Instead, NAEP constructs sets of plausible values designed to represent the distribution of proficiency in the population.[85] A plausible value for an individual is not a scale score for that individual but may be regarded as a representative value from the distribution of potential scale scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the NAEP scales are based on these plausible values. These statistics estimate values that would have been obtained had individual proficiencies been observed — that is, had each student responded to a sufficient number of cognitive questions so that his or her proficiency could be precisely estimated.

---

[85] For theoretical justification of the procedures employed, see Robert J. Mislevy, "Randomization-Based Inferences About Latent Variables from Complex Samples," *Psychometrika*, 56 (2), 177-96, 1988).

For computational details, see *1994 NAEP Technical Report.*

The reading trend scale was constructed based on the 1984 assessment and included all previous reading assessments. The science and mathematics trend scales were developed based on the 1986 science and mathematics assessments, respectively, and also included previous assessments. For the 1994 mathematics, reading, and science trend assessments, separate IRT scales were constructed within each grade. These scales were linked to the previously established scales within each subject area via a common population linking procedure.[86] The initial trend scaling, however, did not include the 1969-70 or 1973 science assessments, or the 1973 mathematics assessment. To provide a link to the early assessments for the nation and for subgroups defined by race/ethnicity, gender, and region at each of three age levels, estimates of average scale scores were extrapolated from previous analyses.

The extrapolated estimates were obtained by assuming that within a given age level the relationship between the logit of a subgroup's average p-value (i.e., average proportion correct) and its respective scale score average was linear and that the same line held for all assessment years and for all subgroups within the age level. Under this assumption, the between-year difference of the scale score averages of a subgroup for a pair of assessment years is equal to a constant (B) times the between-year difference of the logits of the average p-value of that subgroup for the same two years. For each age level, an average p-value estimated using a common set of questions was available for adjacent assessments. For science, these assessments included 1970 to 1973, 1973 to 1977, and 1977 to 1982. The adjacent assessments used for mathematics were 1973, 1978, and 1982. Then, using science as an example, the constant B was estimated by a regression (through the origin) of the difference between scale score averages in 1977 and 1982 on the corresponding difference between the logits of the average p-values for these two years. All subgroups in a given age were included in the regression. For example, the estimate of the 1973 scale score average for a subgroup was then obtained as the sum of the 1977 subgroup scale score average and B times the difference between the logits of the 1973 and 1977 subgroup average p-values (for questions common to 1973 and 1977). The same procedure was used for extrapolating the 1973 average p-value results on the mathematics IRT scale. The only difference was that the average p-value results were for the 1973, 1978, and 1982 assessments in mathematics, rather than for the 1973, 1977, and 1982 assessments. For science, however, results were extrapolated for two

[86] Yamamoto, K. & Mazzeo, J., "Item Response Theory Scale Linking in NAEP," *Journal of Educational Statistics,* Vol. 17, 155-73, 1992.

assessments. Therefore, after estimating the 1973 subgroup scale score average, the 1970 scale score average for the subgroup was estimated by the 1973 scale score average estimate plus B times the difference between the logits of the 1970 and 1973 subgroup average p-values (for questions common to 1970 and 1973).

For the trend writing assessments, a scale ranging from 0 to 500 was created, using a generalized partial-credit (GPC) model.[87] Developed by ETS and first used in 1992, the generalized partial-credit model permits the scaling of tasks scored according to multi-point rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex performance tasks. A separate IRT scale was constructed within each grade. These three within-grade scales were then linked together based on common tasks using the Stocking-Lord transformation.[88] The metric of the resulting linked scales was set to an average of 250 and a standard deviation of 50 across the three grades.

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

Scale Anchoring Analysis

To facilitate interpretation of the NAEP results, the scales are divided into successive levels of performance and a "scale anchoring" process is used to define what it means to score in each of these levels. NAEP's scale anchoring follows an empirical procedure whereby the scaled assessment results are analyzed to delineate sets of questions that discriminate between adjacent performance levels on the scales. For the science, mathematics, and reading trend scales, these levels are 150, 200, 250, 300, and 350. For these five levels,

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

[88] Stocking, M. L. & Lord, F. M., "Developing a Common Metric in Item Response Theory," *Applied Psychological Measurement*, Vol. 7, 201-10, 1983.

questions were identified that were likely to be answered correctly by students performing at a particular level on the scale and much less likely to be answered correctly by students performing at the next lower level.

