DOCUMENT RESUME

ED 430 246 CS 216 717

AUTHOR Ballator, Nada; Farnum, Marisa; Kaplan, Bruce
TITLE NAEP 1996 Trends in Writing: Fluency and Writing Conventions. Holistic and Mechanics Scores in 1984 and 1996.
INSTITUTION Educational Testing Service, Princeton, NJ.; National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-1999-456
ISBN ISBN-0-16-050042-7
PUB DATE 1999-04-00
NOTE 69p.; "In collaboration with Claudia Gentile, Elissa Greenwald, Debra Kline, John Mazzeo, and Ingeborg Novatkoski."
AVAILABLE FROM U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328; ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398; Tel: 877-433-7827 (Toll Free); Web site: http://www.ed.gov/pubs/edpubs.html
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Comparative Analysis; Elementary Secondary Education; Grade 11; Grade 4; Grade 8; Grammar; *Holistic Evaluation; Punctuation; Scores; *Scoring; *Sentence Structure; Spelling; Tables (Data); Test Use; *Writing Achievement; *Writing Evaluation; Writing Research; Writing Skills
IDENTIFIERS National Assessment of Educational Progress

ABSTRACT

Supplementary to "NAEP 1996 Trends in Academic Progress," this report describes two aspects of writing for which change has been measured since 1984: writing fluency as determined by holistic scoring; and mastery of the conventions of written English as determined by mechanics scoring. The introduction discusses the layout and means of evaluation: measuring fluency of writing, mechanics of writing, and expressing the differences in performance. Chapter 1 compares holistic scores in 1984 and 1996 regarding student writing fluency in informative, persuasive, and narrative writing. Chapter 2 compares mechanics scores in 1984 and 1996 to look at overall characteristics of the papers, use of sentence types, control of sentence structure, and control of word-level and punctuation conventions. A summary includes discussion of fluency in writing, grammar, spelling, and punctuation, and observations made from the research. Appendixes present an overview of scoring procedures, scoring guides, and tables with standard errors. (SC)

CS

NAEP 1996 Trends in WRITING

Fluency and Writing Conventions

HOLISTIC AND MECHANICS SCORES IN 1984 AND 1996

U.S. Department of Education
Office of Educational Research and Improvement NCES 1999-456

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 1996 Trends in Writing:

Fluency and Writing Conventions

Holistic and Mechanics Scores in 1984 and 1996

Nada Ballator
Marisa Farnum
Bruce Kaplan

in collaboration with
Claudia Gentile, Elissa Greenwald, Debra Kline,
John Mazzeo, and Ingeborg Novatkoski

April 1999

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

NCES 1999-456

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

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

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

Assessment Division
Peggy G. Carr
Associate Commissioner

April 1999

SUGGESTED CITATION
U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. *NAEP 1996 Trends in Writing: Fluency and Writing Conventions,* NCES 1999–456, by N. Ballator, M. Farnum, and B. Kaplan. Washington, DC: 1999.

FOR MORE INFORMATION
Content contact:
Arnold Goldstein
202–219–1741

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

Education Publications Center (Ed Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794–1244

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

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

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

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-050042-7

Contents

Introduction

The NAEP Long-Term Trend Writing Assessment

The NAEP long-term trend writing assessment provides an important picture of students' progress over time because it compares performance on the same writing tasks, administered in identical fashion to comparable samples of students and yielding comparable scores. There have been six national assessments of writing conducted during the school years ending in 1984, 1988, 1990, 1992, 1994, and 1996. The 1996 assessment included the same set of 12 writing tasks that had been administered in the five previous assessments. Each of these trend assessments was administered to nationally representative samples of students in grades 4, 8, and 11.

Over the past three decades, many teacher educators and classroom teachers have been emphasizing the writing process. The writing process approach focuses on the iterative nature of writing, in which writers plan, write, and revise their ideas in several drafts before a final version is produced. It is during the revision or editing stages of this process that writers focus on correcting grammatical and mechanical errors. Grammatical and mechanical correctness is not viewed as an end in and of itself, but eliminating these errors is an important part of improving the final draft. This report focuses on what changes, if any, have occurred in student writing between 1984 and 1996, the period examined by the NAEP long-term trend writing assessment.

This Report

Results of the 1996 long-term trend writing assessment are reported in two publications. This report describes two aspects of writing for which change has been measured since 1984: writing fluency, as determined by **holistic scoring**; and mastery of the conventions of written English (spelling, punctuation, grammar) as determined by **mechanics scoring**. This report is supplementary to *NAEP 1996 Trends in Academic Progress,*[1] the main report for the NAEP long-term trend assessment. That document reports trends in writing scores since 1984 as determined by **primary trait scoring**. This report presents the results of the

[1] Campbell, J. R., Voelkl, K. E., & Donahue, P. L. (1997). *NAEP 1996 trends in academic progress: Achievement of U.S. students in science, 1969 to 1996; mathematics, 1973 to 1996; reading, 1971 to 1996; and writing, 1984 to 1996* (Publication No. NCES 97-985). Washington, DC: National Center for Education Statistics. This report is frequently referred to as *Trends* in this report. It is available on the Web at http://nces.ed.gov/naep/.

holistic scoring of a subgroup of four of the 12 writing tasks, and the mechanics scoring of two of these four tasks.

The report is organized as follows: Chapter 1 compares student performance on writing tasks in 1984 and 1996 as measured by holistic scoring. Chapter 2 compares students' mastery of the conventions of writing (grammar, punctuation, and spelling) in 1984 and 1996. The brief Summary offers conclusions, and is followed by three appendices. Appendix A contains information about sample sizes and scoring procedures, and Appendix B contains the guides for holistic and mechanics scoring. Appendix C provides the standard errors for the data in the tables contained in the body of the report.

The NAEP long-term trend writing assessments discussed here and in *Trends* should not be confused with the main NAEP writing assessments. The long-term trend writing assessment was begun in 1984, and has presented students with the same writing tasks in the five ensuing assessments. These writing tasks are completely different from the prompts in the main NAEP assessment.[2] The use of different writing prompts, as well as other procedural differences, precludes direct comparison of the results of the long-term trend assessment discussed here with those of the main assessments.

Multiple Tasks and Multiple Measures of Writing

In order to assess students' abilities to write in a variety of formats and genres, the NAEP long-term trend writing assessment asks them to respond to several different tasks in each of three types of writing:

- informative tasks ask students to write descriptions, reports, and analyses;
- persuasive tasks ask students to write convincing letters and arguments; and
- narrative tasks ask students to write stories.

The NAEP long-term trend instrument consists of 12 distinct writing tasks; however, each student who participated in the assessment responded to only a few (usually two) of the 12 tasks. These tasks are assessed using three types of measures:

- primary trait scoring, as described in *NAEP 1996 Trends in Academic Progress*, measures success in accomplishing the specific task, e.g., writing persuasively;
- holistic scoring, reported here, measures fluency in a subgroup of four of the 12 tasks; and

[2] The NAEP long-term trend assessments have been administered in mathematics, science, reading, and writing, to national samples of students. Eighth graders are assessed in the fall, fourth graders in the winter, and eleventh graders in the spring, and the test booklets remain the same over all assessments. In contrast, the main NAEP 1992 Writing Assessment was conducted in the first quarter of 1992 at grades 4, 8, and 12, and the main NAEP 1998 Writing Assessment (based on a new framework) was conducted at grades 4, 8, and 12 in the first quarter of 1998. The 1998 main writing assessment was also administered to students in participating states at grade 8.

- mechanics scoring, also reported here, measures conventions of written English using a subgroup of two of the four holistically scored tasks.

Primary trait scoring is based on established criteria that reflect the success of the student in accomplishing the specific writing task; for primary trait scoring, a unique scoring guide was used for each of the tasks. Student responses to all 12 writing tasks received primary trait scoring as reported in the principal 1996 long-term trend report, *NAEP 1996 Trends in Academic Progress*.

However, there are other aspects of writing that it is also important to assess. For instance, general writing quality or fluency — the student's capacity to organize and develop a written piece, to use correct syntax, and to observe the conventions of standard written English — is important. These aspects of written communication, taken together, are what holistic evaluation of writing addresses.[3]

The long-term trend writing assessment consisted of three distinct parts: primary trait, holistic, and mechanics scoring criteria. First, all 12 of the long-term trend writing tasks were scored using primary trait scoring criteria. The results of this are reported in *NAEP 1996 Trends in Academic Progress* in Chapters 7 and 8 (pages 151-197).[4]

Next, a subgroup of four of these tasks was scored holistically — two tasks at each grade level. Two of the writing tasks were administered at grade 4 only, while the two other tasks were both administered at grades 8 and 11. One of the four is an informative task, one is a narrative task, and two are persuasive tasks. A brief description of each writing task and the grades at which the task was administered are provided in Figure I.1 on the next page. Holistic scoring of these tasks yielded information about students' level of writing fluency, as seen in Tables 1.1 - 1.3. Different scoring guides were used for holistic scoring of narrative, informative, and persuasive tasks, as described in Appendix B.

Lastly, to gain information about students' mastery of the conventions of written English, a subgroup of two of the holistic tasks was scored for mechanics — one at each grade level (see the table on the next page and Tables 2.1 - 2.3). The mechanics scoring involved assessing students' use of standard English sentence structure, rules of agreement, word choice, spelling, and punctuation. It also captured information about the overall length of the students' responses and the number and complexity of the sentences that they used. For mechanics scoring, the same criteria were used to evaluate all tasks. See Appendix B of this report for the mechanics and holistic scoring guides.

---

[3] It should be noted that holistic evaluation depends in part on aspects of writing measured in mechanics scores; Table A.3 and associated text discuss this relationship.

[4] Previous years of the *Trends* report also contain results from holistic and mechanics scoring of the tasks presented here. The 1994 *Trends* is also on the Web, as is the 1996 edition.

Figure I.1 Task by type of writing and summary of wrting tasks scored for fluency (H[a]) and for mechanics (M[b])

Tasks by type of writing	Summary of writing tasks scored for fluency (H[a]) and for mechanics (M[b])	Administered at Grade 4	8	11
INFORMATIVE	**Food on the Frontier** required students to read a social studies passage about frontier life and then to explain why modern-day food differs from frontier food		H	H
PERSUASIVE	**Spaceship** required students 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	H M		
PERSUASIVE	**Recreation Opportunities** required students to take a stand on whether their own town should purchase an abandoned railroad track or a warehouse as a recreation center, to defend their choice, and to refute the alternative choice		H M	H M
NARRATIVE	**Flashlight** required students to write a story about their imagined adventures with a flashlight that has special powers	H		

[a] Holistic scoring (fluency) is described in Chapter 1; the scoring guides are in Appendix B.

[b] Mechanics scoring measures the writer's control of the conventions of written English. Mechanics scoring is described in Chapter 2; the scoring guides are in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Measuring the Fluency of Writing

Holistic scoring is the most commonly used method for evaluating students' writing performance in the United States today. Holistic scoring for NAEP focuses on the writer's fluency in responding to a task relative to the performance of other students at that grade level.[5] Fluency reflects a writer's facility with language both in terms of the development

[5] Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper & L. Odell (Eds.), *Evaluation writing: Describing, measuring, judging*. Urbana, IL: National Council of Teachers of English.

and organization of ideas and in the use of syntax, diction, and grammar. Holistic scoring methods were specifically designed to assess writing fluency. The underlying assumption of holistic scoring is that the whole piece of writing is greater than the sum of its parts. In holistic scoring, readers do not make separate judgments about specific aspects of a written response, but rather consider the overall effect, rating each paper on the basis of its general fluency.

In the NAEP long-term trend assessment, responses to four tasks are scored holistically, two tasks at each of the three grades (the same two tasks are administered at both eighth and eleventh grades). The characteristics of general fluency are assessed on a six-point scale, and described in the holistic scoring guides for narrative, informative, and persuasive writing tasks in Appendix B. In order to make comparisons of students' writing fluency across all six years of the assessment, all papers from the previous years were scored holistically, along with all of the 1996 papers. For each year, approximately 1200 papers[6] from each grade are scored.

As is typical with all holistic scorings, raters are trained on a particular task immediately before scoring the papers written in response to that task (as described in Appendix A). For each task, the papers from all years are randomly mixed and then assigned one of six scores. To detect changes in fluency from one assessment to another, the percentages of papers from each year within a given score category are compared. The comparisons reported here are for the first or base year and the current year, as in previous reports.[7]

Thus, while primary trait scoring is based on specific constant criteria and so permits year-to-year and grade-to-grade comparisons, holistic scoring allows within grade comparisons of relative fluency over all years according to contemporaneous criteria.

Measuring the Mechanics of Writing

Another set of analyses, applied to papers written for two of the tasks (see Figure I.1), focused on the mechanics of students' writing. While error counts do not fully reflect a writer's fluency and competency, many educators, policy makers, and parents are interested in the kinds of surface errors students make as they write.[8] Students' mastery of the sentence-level and word-level conventions of English, as well as their use of correct spelling and punctuation, were examined. (See Appendix A for procedures used in scoring,

---

[6] For the first or base year of the assessment (1984), the number of papers was about half the quantity of that in ensuing years.

[7] For instance, see Campbell, J., Reese, C., O'Sullivan, C., & Dossey, J. (1996). *NAEP 1994 trends in academic progress: Achievement of U.S. students in science, 1969 to 1994; mathematics, 1973 to 1994; reading, 1971 to 1994; and writing, 1984 to 1994* (Publication No. NCES 97-095). Washington, DC: National Center for Education Statistics.

[8] Shaughnessy, M. P. (1977). *Errors and expectations: A guide for the teacher of basic writing*. New York, NY: Oxford University Press.

and Appendix B for the mechanics scoring guide.) In order to examine changes in students' success in using 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.

Expressing the Differences in Performance

Because the analysis is conducted using papers written by students who are part of a sample (rather than from the entire population of fourth, eighth, or eleventh graders in the nation) the numbers reported are necessarily *estimates*. As such, they are subject to a measure of uncertainty. This measure of uncertainty is reflected in the *standard error* of the estimate, which can be seen in Appendix C, in tables paralleling those in the main body of the report. In comparing student performances on a particular characteristic by either number or percentage, it is essential to take into account the standard error, rather than to rely solely on observed similarities or differences. The comparisons discussed in this report and marked with asterisks in the tables are based on statistical tests that consider both the magnitude of the difference between the averages and the standard errors of those statistics.

The statistical tests determine whether the evidence — based on data from the two years — is strong enough to conclude that there is an actual difference. If the evidence is strong (i.e., the difference is statistically significant), statements comparing 1996 with 1984 use terms such as higher, lower, increased, or decreased. The reader is cautioned to rely on the results of the statistical tests, as expressed in the text or as indicated in the tables, rather than on the apparent magnitude of the differences.[9]

The statistical tests employed here used Bonferroni procedures to form confidence intervals for the differences for sets of comparisons. Bonferroni procedures are appropriate for sets or "families" of comparisons, allowing adjustments according to family size to keep the certainty or significance level as specified (that is, a 95 percent certainty or 5 percent significance level). For comparisons in this report, several family sizes were used. Consider, for example, Table 2.1, which presents overall averages in 1984 papers compared with those in 1996 papers. For these across-year comparisons, the family size is 1, and consequently no adjustment is needed. Table 2.1 also presents across-year comparisons for papers in the lower and upper halves of the holistic scale; in this case, each half is a family of 1, so a Bonferroni adjustment is made for a family size of 2. Further information on statistical tests and adjustment procedures are in the *NAEP 1996 Technical Report.*

[9] Standard errors measure the uncertainty that another sample drawn from the same population could have yielded somewhat different results.

Chapter 1

Student Writing Fluency: Holistic Scores in 1984 and 1996

This chapter explores changes in students' writing fluency as measured by their responses to selected informative, persuasive, and narrative writing tasks in the NAEP long-term trend assessment. Using a holistic scoring method, readers made a judgment about the overall quality of a piece of writing rather than about separate aspects of the response, such as its grammatical correctness or pattern of development. The aspects of fluency that readers looked for in judging writing achievement holistically were delineated in a scoring guide that defined the scoring criteria for the assessment. These scoring criteria — organization, coherence, elaboration of ideas, facility with language, and mechanics — are organized on a six-point scale, each describing a different level of competence. The complete holistic scoring guides for the informative, persuasive, and narrative tasks are presented in Appendix B.

In the following tables, responses that received scores in the upper half of the scale (scores of 4, 5, or 6) are examples of better writing fluency. Responses with scores in the lower half of the scale (scores of 3, 2, or 1) demonstrate problems with the aspects of fluency described in the scoring guides in Appendix B. Papers receiving scores of 0 were essentially blank and could not be rated.

Fluency in Informative Writing

"Food on the Frontier" is a task designed to elicit a sample of informative writing from students. Some informative writing requires reporting from personal experience or from given information, and involves descriptions of what happened or what exists. Analytic writing, such as that called for in "Food on the Frontier," asks for an explanation of why something happened as it did or how parts fit together.

The "Food on the Frontier" task was administered to samples of eighth and eleventh graders. Their responses were scored holistically, as a measure of writing fluency. As Table 1.1 shows, there was an improvement in 1996 at grade 8, with an average rating of 3.06 compared to 2.77 in 1984. However, at grade 11, 1984 and 1996 holistic ratings were essentially the same.[10]

[10] To determine whether 1984 and 1996 values were different, standard errors were taken into account (see Table C1.1).

Table 1.1 **Fluency in an informative task, "Food on the Frontier," comparing base year to current year**

THE NATION'S REPORT CARD

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers			
	Grade 8		Grade 11	
	1984	1996	1984	1996
0[a]	5.5	5.0	5.4	5.3
1	11.3	7.3	4.6	4.5
2	22.6	18.0	13.3	9.3
3	31.5	32.3	21.5	25.6
4	21.1	24.1	32.1	30.9
5	6.0	10.3	16.1	18.0
6	1.9	3.0	7.0	6.3
4, 5, or 6	29.0	37.4	55.1	55.2
Average rating	2.77	3.06*	3.46	3.51

[a]Papers receiving scores of 0 were blank or otherwise could not be rated.

* Statistically significant difference from 1984, at the 95 percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Fluency in Persuasive Writing

The primary purpose of persuasive writing is to influence — to change ideas or actions. It is used, for example, to convince others to adopt a particular point of view or a course of action, to refute arguments, or to defend certain positions or behaviors (details about the specific tasks can be found in Figure I.1 on page 4). Two persuasive tasks administered in the long-term trend writing assessment were analyzed holistically: "Spaceship" involved writing to convince others to adopt a particular point of view, and "Recreation Opportunities" involved writing to refute a position. It should be noted that students' responses to persuasive tasks are evaluated on their effectiveness in communicating and supporting an opinion, but not on the merit of the opinion itself. Fourth graders responded to the "Spaceship" task. The "Recreation Opportunities" task was administered to eighth- and eleventh-grade students. The results of these holistic analyses are presented in Table 1.2.

For the "Spaceship" task (fourth grade), the percentage of papers receiving the lowest score, a fluency rating of 1, was lower in 1996 than the percentage of papers receiving that score in 1984 (7.6 percent compared to 12.7 percent). Along with this decrease in the percentage of papers earning a rating of 1, the percentages of papers receiving the higher ratings of 4, 5, or 6 increased in 1996 compared to 1984 (25.3 percent and 18.3 percent, respectively). However, this did not result in a significant change in the average holistic scores in 1984 and 1996 (2.61 and 2.78 points respectively.)

The "Recreation Opportunities" task was administered at grades 8 and 11. The average fluency rating of eighth graders' responses on this task increased from 2.98 in 1984 to 3.21 in 1996. The increases in the percentage of responses receiving a rating of 4 or 5, although neither was significant, contributed to this improvement in fluency since 1984. Among eleventh graders, the relative fluency of written responses on this task remained essentially the same in 1996 as in 1984.[11]

[11] To determine whether 1984 and 1996 values were different, standard errors were taken into account (see Table C1.2).

Table 1.2 Fluency in persuasive tasks, "Spaceship" and "Recreation Opportunities," comparing base year to current year

THE NATION'S REPORT CARD

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers					
	Spaceship		Recreation Opportunities			
	Grade 4		Grade 8		Grade 11	
	1984	1996	1984	1996	1984	1996
0[a]	4.4	6.4	5.8	3.8	2.4	4.0
1	12.7	7.6*	8.8	7.0	2.0	2.5
2	22.4	18.9	12.7	14.5	11.2	9.1
3	42.3	41.8	36.0	29.7	27.5	27.0
4	15.2	20.4	28.8	32.3	37.6	36.5
5	2.6	4.3	6.3	10.3	12.8	15.6
6	0.5	0.6	1.5	2.4	6.6	5.4
4, 5, or 6	18.3	25.3*	36.7	45.1	57.0	57.5
Average rating	2.61	2.78	2.98	3.21*	3.61	3.58

[a] Papers receiving scores of 0 were blank or otherwise could not be rated.

* Statistically significant difference from 1984, at the 95 percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Fluency in Narrative Writing

Narrative writing involves the production of stories or personal essays, and requires observation of people, objects, and/or places. At its best, narrative writing fosters creativity, imagination, and speculation by allowing writers to express their thoughts and emotions, and offers an opportunity for writers to analyze and understand their actions and those of others.

"Flashlight" was a narrative task presented at grade 4. As shown in Table 1.3, the average fluency of fourth graders' written responses to the "Flashlight" task remained essentially the same in 1996 (3.02) as in 1984 (2.82). However, there was an increase in the percentage of papers receiving a holistic rating of 4 (from 15 percent in 1984 to 23 percent in 1996).

Table 1.3 Fluency in a narrative task, "Flashlight," comparing base year to current year

THE NATION'S REPORT CARD

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers	
	Grade 4	
	1984	1996
0[a]	6.2	5.4
1	10.4	6.6
2	24.6	24.2
3	30.1	26.6
4	14.8	22.6*
5	11.0	10.7
6	3.0	3.8
4, 5, or 6	28.8	37.2
Average rating	2.82	3.02

[a] Papers receiving scores of 0 were blank or otherwise could not be rated.

* Statistically significant difference from 1984, at the 95 percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Overall Changes in Fluency from 1984 to 1996

Across the four tasks assessed at 3 grades, the changes in writing fluency since 1984 have been few but positive. For grade 8, there were improvements for both tasks. For the tasks at grades 4 and 11, the average ratings either remained unchanged, or appeared to increase slightly, although this increase was not by a statistically significant amount.

Chapter 2

Grammar, Spelling, and Punctuation: Mechanics Scores in 1984 and 1996

Has student writing improved in terms of writing conventions? Are students consistently creating longer essays with ease and doing so with fewer errors? In an attempt to answer this question, student responses to one task at each grade were selected for further analysis.

The tasks chosen were "Spaceship" at grade 4 and "Recreation Opportunities" at grades 8 and 11. For the detailed analysis of writing mechanics, approximately 500 essays were drawn from the total national sample in each of the long-term trend assessment years. In addition to receiving measures indicative of overall quality as represented by primary trait and holistic scores, each of these papers was analyzed for various aspects of spelling, word choice, punctuation, and syntax.

Because these two writing tasks were not identical across all three grades, improvement in use of writing conventions cannot be measured from one grade to another — nor is growth by grade level the primary issue addressed by the NAEP long-term trend assessment. Since *NAEP 1996 Trends in Academic Progress* reports primary trait scoring for all 12 writing tasks, it can serve as a rough gauge of growth from grade 4 to grade 8 and from grade 8 to grade 11. However, the more important question, for that report as well as this, is one of trend: what changes have occurred at a particular grade level since 1984? Improvement in some measures and decline in others provide a complex answer to the question of how students' writing achievement has changed between 1984 and 1996.

In this Chapter, the evidence of change in students' use of the conventions of written English is examined through the analysis of responses to the "Spaceship" task at grade 4 and the "Recreation Opportunities" task at grades 8 and 11. Students' performance in the base year, 1984, is compared with their performance in the current year, 1996. The emphasis of this chapter is on *average* performance according to mechanics scoring of the students' papers. Included in the tables and in the discussion are mechanics measures for papers judged to be in the lower half of the holistic scale for fluency (with holistic scores of 1, 2, or 3) and in the upper half of the holistic scale for fluency (with holistic scores of 4, 5, or 6), to permit comparisons between lower half or upper half papers in 1984 and 1996.[12] For instance, it might be anticipated that the papers in the upper half of the holistic scale might show an increase in number of sentences and punctuation marks used, as a product of increased emphasis in writing. For the same reason, a differential decrease in

[12] Note that because the mechanics scores are determined only once, in the year of the assessment, the lower- or upper-half assignment of the papers is determined by the papers' holistic scores in the year of the assessment.

the percentage of sentence errors might be expected in the better papers. Comparing papers at the two different levels may provide additional information about whether changes occurred at only one level or at both halves of the holistic scale.