The guidelines used to select such questions were as follows: students at a given level must have *at least* a 65 percent probability of success with the questions, while students at the next lower level have a much lower probability of success (that is, lower than 50 percent); and the difference in probabilities between adjacent levels must exceed 30 percent. For each of the three curriculum areas, subject-matter specialists examined these empirically selected question sets and used their professional judgment to characterize each level. The reading scale anchoring was conducted on the basis of the 1984 assessment, and the scale anchoring for mathematics and science trend reporting was based on the 1986 assessments. Scale anchoring was not used with the writing assessment.

For writing, an item mapping procedure was used to portray the writing skills of students at various scale levels. Writing trend assessment tasks were scored on a three- or four-point scale. As a result of the item mapping procedure, researchers identified, for each task, the points on the NAEP writing scale at which it is estimated that 65 percent of students would write a response scored one or better (unsatisfactory response or better), two or better (minimal response or better), three or better (adequate response or better), and four (elaborated response). For those tasks having only three score points, of course, the highest category is three (adequate response).

NAEP Reporting Groups

This report contains results for the nation and for groups of students within the nation defined by shared characteristics. The subgroups defined by race/ethnicity, parents' education level, gender, and region are defined below.

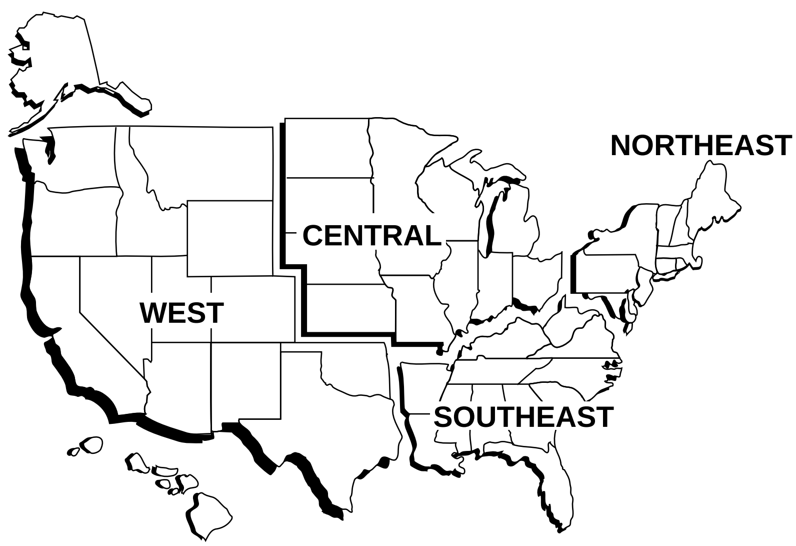
Race/Ethnicity. Results are presented for students in different racial/ethnic groups according to the following mutually exclusive categories: White, Black, Hispanic, Asian/Pacific Islander, and American Indian (including Alaskan Native). Some racial/ethnic results are not reported separately because there were too few students in the group. The data for all students, regardless of whether their racial/ethnic group was reported separately, were included in computing the overall national results.

Parents' Education Level. Students were asked to indicate the extent of schooling for each of their parents: did not finish high school, graduated from high school, had some education after high school, or graduated from college.

The response indicating the higher level of education for either parent was selected for reporting.

Gender. Results are reported separately for males and females. Gender was reported by the student.

Region. The United States was divided into four regions: Northeast, Southeast, Central, and West. States in each region are shown on the map below. Each state except Virginia is contained entirely in one region. The part of Virginia that is part of the Washington, D.C.-Maryland-Virginia metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.