In the next section, the following elements of student papers are examined: length of the papers (number of words and number of sentences) and all types of errors (the total of errors in sentence structure, word choice errors, misspellings, and wrong or missing punctuation). Table 2.1 summarizes the comparisons between student performance in 1984 and in 1996.

Overall Characteristics of Papers

After expert scorers coded students' responses to the "Spaceship" and "Recreation Opportunities" tasks for various aspects of grammar, spelling, and punctuation, their codings were entered into a computer-readable database. Measures including numbers of sentences per essay, words per essay, and words per sentence appear in Table 2.1.

Errors also were coded by the expert scorers, and tabulated by computer. This included all errors in sentence structure, such as agreement errors, awkward sentences, run-on sentences, and sentence fragments; word use errors; punctuation and spelling errors; and incorrect word choices.

Comparing Overall Characteristics (Table 2.1)

- → On average in 1996 papers, two overall measures usually associated with improvement in writing skills (**average number of full words per paper** and **average number of sentences per paper**) increased when compared to average performance in 1984. Both of the increases were at grades 8 and 11 only.
- → **Average number of words per sentence** showed no change between 1984 and 1996, with one exception: for fourth graders' papers at the lower half of the holistic scale, the sentence length increased in 1996.
- → Even though eighth- and eleventh-graders' papers were longer in 1996, their error rates (**average number of all errors per 100 words**) did not change from 1984 to 1996.

Table 2.1 Overall characteristics of papers: 1984 and 1996

THE NATION'S REPORT CARD naep

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of full words per paper	4	1996	35.4	28.4	53.9
		1984	33.8	27.9	47.7
	8	1996	79.4*	59.2*	104.2*
		1984	67.5	51.1	89.5
	11	1996	104.4*	68.6	124.4
		1984	93.3	62.0	115.0
Average number of sentences per paper	4	1996	2.6	2.1	3.9
		1984	2.6	2.2	3.6
	8	1996	5.2*	3.8	6.8*
		1984	4.4	3.4	5.8
	11	1996	6.5*	4.1	7.8*
		1984	5.6	3.7	6.9
Average number of words per sentence	4	1996	16.1	16.1*	16.0
		1984	15.1	14.3	16.8
	8	1996	17.7	18.3	16.9
		1984	17.3	17.5	17.0
	11	1996	17.7	18.8	17.0
		1984	18.2	18.8	17.7
Average number of all errors per 100 words	4	1996	17.2	19.2	11.8
		1984	15.5	17.1	11.6
	8	1996	10.2	12.1	7.9
		1984	9.2	10.3	7.7
	11	1996	7.4	9.2	6.3
		1984	7.0	8.4	6.0

*Statistically significant difference from 1984, at the 95-percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

Definitions of these measures appear in Appendix B.

Note: For the first three characteristics, average number of full words divided by average number of sentences does not necessarily produce average number of words per sentence. Consider a simplified example using four papers, with two papers each containing one sentence with 20 words and two papers each containing 10 sentences with 100 words. In this example, the average number of words would be 60, the average number of sentences would be 5.5, and the average number of words per sentence would be 15. Consequently, in this example, words per paper divided by average number of sentences per paper yields 10.5 (instead of 15) as the average number of words per sentence.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

A Profile by Grade (Table 2.1)

Grade 4

In 1984 and 1996, fourth-grade students on average wrote responses of comparable length in terms of number of words per paper, number of sentences per paper, and number of words per sentence. In terms of all errors per 100 words in student papers, there was no change between 1984 and 1996. The only change noted in this table is an increase in fourth graders' lower-level papers, with more words per sentence in 1996 (16.1) than in 1984 (14.3).

Grade 8

Overall, papers written in 1996 were longer than those written in 1984 in terms of number of full words per response (79.4 words compared to 67.5 in 1984). This increase in number of words can be seen in the papers in the lower half of the holistic scale as well as those in the upper half.

Overall, the papers also showed an increase in the number of sentences (5.2 in 1996 compared to 4.4 in 1984). This increase was also seen in the upper-level papers. Even though the length of the papers increased, the number of all errors per 100 words remained unchanged.

Grade 11

In a comparison of papers written in 1996 to those written in 1984, eleventh-grade students on average wrote longer responses in terms of number of full words per response (104.4 words in 1996 compared to 93.3 in 1984).

Overall, eleventh graders' papers also showed an increase in the number of sentences (6.5 in 1996 compared to 5.6 in 1984). This increase was seen in the upper-level papers but not in the lower-level papers, similar to the pattern seen at grade 8. Just as for grade 8, the number of all errors per 100 words remained unchanged as the number of words increased.

Use of Sentence Types

Students' control of syntax is reflected in the types of sentences they create and the number and types of faulty constructions found in these sentences. To examine changes across time in students' command of sentence structure, simple sentences and complex or compound sentences were identified by scorers, coded, and counted by computer.

Types of sentences

- **Simple** — A sentence that contains a subject and a verb. It may also contain an object, subject complement, phrase, appositive, nominative absolute, or verbal. (See Appendix B for descriptions.) Also, it may have a word group used in dialogue, for emphasis, or as an exclamation that is not an independent clause.
- **Compound** — A sentence containing two or more simple sentences joined by something other than a comma.
- **Complex** (and compound-complex) — A sentence that contains at least one independent clause and one dependent clause.

Table 2.2a first repeats (from Table 2.1) the average number of sentences per paper. These data are also repeated in Table 2.2b to serve as a reminder that when we say, for example, around 50 percent of the sentences were complex or compound, we are talking about only a few sentences — from just over one sentence at grade 4 to around 3 sentences at grade 11. The remainder of Table 2.2a contains the average percentages of sentence types in the student papers.

Comparing Sentence Types (Table 2.2a)

→ As mentioned in the discussion of Table 2.1, the **average number of sentences per paper** was higher in 1996 than in 1984 at grades 8 and 11; this increase is reflected in papers rated in the upper half of the holistic scale. This measure is repeated here to provide a gauge for the numbers of simple and complex or compound sentences per paper.

→ At grade 11 only, the **percentage of simple sentences** increased since 1984; this increase also could be seen in the upper-level papers at grade 11.

→ The **percentage of complex or compound sentences** decreased at grades 8 and 11. The decrease at grade 11 was also seen in upper-level papers.

Table 2.2a **Number and types of sentences: 1984 and 1996**

THE NATION'S REPORT CARD NAEP

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of sentences per paper[1]	4	1996	2.6	2.1	3.9
		1984	2.6	2.2	3.6
	8	1996	5.2*	3.8	6.8*
		1984	4.4	3.4	5.8
	11	1996	6.5*	4.1	7.8*
		1984	5.6	3.7	6.9
Percentage of simple sentences	4	1996	21.9	20.5	25.6
		1984	23.1	23.3	22.2
	8	1996	34.9	30.0	40.9
		1984	33.7	32.3	35.6
	11	1996	40.9*	31.8	46.0*
		1984	35.7	32.7	37.8
Percentage of complex or compound sentences	4	1996	52.0	52.6	50.3
		1984	54.8	56.4	51.2
	8	1996	44.8*	45.7	43.8
		1984	49.8	50.6	48.6
	11	1996	44.0*	50.5	40.4*
		1984	52.4	54.3	51.1

* Statistically significant difference from 1984, at the 95-percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

[1] Average number of sentences per paper is repeated in this table because it can be used to estimate the numbers implied by the percentages of the different sentence types that follow.
Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Control of Sentence Structure

Sentence construction errors were also identified, coded, and tallied; four types of construction errors in the 1984 and 1996 student essays are reported here — run-on sentences, sentence fragments, agreement errors, and awkward sentences. It is important to remember that students' responses to the assessment tasks are examples of first draft writing. There is little time within the assessment period for students to proofread and correct their work. Therefore, a certain number of sentence-level errors, as well as other types of errors, can be expected.

Sentence structure errors

- **Run-On Sentences:**
  Fused — A sentence containing two or more independent clauses with no punctuation or conjunction separating them.
  On and on — A sentence consisting of four or more independent clauses strung together with conjunctions.
  Comma splice — A sentence containing two or more independent clauses separated by a comma instead of a semicolon or a coordinating conjunction.
- **Sentence Fragment** — A word group, other than an independent clause, written and punctuated as a sentence.
- **Awkward Sentence** (The awkward categories are listed in order of category precedence, since only one score was given to a sentence.)
  Faulty parallelism — A parallel construction that is semantically or structurally dysfunctional.
  Unclear pronoun reference — A pronoun's antecedent is unclear.
  Illogical construction — A sentence containing a faulty modification or a dangling modifier or a functionally misarranged or misproportioned sentence.
  Other dysfunctions — A sentence omitting a word, containing an extra word, or using a split construction that definitely detracts from readability.

Comparing Control of Sentence Structure (Table 2.2b)

- → For all three grades, the **percentage of run-on sentences** was equivalent in students' papers in 1984 and in 1996.
- → At the fourth and eighth grades, the average **percentage of sentence fragments** in students' papers was higher in 1996 than in 1984 papers; papers with scores in the lower half of the holistic scale also follow this pattern. In eleventh graders' papers, the percentage of sentence fragments was essentially the same in 1984 and 1996.
- → At the eighth and eleventh grades, the **percentage of awkward sentences** decreased. This is seen in the upper-level papers, although at grade eight, even the lower-level papers showed a decrease in the percentage of awkward sentences. At grade four, this measure was comparable in 1984 and 1996.

A Profile by Grade (Tables 2.2a and 2.2b)

Grade 4

The fourth graders' papers showed no increase in the number of sentences in their essays in 1996 compared to 1984. However, these papers averaged a higher percentage of sentence fragments in 1996, as did the papers at the lower half of the holistic scale.

Grade 8

Eighth graders' papers in 1996 contained more sentences on average, a pattern seen in papers in the upper half of the holistic scale. On average, the percentage of complex or compound sentences decreased in 1996 compared to 1984 papers. The percentage of sentence fragments increased in 1996 on average, as in papers in the lower half of the holistic scale. The percentage of awkward sentences decreased overall, and at both halves of the holistic scale.

Grade 11

A comparison of papers written in 1984 and 1996 at grade 11 reveals that the number of sentences per paper increased overall, as well as for papers in the upper half of the holistic scale. The percentage of complex or compound sentences decreased on average and for papers in the upper half of the holistic scale. Encouragingly, the percentage of awkward sentences decreased in 1996 on average and also in the upper-level papers.

Grade 11 performance differed from grades 4 and 8 in that the percentage of simple sentences increased at grade 11 but not at the other grades, and the percentage of sentence fragments was unchanged at grade 11, while it increased for the other two grades.

Table 2.2b Control of sentence structure: 1984 and 1996

THE NATION'S REPORT CARD

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of sentences per paper[1]	4	1996	2.6	2.1	3.9
		1984	2.6	2.2	3.6
	8	1996	5.2*	3.8	6.8*
		1984	4.4	3.4	5.8
	11	1996	6.5*	4.1	7.8*
		1984	5.6	3.7	6.9
Percentage of run-on sentences	4	1996	14.9	15.8	12.5
		1984	15.3	14.5	17.3
	8	1996	9.2	12.0	5.7
		1984	7.2	8.6	5.4
	11	1996	6.2	8.7	4.9
		1984	4.6	6.4	3.4
Percentage of sentence fragments	4	1996	5.8*	6.1*	5.0
		1984	3.2	3.2	3.1
	8	1996	5.2*	6.4*	3.6
		1984	3.3	2.7	4.0
	11	1996	3.8	5.7	2.8
		1984	3.0	3.0	3.0
Percentage of awkward sentences	4	1996	26.1	28.7	19.2
		1984	25.4	28.2	18.6
	8	1996	22.5*	26.0*	18.1*
		1984	32.3	36.9	26.1
	11	1996	19.0*	26.2	15.0*
		1984	31.2	36.0	27.7

*Statistically significant difference from 1984, at the 95-percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

[1] Average number of sentences per paper is repeated in this table because it can be used to estimate the numbers implied by the percentages for the different sentence types that follow.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Control of Word-Level and Punctuation Conventions

Students' control of the word-level conventions of written English may be seen in the data for word choice and spelling errors that are presented in Table 2.3a. Punctuation was also 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. Word-level measures shown in Table 2.3a include:

- **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.
- **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.

The measures shown in Table 2.3b include:

- **Punctuation** — Every error of commission and error of omission was coded for commas, dashes, quotation marks, semicolons, apostrophes, and end marks (periods, exclamation points, and question marks). Because standard punctuation rules may vary over time, informal rules of usage were applied, especially when coding the errors in comma usage, with the writer receiving the benefit of any doubt.

Comparing Word-Level Conventions (Table 2.3a)

→ Only at grade 4 did the **percentage of incorrect word choices** change from 1984; it increased in both upper- and lower-level papers as well as for the overall group.

→ The **percentage of spelling errors** remained the same as in 1984, at all three grades.

In Table 2.3a, the average number of full words per paper is repeated from Table 2.1 to provide a gauge of the actual numbers of incorrect words or spelling errors in students' papers.

Table 2.3a **Control of word-level conventions: 1984 and 1996**

THE NATION'S REPORT CARD NAEP

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of full words per paper[1]	4	1996	35.4	28.4	53.9
		1984	33.8	27.9	47.7
	8	1996	79.4*	59.2*	104.2*
		1984	67.5	51.1	89.5
	11	1996	104.4*	68.6	124.4
		1984	93.3	62.0	115.0
Percentage of incorrect word choices	4	1996	1.8*	2.0*	1.2*
		1984	0.8	0.9	0.4
	8	1996	0.8	1.1	0.5
		1984	0.7	0.7	0.6
	11	1996	0.5	0.6	0.5
		1984	0.6	0.8	0.5
Percentage of spelling errors	4	1996	7.8	8.7	5.4
		1984	8.4	9.0	6.8
	8	1996	3.9	4.7	2.9
		1984	3.7	4.2	3.1
	11	1996	2.7	3.2	2.4
		1984	2.3	2.7	2.0

* Statistically significant difference from 1984, at the 95-percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

[1] Average number of full words per paper is repeated in this table because it can be used to estimate the numbers implied by the percentages of the different sentence types that follow.
Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Comparing Punctuation Conventions (Table 2.3b)

- → The **average number of punctuation marks** per paper serves as a way to estimate the punctuation use and error rates that follow it. At grades 8 and 11 only, the average number of punctuation marks was higher in 1996 than in 1984. This coincides with the increase in number of sentences and number of words at grades 8 and 11. There was no change at grade 4 in average punctuation marks or paper length (number of sentences and number of words).
- → Although the longer essays led to more punctuation, the **punctuation error rate (not including omissions)** declined for grades 8 and 11, compared to 1984. This decline of punctuation errors was reflected in the lower-level papers at grades 8 and 11 as well as in the upper-level papers at grade 11. There was no change at grade 4.
- → The **punctuation omission rate** has not changed at any of the three grades.
- → The **comma and dash use rate** has stayed the same since 1984, and the **comma and dash error rate** has changed only at grade 4, where it has increased since 1984.
- → The **end mark use rate** has increased only at grade 11, and the **end mark error rate** decreased only at grade 8. However, the **percentage of sentences with end mark errors** decreased at grades 4 and 11, but not at grade 8.
- → Although the **other punctuation use rate** increased at both grade 4 and grade 11, the **other punctuation error rate** was unchanged at any of the three grades.

Table 2.3b Control of punctuation conventions: 1984 and 1996

THE NATION'S REPORT CARD

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of punctuation marks used[1]	4	1996	3.3	2.6	5.3
		1984	3.0	2.4	4.6
	8	1996	7.8*	5.1	11.1*
		1984	6.3	4.5	8.8
	11	1996	10.2*	6.0	12.6*
		1984	8.2	5.0	10.4
Punctuation error rate (not including omissions)	4	1996	0.2	0.3	0.2
		1984	0.3	0.3	0.3
	8	1996	0.2*	0.2*	0.2
		1984	0.5	0.5	0.4
	11	1996	0.2*	0.1*	0.2*
		1984	0.4	0.4	0.3
Punctuation omission rate	4	1996	2.3	2.5	1.7
		1984	2.3	2.8	1.4
	8	1996	1.4	1.7	1.2
		1984	1.3	1.5	1.1
	11	1996	1.3	1.7	1.1
		1984	1.3	1.8	0.9
Comma and dash use rate	4	1996	0.8	0.8	0.9
		1984	0.6	0.5	1.0
	8	1996	2.4	1.8	3.1
		1984	2.3	2.1	2.6
	11	1996	2.6	1.6	3.1
		1984	2.5	1.9	2.8
Comma and dash error rate	4	1996	1.2*	1.1*	1.3
		1984	0.8	0.7	0.9
	8	1996	1.0	1.0	0.9
		1984	1.0	1.0	0.9
	11	1996	0.8	0.8	0.9
		1984	0.9	1.1	0.8

(continued)

Table 2.3b Control of punctuation conventions: 1984 and 1996 (continued)

THE NATION'S REPORT CARD

Mechanics measure	Grade	Year	Holistic scores: Overall Average	Holistic scores: 1,2,3 Lower half of scale	Holistic scores: 4,5,6 Upper half of scale
End mark use rate	4	1996	7.0	6.8	7.5
		1984	6.9	6.9	7.0
	8	1996	6.3	6.2	6.4
		1984	6.1	5.9	6.4
	11	1996	6.0*	5.6	6.2*
		1984	5.6	5.5	5.6
End mark error rate	4	1996	1.1	1.4	0.4
		1984	1.6	2.1	0.5
	8	1996	0.5*	0.6	0.3
		1984	0.7	0.9	0.4
	11	1996	0.5	1.0	0.3
		1984	0.7	1.1	0.4
Percentage of sentences with end mark errors	4	1996	13.0*	15.3*	6.8
		1984	19.0	23.5	8.8
	8	1996	8.8	11.3	5.7
		1984	10.7	13.4	7.2
	11	1996	6.3*	8.6	5.0
		1984	10.3	14.4	7.4
Other punctuation use rate	4	1996	1.6*	1.4	2.1
		1984	1.2	1.0	1.5
	8	1996	0.9	0.8	0.9
		1984	0.8	0.7	1.0
	11	1996	1.0*	1.5	0.8
		1984	0.6	0.6	0.7
Other punctuation error rate	4	1996	0.3	0.3	0.2
		1984	0.3	0.3	0.2
	8	1996	0.2	0.3	0.2
		1984	0.2	0.1	0.2
	11	1996	0.1	0.1	0.1
		1984	0.2	0.1	0.2

* Statistically significant difference from 1984, at the 95-percent certainty level. In comparing two estimates, one must use the standard error of the difference (see Appendix C for details).

[1] Average number of punctuation marks per paper is repeated in this table because it can be used to estimate the numbers implied by the percentages of the different sentence types that fallow.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

A Profile by Grade (Tables 2.3a and 2.3b)

Grade 4

Fourth graders' papers showed an increase in percentage of incorrect word choices in 1996. This decline was reflected in the average of all papers as well as in the lower- and upper-level papers. For spelling errors, there was no change for fourth graders in comparison with 1984.

The comma and dash error rate increased and the percentage of sentences with end mark errors decreased in 1996. The increase in the rate of use of other punctuation did not result in an increase in the error rate for other types of punctuation.

Grade 8

For grade 8, there was no change in percentage of incorrect word choice or spelling errors, even though the essays were longer than in 1984. The number of punctuation marks used increased for eighth graders in 1996, while the punctuation error rate, not including omissions, declined. The only other change in the components of punctuation error rates was a decline in end mark error rate.

Grade 11

The rate of spelling errors did not change even though the number of words increased in 1996. With the increase in paper length, the number of all punctuation marks used, the use rates of end marks and of other punctuation have all risen compared to 1984. At the same time, the overall punctuation error rate (not including omissions) has declined. Although the rate of end mark use has risen, the end mark error rate was unchanged, and the percentage of sentences with end mark errors was lower than in 1984.

Summary

Fluency in Writing

Modest improvements in writing fluency between 1984 and 1996 are seen in fourth- and eighth-grade students' essays (Table 1.1 – Table 1.3).

At grade 4, holistic scoring of the persuasive task "Spaceship" showed no overall increase in students' writing fluency between 1984 and 1996. However, there was a significant increase in the percentage of papers rated in the upper half of the holistic scale (that is, papers receiving a rating of 4, 5, or 6). Fourth graders writing for the narrative task "Flashlight" showed an increase in the percentage of papers receiving a rating of 4, but no change in the overall rating of performance between 1984 and 1996.

Eighth graders' essays showed improvement in 1996 on both of the tasks analyzed holistically, the informative task "Food on the Frontier" and the persuasive task "Recreation Opportunities." At grade 11, no change was seen in writing fluency on either of these tasks when comparing 1996 papers to those written in 1984.

Grammar, Spelling, and Punctuation

Differences in the use of grammar, spelling, and punctuation conventions between 1984 and 1996 were primarily in the direction of improvement at grades 8 and 11. For both 8th and 11th graders, the percentage of awkward sentences and punctuation error rates decreased, even as papers contained more sentences and more words. But there was a more mixed picture at grade 4: Fourth graders showed a decrease in one kind of error but an increase in three other kinds of errors.

English language conventions were examined in papers written in 1984 and 1996 for the task "Spaceship" at grade 4 and for the task "Recreation Opportunities" at grades 8 and 11. A subsample of papers from 1984 and 1996 had been coded by experts so that students' control of the conventions of the English language could be analyzed. Many of the results of these analyses appear in Tables 2.1 - 2.3. Overall, these indicators of performance at the three grades suggest that there have been some changes in students' mastery of English language conventions between 1984 and 1996.

Table 2.1 shows that the number of words and sentences written by eighth- and eleventh-grade students has increased since 1984. Over the same period, however, there has been no change in the rate of errors (number of errors per 100 words) in all three grades.

While there were increases in percentages of sentence fragments in fourth- and eighth-grade papers (Table 2.2b), there were declines in the percentage of awkward sentences in eighth and eleventh graders' papers compared to 1984.

At all grades, the percentage of spelling errors has remained unchanged, comparing 1984 to 1996 (Table 2.3a). The percentage of incorrect word choices is unchanged in grades 8 and 11 but has increased at grade 4. From Table 2.3b, at grades 8 and 11, the punctuation error rate decreased while the number of punctuation marks used per paper increased.

Figure S.1 provides a synopsis of Tables 2.1 – 2.3, comparing student use of grammar, punctuation, and spelling conventions in 1984 and 1996. The column on the left indicates which table contains the data for the comparison. Measures in the first section are characteristics desirable for these NAEP long-term trend writing tasks. An increase in prevalence in 1996 compared to 1984 is desirable, and "increase" is shown in bold text. Decreases in 1996 compared to 1984 are undesirable, and those cells are shaded. The middle section contains characteristics that are neutral; that is, changes in these measures are of interest, but there is no clear advantage or disadvantage to either increase or decrease. The lower section contains measures of writing error, with the notation of increase or decrease. In this section, decreases are desirable and are in bold text, while increases (undesirable) are shaded. Throughout the table, empty cells indicate that no statistically significant change occurred in 1996 compared to 1984.

Of the measures of students' control of writing reported here, at grade 4, one measure of writing error showed improvement (that is, the percentage of sentences with end mark errors declined), while three showed increases in error rate (that is, the prevalence of sentence fragments, incorrect word choice, and comma/dash errors increased). At grade 4, there was no change in most characteristics reported here. At grades 8 and 11, students are writing more in 1996, although the rate at which they use more sophisticated sentence constructions has decreased. At grade 8, the good news was that two desirable characteristics improved and three errors decreased, and only two changes indicated problems (that is, the use of complex sentence structures decreased while sentence fragments increased). At grade 11, improvement occurred in two desirable characteristics and for three types of error, while only one change in a desirable characteristic indicated a problem (that is, the use of complex sentence structures decreased).