Estimating Variability

The statistics presented in this report are estimates of group and subgroup performance based on samples of students, rather than the values that could be calculated if every student in the nation answered every assessment question. It is therefore important to have measures of the degree of uncertainty of the estimates. Accordingly, in addition to providing estimates of percentages of students and their average scale score, this report provides information about the uncertainty of each statistic.

Two components of uncertainty are accounted for in the variability of statistics based on scale scores: the uncertainty due to sampling only a relatively small number of students and the uncertainty due to sampling only a relatively small number of questions. The variability of estimates of percentages of students having certain background characteristics or answering a certain cognitive question correctly is accounted for by the first component alone. Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. For this reason, NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any information about students that can be observed without error, but each student typically responds to so few questions within any content area that the scale score for any single student would be imprecise. In this case, using plausible values technology makes it possible to describe the performance of groups and subgroups of students, but the underlying imprecision that makes this step necessary adds an additional component of variability to statistics based on NAEP scale scores.[89]

The reader is reminded that, like those from all surveys, NAEP results are also subject to other kinds of errors including the effects of necessarily imperfect adjustments for student and school nonresponse and other largely unknowable effects associated with the particular instrumentation and data collection methods used. Nonsampling errors can be attributed to a number of sources: inability to obtain complete information about all selected students in all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording,

---

[89] For further details, see Johnson, E. G., "Considerations and Techniques for the Analysis of NAEP Data," in *Journal of Educational Statistics* (December 1989).

coding, or scoring data; and other errors of collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate. By their nature, the impacts of such error cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

Drawing Inferences from the Results

The use of confidence intervals, based on the standard errors, provides a way to make inferences about the population averages and percentages in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample scale score average ± 2 standard errors represents about a 95 percent confidence interval for the corresponding population quantity. This means that with approximately 95 percent certainty, the average performance of the entire population of interest is within ± 2 standard errors of the sample average.

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

$$\text{Average} \pm 2 \text{ standard errors} = 256 \pm 2\,(1.2) = 256 \pm 2.4 =$$
$$256 - 2.4 \text{ and } 256 + 2.4 = 253.6,\ 258.4$$

Thus, one can conclude with close to 95 percent certainty that the average scale score for the entire population of students in that group is between 253.6 and 258.4.

Similar confidence intervals can be constructed for percentages, provided that the percentages are not extremely large (greater than 90) or extremely small (less than 10). For extreme percentages, confidence intervals constructed in the above manner may not be appropriate, and procedures for obtaining accurate confidence intervals are quite complicated.

To determine whether there is a real difference between the average scale score (or percentage 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 average scale scores 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 squaring each group's standard error, summing these squared standard errors, and then taking the square root of this sum. This procedure produces a conservative estimate of the standard error of the difference, since the estimates of the group averages or percentages will be positively correlated to an unknown extent due to the sampling plan. Direct estimation of the standard errors of all reported differences would involve a heavy computational burden.

Similar to the manner in which the standard error for an individual group average or percentage is used, the standard error of the difference can be used to help determine whether differences between assessment years are real. If one wants to hold the certainty level for a specific set of comparisons at a particular level (e.g. .95), adjustments (called multiple-comparisons procedures) need to be made. The set of comparisons is referred to as a "family," and the typical family involves all years in the assessment, even for tables in which only the first and last years' results are reported. One such procedure — the Bonferroni method — was used to form confidence intervals for the trend differences between 1994 and each previous assessment year, as well as between the first and each successive year.

Multiple-comparisons procedures are useful for controlling the overall Type I error rate for a defined set of hypothesis tests. However, especially when the number of potential comparisons which could be made is large, as in NAEP data, this protection comes at the substantial loss of power in detecting specific consistent patterns in the data. For example, more powerful and complex tests of significance designed to identify consistent patterns in the data might judge that two groups were significantly different when a Bonferroni multiple-comparisons procedure would not.

One such test of patterns in NAEP data is the test of linear and quadratic trends applied to the trend data for the nation and selected subpopulations. The linear and quadratic components of the trend in average scale scores for a given subject area and age group were estimated by applying two sets of contrasts to the set of average scale scores by year. The linear component of the trend was estimated by the sum $b_1 = \sum c_j x_j$, where the x_j are the average scale scores by year and the c_j are defined such that b_1 corresponds to the slope of an unweighted regression of the average scale scores on the assessment year. The quadratic component was estimated by the sum $b_2 = \sum d_j x_j$, in which the d_j are formally orthogonal to the c_j and are defined such that b_2 is the quadratic term in the unweighted regression of the average scale scores on the assessment year and the square of the assessment year. The statistical significance of b_1 and b_2 was evaluated by comparing each estimate to its estimated standard error. The standard error of b_1 was estimated as the square root of the sum $\sum c_j^2\ SE_j^2$, in which SE_j is the estimated standard error of x_j. The estimated standard error of the b_2 was analogously defined. The reader is cautioned that some averages and standard errors in this report may differ slightly from values reported in previous trend reports because of a slight modification of procedures. The method used to round off numbers to the number of reported decimal places was modified to conform to NCES standards.

ACKNOWLEDGMENTS

The work presented herein represents the efforts of the hundreds of individuals who are necessary to implement a project of this scope across several decades. From the considerable expertise, energy, and dedication required to develop and conduct the NAEP trend assessments to that necessary to analyze and report the data, many persons have made important and substantial contributions. Most importantly, NAEP is grateful to students and school staff who made the trend assessments possible.

The National Center for Education Statistics (NCES), in the Office of Educational Research and Improvement of the U.S. Department of Education funded the NAEP trend assessments. Pascal Forgione, Commissioner, and the NCES staff — particularly Gary Phillips, Peggy Carr, Arnold Goldstein, and Susan Ahmed — worked closely and collegially with ETS, Westat, and NCS staff and played a crucial role in all aspects of the program. The NAEP 1994 assessments and reports also benefited from the constant support and guidance of Emerson Elliot, past commissioner of NCES. The members of the National Assessment Governing Board (NAGB) and the NAGB staff provided invaluable advice throughout.

The NAEP project at ETS is directed by Paul Williams and resides in the Center for the Assessment of Educational Progress (CAEP), managed by Archie Lapointe and Paul Williams. Steve Lazer managed test development activities, and Jules Goodison managed operational aspects of the trend assessment. The sampling and data collection activities were carried out by Westat under the direction of Rene Slobasky, Nancy Caldwell, and Keith Rust. Printing, distribution, scoring, and processing activities were conducted by NCS, under the supervision of Judy Moyer, Brad Thayer, Mathilde Kennel, Linda Reynolds, and Barbara Price.

Statistical and psychometric activities were led by Hua Hua Chang, Jo-Lin Liang, Spence Swinton, and Eiji Muraki under the direction of Eugene Johnson, John Mazzeo, and Jim Carlson. Nancy Allen, John Donoghue, and Frank Jenkins provided insightful advice as the analyses progressed. Analysis activities were managed by John Barone, with the support of Alfred Rogers and Debbie Kline. David Freund, Steve Isham, Bruce Kaplan, and Edward Kulick performed the analyses, assisted by Lois Worthington, Norma Norris, Minhwei Wang, Steve Wang, Janet Chen, Hong Zhou, Phillip Leung, Mike Narcowich, and Craig Pizzuti. The authors of the report received considerable editorial and quality control help from Patricia Donahue, Lynn Jenkins, Karen Miller, Alexandra Beatty, Fiona Herr, and Christy Schwager. Carol Errickson and Sharon Davis-Johnson oversaw the production aspects of this report, with the ETS Publications Division and Regency InfoGraphics providing expert and efficient composition services.

Many thanks are due to the numerous reviewers, both internal and external to ETS and NCES. The comments and critical feedback provided by the following reviewers are reflected in the final version of this report: Susan Ahmed, David Baker, Peggy Carr, Mary Lyn Bourque, Larry Feinberg, Alan Ginsburg, Steven Gorman, Eugene Johnson, Andrew Kolstad, Steve Lazer, John Mazzeo, Laurence Peters, Deborah Spitz, Larry Suter, Jerry West, and Shi-Chang Wu.