Figure S.1 Average change from 1984 to 1996 in writing mechanics measures in this report

THE NATION'S REPORT CARD

Table number	Mechanics measures	Grade 4	Grade 8	Grade 11
	DESIRABLE CHARACTERISTICS—Increase shows improvement			
2.1	Average number of full words per paper		**increase**	**increase**
2.1	Average number of sentences per paper		**increase**	**increase**
2.1	Average number of words per sentence			
2.2a	Percentage of complex or compound sentences		decrease	decrease
	NEUTRAL CHARACTERISTCS			
2.2a	Percentage of simple sentences			increase
2.3b	Average number of punctuation marks used		increase	increase
2.3b	Comma and dash use rate			
2.3b	End mark use rate			increase
2.3b	Other punctuation use rate	increase		increase
	WRITING ERRORS—decrease shows improvement			
2.1	Average number of all errors per 100 words			
2.2b	Percentage of run-on sentences			
2.2b	Percentage of sentence fragments	increase	increase	
2.2b	Percentage of awkward sentences		**decrease**	**decrease**
2.3a	Percentage of incorrect word choice	increase		
2.3a	Percentage of spelling errors			
2.3b	Punctuation error rate (without omissions)		**decrease**	**decrease**
2.3b	Punctuation omission rate			
2.3b	Comma and dash error rate	increase		
2.3b	End mark error rate		**decrease**	
2.3b	Percentage of sentences with end mark errors	**decrease**		**decrease**
2.3b	Other punctuation error rate			

Note: Shading indicates a decline in performance and bold text indicates an improvement since 1984. If neither increase nor decrease is shown, there was no statistically significant change.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Observations

Students at all three grade levels wrote at least as fluently in 1996 as they did in 1984, while students at eighth grade demonstrated improved fluency on the informative and persuasive tasks.

When writing mechanics are considered, the overall picture is of improvement in grades 8 and 11, but there are several declines at grade 4. In the eighth and eleventh grades, students wrote more, as indicated by the increase in the number of words and sentences in their responses, while demonstrating no increase in the number of errors per 100 words. A summary of the measures can be seen in the previous table.

Thus, increased instructional emphasis on writing processes over the 12 years between 1984 and 1996 appears associated with modest improvements in students' mastery of the conventions of written English at grades 8 and 11. During this time period, the overall fluency of eighth graders' writing has also improved. It appears that the process approach to writing, in which planning, writing, and revision through several drafts are practiced, gives students the opportunity to write more and to employ editing strategies, which in turn affords them the opportunity to improve their mastery of the writing conventions reported here.

Appendix A: Procedures

Overview of Holistic and Mechanics Scoring Procedures in the NAEP 1996 Long-Term Trend Writing Assessment

This appendix provides more detailed information about the methods and procedures used in NAEP's 1996 long-term trend assessments in writing, in particular concerning writing fluency and mechanics. The *NAEP 1996 Technical Report* contains more extensive information about the long-term trend writing assessment.

The first year of the administration and reporting of this writing assessment was 1984; it was administered and reported again in 1988, 1990, 1992, and 1994. For each ensuing year, the base year and the current year have been reported. This NAEP long-term trend report is based on two writing assessments, that conducted in 1983-1984 and the assessment conducted during the 1995-1996 school year.

In both assessments, the same tasks were administered in the same manner to comparable samples of students (see Table A.1 for sample sizes). 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' success in writing for various purposes. Related objectives were to evaluate the extent to which students managed the writing process and controlled the forms of written language.[13] At each grade, six different writing tasks were administered.

[13] Educational Testing Service (1987). *Writing objectives: 1988 assessment.* Princeton, NJ: Author.

Table A.1 Sample sizes for the writing trend assessments by task and scoring method

THE NATION'S REPORT CARD NAEP

Writing Task	1984 Grade 4	1984 Grade 8	1984 Grade 11	1996 Grade 4	1996 Grade 8	1996 Grade 11
INFORMATIVE			Food on the Frontier			
Scoring:						
primary trait	—	603	629	—	1275	1151
holistic	—	637	663	—	1321	1194
PERSUASIVE			Spaceship			
Scoring:						
primary trait	611	—	—	1212	—	—
holistic	636	—	—	1250	—	—
mechanics	506	—	—	515	—	—
PERSUASIVE			Recreation Opportunities			
Scoring:						
primary trait	—	494	521	—	1315	1235
holistic	—	531	536	—	1350	1269
mechanics	—	473	517	—	537	495
IMAGINATIVE			Flashlight			
Scoring:						
primary trait	609	—	—	603	—	—
holistic	649	—	—	636	—	—

Note: The sample sizes for 1984 holistic scoring reflect numbers of papers rescored in 1996, and so may differ sightly from numbers in previous reports. Sample sizes for primary trait papers are for those scored in the specified year; sample sizes for 1984 mechanics papers were for 1984 papers scored in 1988 as detailed in *Focusing the New Design: The NAEP 1988 Technical Report.*

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Scoring the Writing Tasks

Materials from NAEP 1996 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 essay questions were evaluated by trained staff using guidelines prepared by NAEP. Each processing activity was conducted with rigorous quality control. An overview of the professional scoring for writing as well as for mathematics and reading is in the Procedural Appendix in the *NAEP 1996 Trends in Academic Progress* (no constructed-response questions were scored for science).

Holistic Scoring

Selected tasks included in the trend assessment were scored holistically for overall fluency. 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 six-point scale. (See the scoring guides for informative, narrative, and persuasive writing in Appendix B.)

A chief reader and assistant chief reader, chosen for their expertise in holistic scoring, anchored the holistic scale. They, together with the table leaders and ETS staff members, studied the pool of 1996 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 all six 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 changes 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, 1994, and 1996 assessments were randomly mixed prior to scoring. Thus, if more responses from one or another assessment were judged to be in the upper half of the scale, the results would indicate changes across time in overall writing fluency.

Agreement Among Raters (Table A.2)

In order to ensure that all readers conform to the standards presented in the scoring guide, 20 percent of the 1996 responses were scored by a second reader to provide information on interrater agreement.

The **exact agreement** category shows the rate at which both the first and second reader awarded the same score to an essay. Exact agreement rates in the 50s are typical for holistic scorings using six-level rubrics. **Adjacent agreement** shows the rate at which scores given by the first and second readers were no more than one score point apart (e.g., the first reader assigned a score of 4 and the second reader assigned a score of 5 or 3). The criterion for an acceptable level of adjacent agreement for six-level holistic scorings is usually 80 percent. **Interrater reliability** correlation demonstrates the extent to which readers agreed with one another overall. Interrater correlations of .80 or above are usually accepted as indicating a sufficient degree of reliability for six-level holistic scorings. Thus, the data in Table A.2 indicate that the holistic scoring met reasonable criteria for interrater reliability for a holistic scoring using a six-level holistic rubric (C. A. Gentile, personal communication, January 25, 1999).[14]

[14] Most testing programs do not publish their rates of desired or acceptable interrater agreement. However, most state testing programs that use holistic scoring to evaluate students' writing achievement strive to have adjacent agreement percentages at 80 or above and interrater correlations of .80 or above.

Table A.2

Agreement and reliability of holistic scores in 1996: A second rater rescored 20% of the papers

Percentage exact agreement, percentage adjacent scores, and correlation of both scorers' ratings overall

THE NATION'S REPORT CARD

	Grade 4			Grade 8			Grade 11		
	Percentage		Correlation	Percentage		Correlation	Percentage		Correlation
Writing task	Exact agreement between first and second rater	Adjacent scores by first and second rater	Interrater reliability	Exact agreement between first and second rater	Adjacent scores by first and second rater	Interrater reliability	Exact agreement between first and second rater	Adjacent scores by first and second rater	Interrater reliability
Spaceship	59.6	94.0	0.86	—	—	—	—	—	—
Flashlight	55.3	93.7	0.82	—	—	—	—	—	—
Recreation Opportunities	—	—	—	58.9	95.5	0.85	53.7	95.1	0.82
Food on the Frontier	—	—	—	53.2	93.9	0.81	50.2	91.2	0.82

—The task was not administered at this grade.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Correlation Between Holistic and Primary Trait Scores (Table A.3)

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.3, the correlations range from 0.34 to 0.75. 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. The holistic score provides a general measure of writing fluency, since the scores raters assign are affected by the writer's attention to organization, adherence to the conventions of written English, word choice, and quality of ideas.

Table A.3 **Holistic and primary trait scores: Correlation between two types of scores in 1984 and 1996**

THE NATION'S REPORT CARD

Writing task	1984 Papers			1996 Papers		
	Grade 4	Grade 8	Grade 11	Grade 4	Grade 8	Grade 11
Spaceship	0.66	—	—	0.73	—	—
Flashlight	0.64	—	—	0.75	—	—
Recreation Opportunities	—	0.34	0.48	—	0.57	0.60
Food on the Frontier	—	0.43	0.43	—	0.66	0.68

—The task was not administered at this grade.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

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 1996 writing responses using the mechanics scoring criteria it used to evaluate writing responses from the 1984, 1988, 1990, 1992, and 1994 assessments.[15] 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 500 responses to each task at each grade level was selected for evaluation (see Table A.1). Readers were trained by practicing on a 10-percent sample of the 1996 papers. Another 10-percent sample of essays previously scored for mechanics from the 1988, 1990, 1992, and 1994 assessments was rescored for reliability. A comparison of the 1996 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 in Appendix B.

Two raters scored the papers and the table leader resolved any discrepancies between the raters. 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.

Further details of writing scoring are available in *Focusing the New Design: The NAEP 1988 Technical Report*[16] and in the *NAEP 1994 Trends in Academic Progress*.[17]

---

[15] Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1987). *Grammar, punctuation, and spelling: Controlling the conventions of written English*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

[16] Johnson, E. G. (1990). Data analysis for the writing assessment. In E. G. Johnson & R. Zwick (Eds.), *Focusing the new design: The NAEP 1988 technical report* (pp. 267-296). Washington, DC: National Center for Education Statistics.

[17] Campbell, J. R., Reese, C. M., O'Sullivan, C., & Dossey, J. A. (1996). *NAEP 1994 trends in academic progress: Achievement of U.S. students in science, 1969 to 1994; mathematics, 1973 to 1994; reading, 1971 to 1994; and writing, 1984 to 1994* (Publication No. NCES 97-095). Washington, DC: National Center for Education Statistics.

Appendix B: Scoring Guides

NAEP HOLISTIC SCORING GUIDE

Informative Task

Scores

6 A **6** essay demonstrates a high degree of competence in response to the prompt, but may have a few minor errors.

An essay in this category generally has the following features:

- is clearly well organized and coherently developed
- clearly explains or elaborates on key ideas
- clearly displays facility in the use of language
- is generally free from errors in mechanics, usage, and sentence structure

5 A **5** essay demonstrates clear competence in response to the prompt, but may have minor errors.

An essay in this category generally has the following features:

- is generally well organized and coherently developed
- explains or illustrates key ideas
- displays facility in the use of language
- contains few errors in mechanics, usage, and sentence structure

4 A **4** essay demonstrates competence in response to the prompt.

An essay in this category generally has the following features:

- is adequately organized and developed
- partially explains or illustrates ideas
- demonstrates adequate facility with language
- may display some errors in mechanics, usage, or sentence structure, but not a consistent pattern of such errors

3 A **3** essay demonstrates some degree of competence in response to the prompt, but is clearly flawed.

An essay in this category reveals one or more of the following weaknesses:

- is inadequately organized or developed
- inadequately explains or illustrates ideas
- demonstrates inappropriate use of language
- reveals a pattern or accumulation of errors in mechanics, usage, or sentence structure

2 A **2** essay demonstrates only limited competence and is seriously flawed.

An essay in this category reveals one or more of the following weaknesses:

- lacks organization or development
- presents no key ideas or does not explain or illustrate ideas given
- displays serious or persistent errors in use of language
- displays serious errors in mechanics, usage, or sentence structure

1 A **1** essay demonstrates fundamental deficiencies in writing skills.

An essay in this category reveals one or more of the following weaknesses:

- is undeveloped
- is incoherent
- contains serious and persistent writing errors

Persuasive Task

Scores

6 A **6** essay demonstrates a high degree of competence in response to the task, but may have a few errors.

An essay in this category generally has the following features:

- is clearly well organized and coherently developed
- supports a stand with a well-developed argument
- clearly demonstrates facility in the use of language
- is generally free from errors in mechanics, usage, and sentence structure

5 A **5** essay demonstrates clear competence in response to the assignment, but may have minor errors.

An essay in this category generally has the following features:

- is generally well organized and coherently developed
- supports a stand with an argument or interrelated list of ideas
- demonstrates facility in the use of language
- contains few errors in mechanics, usage, and sentence structure

4 A **4** essay demonstrates competence in response to the task.

An essay in this category generally has the following features:

- is adequately organized and developed
- supports a stand with a few reasons
- demonstrates adequate facility in the use of language
- may display some errors in mechanics, usage, or sentence structure, but not a consistent pattern of such errors

3 A **3** essay demonstrates some degree of competence in response to the task, but is clearly flawed.

An essay in this category reveals one or more of the following weaknesses:

- is inadequately organized or developed
- inadequately supports a stand
- demonstrates inappropriate use of language
- reveals a pattern or accumulation of errors in mechanics, usage, or sentence structure

2 A **2** essay demonstrates only limited competence and is seriously flawed.

An essay in this category reveals one or more of the following weaknesses:

- lacks organization or development
- does not take a stand or does not provide support consistent with stand taken
- displays serious or persistent errors in use of language
- displays serious errors in mechanics, usage, sentence structure, or word choice

1 A **1** essay demonstrates fundamental deficiencies in writing skills.

An essay in this category reveals one or more of the following weaknesses:

- is undeveloped
- is incoherent
- contains serious and persistent writing errors

Narrative Task

Scores

6 A **6** story demonstrates a high degree of competence in response to the prompt but may have a few minor errors.

A story in this category generally has the following features:

- is well developed with a clear narrative structure
- contains considerable detail that enriches the narrative
- clearly demonstrates facility in the use of language
- is generally free from errors in mechanics, usage, and sentence structure

5 A **5** story demonstrates clear competence in response to the prompt, but may have minor errors.

A story in this category generally has the following features:

- is developed with a clear narrative structure
- contains details that contribute effectively to the narrative
- demonstrates facility in the use of language
- contains few errors in mechanics, usage, and sentence structure

4 A **4** story demonstrates competence in response to the prompt.

A story in this category generally has the following features:

- is adequately developed, but may have occasional weaknesses in narrative structure
- contains details that contribute to the narrative
- demonstrates adequate facility in the use of language
- may display some errors in mechanics, usage, or sentence structure, but not a consistent pattern or accumulation of such errors

3 A **3** story demonstrates some degree of competence in response to the prompt but is clearly flawed.

A story in this category reveals one or more of the following weaknesses:

- is somewhat developed, but lacks clear narrative structure
- contains few details that contribute to the narrative
- demonstrates inappropriate use of language
- reveals a pattern or accumulation of errors in mechanics, usage, or sentence structure

2 A **2** story demonstrates only limited competence and is seriously flawed.

A story in this category reveals one or more of the following weaknesses:

- lacks development and/or narrative structure
- contains little or no relevant detail
- displays serious or persistent errors in use of language
- displays serious errors in mechanics, usage, or sentence structure

1 A **1** story demonstrates fundamental deficiencies in writing skills

A story in this category reveals one or more of the following weaknesses:

- is undeveloped
- is incoherent
- contains serious and persistent writing errors

NAEP MECHANICS SCORING GUIDE

I. Sentence Types

1. Simple — A sentence that contains a subject and a verb. It may also contain an object, subject complement, phrase, appositive, nominative absolute, or verbal.[18] Also, it may have 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.
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

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.

---

[18] **A subject complement** is a word or group of words functioning as an adjective or noun that is used in the predicate and describes or is identified with the subject; for example, "As the travelers became sleepy ..." An **appositive** is a noun or noun phrase that identifies another noun or pronoun that immediately precedes it; for example, "Washington, our first president, ..." A **nominative absolute** is a construction of a noun, noun phrase, or pronoun in the nominative case, followed by a predicate lacking a finite verb, used as a loose modifier of the whole sentence; for example, "The play done, the audience left ..." A **verbal** is a word that is derived from a verb but that functions in some other way (gerund, infinitive, participle); for example, "We need to translate the article ..."

c. Illogical construction — A sentence containing a faulty modification or a dangling modifier, or a functionally misarranged or misproportioned sentence.

d. Other dysfunctions — A sentence omitting a word or containing an extra word or using 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.

V. Writing Mechanics Score Characteristics

After the essays were read and marked, the essay text and scoring symbols were entered into the computer, producing a comprehensive set of variables containing elements referred to as SCORE, CODE, PUNC, and PUNCER. These elements form mechanics characteristics, many of which are reported in Chapter 2. The characteristics are listed below, with table numbers for those used in Chapter 2. Following this list is a brief explanation of the elements composing the characteristics, with definitions.

Table #	Characteristic	Description
2.1, 2.3, 2.3a	1 Number of full words per paper	= SCORE 2
	2 Number of letters per word	= SCORE 3
	3 Number of T-units	= SCORE 4
2.1, 2.2a, 2.2b	4 Number of sentences	= SCORE 5
	5 Number of good sentences	= SCORE 6
	6 Number of bad sentences	= SCORE 7
	7 Percent of sentences with agreement errors	= SCORE 8
2.2b	8 Percent of awkward sentences	= SCORE 9
2.2a	9 Percent of simple sentences	= SCORE 10
	10 Percent of compound sentences	= SCORE 11
2.2a	11 Percent of complex, compound sentences	= SCORE 12
2.2b	12 Percent of run-on sentences	= SCORE 13
2.2b	13 Percent of incorrect sentence fragments	= SCORE 14
	14 Percent of good sentences	= SCORE 15
	15 Percent of bad sentences	= SCORE 16
2.3a	16 Percent of spelling errors	= SCORE 18
2.3b	17 Number of punctuation marks used	= SCORE 20
	18 Ratio of punctuation errors to punctuation used	= SCORE 21
	19 Number of words per T-unit	= SCORE 22
2.1	20 Number of words per sentence	= SCORE 23
	21 Number of sentences with agreement errors	= CODE 2
	22 Number of awkward sentences	= CODE 3
	23 Number of simple sentences	= CODE 4
	24 Number of compound sentences	= CODE 5
	25 Number of complex, compound sentences	= CODE 6
	26 Number of run-on sentences	= CODE 7
	27 Number of incorrect sentence fragments	= CODE 8
	28 Number of spelling errors	= CODE 9
	29 Number of incorrect word choices	= CODE 10
	30 Number of capitalization errors	= CODE 11
	31 Number of punctuation errors	= CODE 14
	32 Number of end mark errors	= CODE 15

Table #	Characteristic	Description
	33 Number of comma and dash errors	= CODE 16
	34 Number of other punctuation errors	= CODE 17
	35 Mean number of errors	= SUM of CODE(1) through CODE(12) + Code(14)
	36 Number of end marks used	= SUM of PUNC(3) + PUNC(7) + PUNC(9)
	37 Mean number of commas & dashes used	= SUM of PUNC(8) + PUNC(11)
	38 Number of other punctuation marks used	= SUM of PUNC(1+2+4+5+6+10+12)
2.3a	39 Percent of incorrect word choice	= SCORE 19
2.3b	40 Percent of sentences with end mark errors	= if (SCORE (5).GT. 0) X(40) = CODE(15)/SCORE(5)*100
	41 Number of omitted punctuation marks	= PUNCER(1,12)
	42 Number of wrong or unnecessary punctuation marks	= PUNCER(12,12)-PUNCER(1,12)
	43 Percent of capitalization errors	= CODE(11)/SCORE(2)*100
	44 Total number of errors	= SUM of CODE (2,3,7,8,9,10,11,14)
2.1	45 Error rate	= X(44)/SCORE(2)*100
	46 Punctuation error and omission rate	= CODE(14)/SCORE(2)*100
2.3b	47 Punctuation omission rate	= PUNCER(1,12)/SCORE(2)*100
2.3b	48 Punctuation error rate (not including omissions)	= PUNCER(12,12)-PUNCER(1,12)/SCORE(2)*100
2.3b	49 End mark error rate	= CODE(15)/SCORE(2)*100
2.3b	50 Comma and dash error rate	= CODE(16)/SCORE(2)*100
2.3b	51 Other punctuation error rate	= CODE(17)/SCORE(2)*100
2.3b	52 End mark use rate	= SUM of PUNC(3,7,9) / SCORE(2)*100
2.3b	53 Comma and dash use rate	= PUNC(8)+PUNC(11)/ SCORE(2)*100
2.3b	54 Other punctuation use rate	= SUM of PUNC(1,2,4,5,6,10,12)/ SCORE(2)*100

VI. Elements in Mechanics Score Characteristics

Some of the characteristics used in this report to describe students' writing abilities were taken directly from a SCORE or CODE created by the computer routine. Others, however, were derived from some combination of variables . which may include values from PUNC or PUNCER. The elements of the mechanics characteristics listed above follow.

CODE is an array containing 34 variables, the contents of which are described below.

CODE	Occurrences of:
1	One word written as two
2	A - Agreement errors
3	W - Awkward sentences
4	S - Simple sentences
5	C - Compound sentences
6	X - Complex or compound-complex sentences
7	R - Run-on sentences
8	F - Incorrect sentence fragments
9	E - Spelling errors
10	K - Incorrect word choices
11	T - Capitalization errors
12	L - Illegible words
13	Blanks within asterisk-bounded strings
14	P - Punctuation errors
15	. - end mark errors
16	, - comma and dash errors
17	- - other punctuation errors
18	Asterisk-bounded strings
19	. - preceding P symbol
20	, - preceding P symbol
21	- - preceding P symbol
22	. - following P symbol
23	, - following P symbol
24	- - following P symbol
25-34	A,W crossed with S,C,X,R,F

PUNC is a 16-word array containing punctuation used by the student and delimiter counts. The counts in PUNC(1-12,15) are from the student's writing sample and do not include symbols or punctuation that were placed in the writing sample by NAEP scoring staff. PUNC(14 & 15) are counts of the scoring delimiters (* and <) used by NAEP scoring staff. The contents of PUNC are described below.

PUNC	**Occurrences of:**
1-12	( ) ! / : ; ? , ." - '
	Counts of actual punctuation used by the student
13	< NAEP scoring flag
14	* NAEP scoring flag
15	Nontrailing blanks (=0 when text is all blanks)
16	=1 if no delimiter is observed at the end of text

PUNCER is a 12 x 12 array containing 144 punctuation error counts. These punctuation error counts are obtained when a **P** (punctuation error) code is identified in an asterisk-bounded string. NAEP scorers placed the incorrect punctuation mark before the **P**, and the correct punctuation after the **P**. The rows in PUNCER represent any incorrect punctuation that was used; the columns represent the correct punctuation. As an example, PUNCER(3,2) would contain a count of the number of times a given student used a comma when the correct mark should have been a period. The contents of the rows and columns of PUNCER are described below.

PUNCER

Row/Column

Location	Occurrences of:
1	NULL (represents the absence of a punctuation symbol)
2	. (represents an end mark of . ? or !)
3	, (represents commas and dashes)
4	- (represents other punctuation " ' : ;)
5	., (both symbols used)
6	.- (both symbols used)
7	,. (both symbols used)
8	,- (both symbols used)
9	-. (both symbols used)
10	-, (both symbols used)
11	All other combinations of the three allowed punctuation symbols ., -
12	Sum of row/column counts

SCORE is an array containing summary counts, percentages, and scores derived from a student's writing sample. The contents of SCORES are described below.

SCORE	Description
1	Number of letters (adjusted for illegibles)
2	Number of words (adjusted for double words and illegibles)
3	Average number of letters per word
4	T-units score
5	Number of sentences
6	Number of good sentences
7	Number of bad sentences
8	Percentage of sentences with agreement errors
9	Percentage of awkward sentences
10	Percentage simple sentences
11	Percentage of compound sentences
12	Percentage of complex sentences
13	Percentage of run-on sentences
14	Percentage of fragmented sentences
15	Percentage of good sentences
16	Percentage of bad sentences
17	Percentage of end mark error sentences
18	Percentage of spelling errors
19	Percentage of word choice errors
20	Total number of punctuation marks used
21	Ratio of punctuation errors/punctuation used
22	Number of words/T-units
23	Number of words/number of sentences

The numbers of letters and words returned in the variables SCORE(1) and SCORE(2) were adjusted when a student's writing sample contained illegible words. The number of letters was adjusted by computing the average number of letters per word and adding the product of the number of illegibles and the average number of letters to SCORE(1). The number of words was adjusted by adding the number of illegibles to SCORE(2). The routine also adjusted the counts of words and letters when it encountered certain text strings. More detail can be found in the user guide, obtainable on request from the authors.

Appendix C: Tables with Standard Errors

The comparisons presented in this report are based on statistical tests that consider the magnitude of the difference between group averages or percentages and the standard errors of those statistics. This appendix contains the tables in Chapter 1 and Chapter 2, with standard errors included to the right of each statistic.

Table C1.1 Fluency in an informative task, "Food on the Frontier," comparing base year to current year with standard errors

THE NATION'S REPORT CARD NAEP

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers			
	Grade 8		Grade 11	
	1984	1996	1984	1996
0	5.5 (1.1)	5.0 (0.6)	5.4 (0.8)	5.3 (0.8)
1	11.3 (1.7)	7.3 (1.1)	4.6 (1.0)	4.5 (0.6)
2	22.6 (1.9)	18.0 (1.3)	13.3 (1.9)	9.3 (0.9)
3	31.5 (2.1)	32.3 (1.6)	21.5 (2.2)	25.6 (1.8)
4	21.1 (2.1)	24.1 (1.3)	32.1 (1.9)	30.9 (1.3)
5	6.0 (1.2)	10.3 (1.3)	16.1 (1.4)	18.0 (1.2)
6	1.9 (0.7)	3.0 (0.6)	7.0 (0.9)	6.3 (0.9)
4, 5, or 6	29.0 (2.4)	37.4 (2.0)	55.1 (2.5)	55.2 (1.6)
Average rating	2.77 (0.06)	3.06 (0.06)*	3.46 (0.07)	3.51 (0.05)

*Statistically significant difference from 1984.

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.

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C1.2 Fluency in persuasive tasks, "Spaceship" and "Recreation Opportunities," comparing base year to current year, with standard errors

THE NATION'S REPORT CARD

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers					
	Spaceship		Recreation Opportunities			
	Grade 4		Grade 8		Grade 11	
	1984	1996	1984	1996	1984	1996
0	4.4 (0.8)	6.4 (0.8)	5.8 (0.9)	3.8 (0.6)	2.4 (0.6)	4.0 (0.6)
1	12.7 (1.4)	7.6 (0.8)*	8.8 (1.4)	7.0 (0.6)	2.0 (0.7)	2.5 (0.6)
2	22.4 (1.8)	18.9 (1.3)	12.7 (1.5)	14.5 (0.9)	11.2 (1.4)	9.1 (1.0)
3	42.3 (2.5)	41.8 (1.8)	36.0 (2.3)	29.7 (2.0)	27.5 (2.2)	27.0 (1.7)
4	15.2 (1.3)	20.4 (1.3)	28.8 (1.9)	32.3 (1.4)	37.6 (2.1)	36.5 (1.7)
5	2.6 (0.7)	4.3 (0.6)	6.3 (1.5)	10.3 (1.3)	12.8 (1.7)	15.6 (1.4)
6	0.5 (0.3)	0.6 (0.2)	1.5 (0.5)	2.4 (0.4)	6.6 (1.3)	5.4 (0.8)
4, 5, or 6	18.3 (1.6)	25.3 (1.6)*	36.7 (2.4)	45.1 (1.9)	57.0 (2.3)	57.5 (2.1)
Average rating	2.61 (0.04)	2.78 (0.04)	2.98 (0.06)	3.21 (0.04)*	3.61 (0.06)	3.58 (0.06)

*Statistically significant difference from 1984.

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.

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C1.3 Fluency in a narrative task, "Flashlight," comparing base year to current year with standard errors

THE NATION'S REPORT CARD

Holistic rating	Percentage of papers at each holistic rating; average holistic rating for all papers	
	Grade 4	
	1996	1994
0	6.2 (1.2)	5.4 (1.1)
1	10.4 (1.2)	6.6 (1.1)
2	24.6 (2.2)	24.2 (2.5)
3	30.1 (1.7)	26.6 (1.8)
4	14.8 (1.5)	22.6 (1.9)*
5	11.0 (1.3)	10.7 (1.4)
6	3.0 (0.8)	3.8 (0.9)
4, 5, or 6	28.8 (2.5)	37.2 (2.4)
Average rating	2.82 (0.08)	3.02 (0.08)

*Statistically significant difference from 1984.

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.

Note: Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C2.1

Overall characteristics of papers: 1984 and 1996 with standard errors

THE NATION'S REPORT CARD NAEP

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of full words per paper	4	1996	35.4 (1.17)	28.4 (1.04)	53.9 (2.65)
		1984	33.8 (0.96)	27.9 (0.83)	47.7 (1.82)
	8	1996	79.4 (2.65)*	59.2 (2.10)*	104.2 (3.82)*
		1984	67.5 (1.87)	51.1 (1.38)	89.5 (2.92)
	11	1996	104.4 (3.40)*	68.6 (2.93)	124.4 (3.60)
		1984	93.3 (2.34)	62.0 (2.15)	115.0 (2.88)
Average number of sentences per paper	4	1996	2.6 (0.09)	2.1 (0.07)	3.9 (0.25)
		1984	2.6 (0.10)	2.2 (0.09)	3.6 (0.20)
	8	1996	5.2 (0.19)*	3.8 (0.13)	6.8 (0.29)*
		1984	4.4 (0.13)	3.4 (0.11)	5.8 (0.23)
	11	1996	6.5 (0.22)*	4.1 (0.17)	7.8 (0.27)*
		1984	5.6 (0.15)	3.7 (0.17)	6.9 (0.21)
Average number of words per sentence	4	1996	16.1 (0.47)	16.1 (0.55)*	16.0 (0.69)
		1984	15.1 (0.35)	14.3 (0.37)	16.8 (0.86)
	8	1996	17.7 (0.53)	18.3 (0.81)	16.9 (0.38)
		1984	17.3 (0.38)	17.5 (0.54)	17.0 (0.56)
	11	1996	17.7 (0.42)	18.8 (0.86)	17.0 (0.43)
		1984	18.2 (0.43)	18.8 (0.81)	17.7 (0.38)
Average number of all errors per 100 words	4	1996	17.2 (0.67)	19.2 (0.85)	11.8 (0.85)
		1984	15.5 (0.61)	17.1 (0.84)	11.6 (0.55)
	8	1996	10.2 (0.59)	12.1 (0.73)	7.9 (0.69)
		1984	9.2 (0.34)	10.3 (0.51)	7.7 (0.27)
	11	1996	7.4 (0.39)	9.2 (0.69)	6.3 (0.36)
		1984	7.0 (0.17)	8.4 (0.29)	6.0 (0.27)

*Statistically significant difference from 1984 at the 95 percent certainty level. The standard errors of the estimated statistics appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Definitions of these measures appear in Appendix B.

Note: For the first three characteristics, average number of full words divided by average number of sentences does not necessarily produce average number of words per sentence. Consider a simplified example using four papers, with two papers each containing one sentence with 20 words and two papers each containing 10 sentences with 100 words. In this example, the average number of words would be 60, the average number of sentences would be 5.5, and the average number of words per sentence would be 15. Consequently, in this example, words per paper divided by average number of sentences per paper yields 10.5 (instead of 15) as the average number of words per sentence.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C2.2a

Number and types of sentences: 1984 and 1996, with standard errors

THE NATION'S REPORT CARD NAEP

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of sentences per paper	4	1996	2.6 (0.09)	2.1 (0.07)	3.9 (0.25)
		1984	2.6 (0.10)	2.2 (0.09)	3.6 (0.20)
	8	1996	5.2 (0.19)*	3.8 (0.13)	6.8 (0.29)*
		1984	4.4 (0.13)	3.4 (0.11)	5.8 (0.23)
	11	1996	6.5 (0.22)*	4.1 (0.17)	7.8 (0.27)*
		1984	5.6 (0.15)	3.7 (0.17)	6.9 (0.21)
Percent of simple sentences	4	1996	21.9 (1.49)	20.5 (1.78)	25.6 (2.97)
		1984	23.1 (1.53)	23.3 (1.69)	22.2 (22.2)
	8	1996	34.9 (1.31)	30.0 (1.71)	40.9 (1.51)
		1984	33.7 (1.49)	32.3 (2.00)	35.6 (2.51)
	11	1996	40.9 (1.15)*	31.8 (1.80)	46.0 (1.45)*
		1984	35.7 (1.44)	32.7 (2.35)	37.8 (1.49)
Percent of complex or compound sentences	4	1996	52.0 (1.66)	52.6 (1.89)	50.3 (3.25)
		1984	54.8 (1.88)	56.4 (2.43)	51.2 (2.80)
	8	1996	44.8 (1.34)*	45.7 (2.09)	43.8 (1.61)
		1984	49.8 (1.36)	50.6 (2.05)	48.6 (1.75)
	11	1996	44.0 (1.27)*	50.5 (2.06)	40.4 (1.35)*
		1984	52.4 (1.42)	54.3 (2.43)	51.1 (1.29)

*Statistically significant difference from 1984 at the 95 percent certainty level. The standard errors of the estimated statistics appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C2.2b

Control of sentence structure: 1984 and 1996 with standard errors

THE NATION'S REPORT CARD

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of sentences per paper	4	1996	2.6 (0.09)	2.1 (0.07)	3.9 (0.25)
		1984	2.6 (0.10)	2.2 (0.09)	3.6 (0.20)
	8	1996	5.2 (0.19)*	3.8 (0.13)	6.8 (0.29)*
		1984	4.4 (0.13)	3.4 (0.11)	5.8 (0.23)
	11	1996	6.5 (0.22)*	4.1 (0.17)	7.8 (0.27)*
		1984	5.6 (0.15)	3.7 (0.17)	6.9 (0.21)
Percent of run-on sentences	4	1996	14.9 (1.39)	15.8 (1.70)	12.5 (2.52)
		1984	15.3 (1.46)	14.5 (1.71)	17.3 (2.89)
	8	1996	9.2 (1.04)	12.0 (1.46)	5.7 (1.31)
		1984	7.2 (0.91)	8.6 (1.60)	5.4 (0.89)
	11	1996	6.2 (0.96)	8.7 (2.19)	4.9 (0.80)
		1984	4.6 (0.68)	6.4 (1.14)	3.4 (0.77)
Percent of sentence fragments	4	1996	5.8 (0.65)*	6.1 (0.72)	5.0 (1.19)
		1984	3.2 (0.47)	3.2 (0.61)*	3.1 (0.83)
	8	1996	5.2 (0.57)*	6.4 (0.90)	3.6 (0.60)
		1984	3.3 (0.53)	2.7 (0.55)*	4.0 (1.03)
	11	1996	3.8 (0.47)*	5.7 (1.23)	2.8 (0.38)
		1984	3.0 (0.42)	3.0 (0.59)	3.0 (0.61)
Percent of awkward sentences	4	1996	26.1 (1.78)	28.7 (2.28)	19.2 (2.25)
		1984	25.4 (2.16)	28.2 (2.56)	18.6 (2.65)
	8	1996	22.5 (1.29)*	26.0 (2.04)*	18.1 (1.74)*
		1984	32.3 (1.46)	36.9 (2.55)	26.1 (1.62)
	11	1996	19.0 (1.37)*	26.2 (3.21)	15.0 (1.28)*
		1984	31.2 (1.67)	36.0 (3.28)	27.7 (1.51)

*Statistically significant difference from 1984 at the 95 percent certainty level. The standard errors of the estimated statistics appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C2.3a

Control of word-level conventions: 1984 and 1996, with standard errors

THE NATION'S REPORT CARD NAEP

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of full words per paper	4	1996	35.4 (1.17)	28.4 (1.04)	53.9 (2.65)
		1984	33.8 (0.96)	27.9 (0.83)	47.7 (1.82)
	8	1996	79.4 (2.65)*	59.2 (2.10)*	104.2 (3.82)*
		1984	67.5 (1.87)	51.1 (1.38)	89.5 (2.92)
	11	1996	104.4 (3.40)*	68.6 (2.93)	124.4 (3.60)
		1984	93.3 (2.34)	62.0 (2.15)	115.0 (2.88)
Percent of incorrect word choice	4	1996	1.8 (0.19)*	2.0 (0.25)*	1.2 (0.22)*
		1984	0.8 (0.08)	0.9 (0.11)	0.4 (0.08)
	8	1996	0.8 (0.09)	1.1 (0.12)	0.5 (0.11)
		1984	0.7 (0.07)	0.7 (0.10)	0.6 (0.08)
	11	1996	0.5 (0.05)	0.6 (0.09)	0.5 (0.05)
		1984	0.6 (0.06)	0.8 (0.12)	0.5 (0.05)
Percent of spelling errors	4	1996	7.8 (0.31)	8.7 (0.41)	5.4 (0.53)
		1984	8.4 (0.44)	9.0 (0.61)	6.8 (0.56)
	8	1996	3.9 (0.23)	4.7 (0.35)	2.9 (0.24)
		1984	3.7 (0.21)	4.2 (0.32)	3.1 (0.19)
	11	1996	2.7 (0.20)	3.2 (0.37)	2.4 (0.21)
		1984	2.3 (0.14)	2.7 (0.19)	2.0 (0.16)

*Statistically significant difference from 1984 at the 95 percent certainty level. The standard errors of the estimated statistics appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Table C2.3b

Control of punctuation conventions: 1984 and 1996, with standard errors

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
Average number of punctuation marks used	4	1996	3.3 (0.11)	2.6 (0.12)	5.3 (0.30)
		1984	3.0 (0.13)	2.4 (0.12)	4.6 (0.30)
	8	1996	7.8 (0.36)*	5.1 (0.24)	11.1 (0.58)*
		1984	6.3 (0.22)	4.5 (0.21)	8.8 (0.33)
	11	1996	10.2 (0.41)*	6.0 (0.37)	12.6 (0.48)*
		1984	8.2 (0.21)	5.0 (0.28)	10.4 (0.33)
Punctuation error rate (not including omissions)	4	1996	0.2 (0.07)	0.3 (0.09)	0.2 (0.07)
		1984	0.3 (0.07)	0.3 (0.09)	0.3 (0.10)
	8	1996	0.2 (0.03)*	0.2 (0.06)*	0.2 (0.03)
		1984	0.5 (0.06)	0.5 (0.09)	0.4 (0.08)
	11	1996	0.2 (0.03)*	0.1 (0.04)*	0.2 (0.04)*
		1984	0.4 (0.04)	0.4 (0.08)	0.3 (0.05)
Punctuation omission rate	4	1996	2.3 (0.23)	2.5 (0.30)	1.7 (0.18)
		1984	2.3 (0.19)	2.8 (0.25)	1.4 (0.22)
	8	1996	1.4 (0.10)	1.7 (0.15)	1.2 (0.11)
		1984	1.3 (0.09)	1.5 (0.13)	1.1 (0.09)
	11	1996	1.3 (0.17)	1.7 (0.45)	1.1 (0.08)
		1984	1.3 (0.11)	1.8 (0.25)	0.9 (0.11)
Comma and dash use rate	4	1996	0.8 (0.10)	0.8 (0.11)	0.9 (0.21)
		1984	0.6 (0.09)	0.5 (0.10)	1.0 (0.22)
	8	1996	2.4 (0.16)	1.8 (0.22)	3.1 (0.19)
		1984	2.3 (0.12)	2.1 (0.21)	2.6 (0.19)
	11	1996	2.6 (0.11)	1.6 (0.24)	3.1 (0.15)
		1984	2.5 (0.14)	1.9 (0.15)	2.8 (0.23)
Comma and dash error rate	4	1996	1.2 (0.08)*	1.1 (0.10)*	1.3 (0.15)
		1984	0.8 (0.10)	0.7 (0.13)	0.9 (0.17)
	8	1996	1.0 (0.07)	1.0 (0.10)	0.9 (0.07)
		1984	1.0 (0.10)	1.0 (0.15)	0.9 (0.09)
	11	1996	0.8 (0.06)	0.8 (0.12)	0.9 (0.08)
		1984	0.9 (0.06)	1.1 (0.11)	0.8 (0.09)

(continued)

Table C2.3b

Control of punctuation conventions: 1984 and 1996, with standard errors (continued)

THE NATION'S REPORT CARD

Mechanics measure	Grade	Year	Overall Average	Holistic scores 1,2,3 Lower half of scale	Holistic scores 4,5,6 Upper half of scale
End mark use rate	4	1996	7.0 (0.18)	6.8 (0.21)	7.5 (0.43)
		1984	6.9 (0.20)	6.9 (0.26)	7.0 (0.27)
	8	1996	6.3 (0.15)	6.2 (0.21)	6.4 (0.15)
		1984	6.1 (0.12)	5.9 (0.19)	6.4 (0.14)
	11	1996	6.0 (0.13)*	5.6 (0.24)	6.2 (0.14)*
		1984	5.6 (0.13)	5.5 (0.25)	5.6 (0.12)
End mark error rate	4	1996	1.1 (0.19)	1.4 (0.27)	0.4 (0.10)
		1984	1.6 (0.16)	2.1 (0.21)	0.5 (0.09)
	8	1996	0.5 (0.05)*	0.6 (0.09)	0.3 (0.04)
		1984	0.7 (0.08)	0.9 (0.13)	0.4 (0.06)
	11	1996	0.5 (0.15)	1.0 (0.40)	0.3 (0.04)
		1984	0.7 (0.10)	1.1 (0.25)	0.4 (0.05)
Percent of sentence with end mark errors	4	1996	13.0 (1.40)*	15.3 (1.70)*	6.8 (1.77)
		1984	19.0 (1.59)	23.5 (2.00)	8.8 (1.81)
	8	1996	8.8 (0.95)	11.3 (1.68)	5.7 (0.80)
		1984	10.7 (1.02)	13.4 (1.71)	7.2 (0.94)
	11	1996	6.3 (0.84)*	8.6 (2.12)	5.0 (0.65)
		1984	10.3 (0.82)	14.4 (1.83)	7.4 (0.91)
Other punctuation use rate	4	1996	1.6 (0.14)*	1.4 (0.15)	2.1 (0.33)
		1984	1.2 (0.11)	1.0 (0.13)	1.5 (0.22)
	8	1996	0.9 (0.07)	0.8 (0.10)	0.9 (0.09)
		1984	0.8 (0.08)	0.7 (0.11)	1.0 (0.12)
	11	1996	1.0 (0.14)*	1.5 (0.38)	0.8 (0.06)
		1984	0.6 (0.05)	0.6 (0.09)	0.7 (0.06)
Other punctuation error rate	4	1996	0.3 (0.05)	0.3 (0.07)	0.2 (0.06)
		1984	0.3 (0.05)	0.3 (0.07)	0.2 (0.08)
	8	1996	0.2 (0.04)	0.3 (0.06)	0.2 (0.05)
		1984	0.2 (0.03)	0.1 (0.04)	0.2 (0.04)
	11	1996	0.1 (0.02)	0.1 (0.04)	0.1 (0.03)
		1984	0.2 (0.02)	0.1 (0.05)	0.2 (0.03)

* Statistically significant difference from 1984 at the 95 percent certainty level. The standard errors of the estimated statistics appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Definitions of these measures appear in Appendix B.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

Acknowledgments

This special adjunct report on the NAEP 1996 long-term trend holistic and mechanics scoring of student writing is the culmination of the work of many individuals who contributed their considerable knowledge, experience, and creativity.

The 1996 trend assessments were a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). Most importantly, NAEP is grateful to the students and school staff who made the assessment possible.

The 1996 NAEP long-term trend assessments were funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. The Commissioner of Education Statistics, Pascal D. Forgione, and the NCES staff — Peggy Carr, Arnold Goldstein, Steven Gorman — worked closely and collegially with the authors to produce this report.

Activities for the writing holistic and mechanics scoring were conducted at NCS under the management of Charles Brungardt, with help from Betsy Hickok, John Anderson, Mathilde Kennel, and Connie Smith. Training was done at NCS by ETS staff John Kennedy (holistic trainer) and Claudia Gentile (mechanics trainer). Claudia Gentile also provided overall direction for both holistic and mechanics scoring.

At ETS, the holistic and mechanics analyses presented in this report were led by Bruce Kaplan, with the collaboration of Dave Freund, Ingeborg Novatkoski, Katharine Pashley, and Norma Norris, and with the assistance of Shuyi Hua.

The design and production of the report was overseen by Carol Errickson. The considerable production efforts were primarily handled by Karen Damiano, whose unflagging attention to detail was of enormous value. The production of the World Wide Web version of this report was led by Pat O'Reilly with assistance from Debra Kline.

During the early stages of the report, crucial guidance was provided by John Mazzeo. Claudia Gentile gave generously of her time in explaining important aspects of the scoring and in rewriting, with wonderful clarity, sections to reflect her knowledge. Elissa Greenwald's comprehensive understanding of writing pedagogy and assessment, as well as her skill as a practitioner of writing, dramatically increased the readability of the report. Debra Kline's careful reading of earlier versions forced needed consistency on the reporting of two very different measures. Shari Santapau polished the final version.

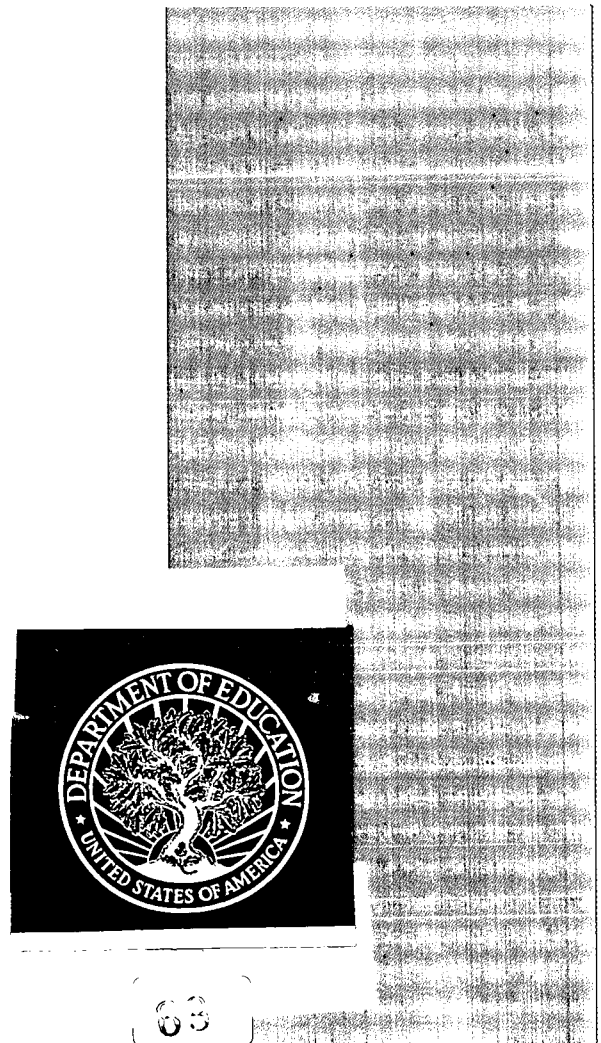
Many thanks are due to the numerous reviewers, both external and internal to NCES, who contributed valuable comments and critical feedback. The reviewers were guided by Arnold Goldstein, NCES, who contributed much to the shaping of the report and also obtained important input from Timothy Shanahan from the University of Illinois, Chicago, and Richard Sterling from the University of California, Berkeley. The report was reviewed closely by Mary Crovo at NAGB, and by Janis Brown, Andrew Kolstad, Sheida White, and Shi-Chang Wu at NCES. Robert Burton and Lee M. Hoffman, also from NCES, contributed importantly to the adjudication process. Christina Kary, of the Education Statistics Services Institute, clarified the adjudication decisions. Finally, Matthew Onstott, from the Office of the Undersecretary, Department of Education, asked questions which helped clarify several issues. Without guidance from these reviewers, this report would have far less utility as documentation for two very interesting facets of the long-term trend writing assessment.

United States
Department of Education
Washington, DC 20208-5653

Official Business
Penalty for Private Use, $300

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

Standard Mail (A)



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").