NCES
National Center for Education Statistics

The Nation's Report Card NAEP TM

U.S. Department of Education
Institute of Education Sciences
NCES 2006–469

The Nation's Report Card

Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading

The National Assessment of Educational Progress

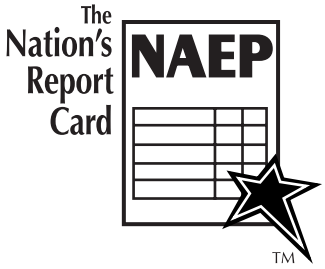
What is The Nation's Report Card™?

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

NAEP is a congressionally mandated project of the National Center for Education Statistics within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations.

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

The National Assessment Governing Board



U.S. Department of Education
Institute of Education Sciences
NCES 2006–469

The Nation's Report Card

Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading

October 2005

Mary C. Daane
Jay R. Campbell
Wendy S. Grigg
Madeline J. Goodman
Andreas Oranje
Educational Testing Service

Arnold Goldstein
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark S. Schneider
Commissioner

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

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

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

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

October 2005

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

Suggested Citation
Daane, M.C., Campbell, J.R., Grigg, W.S., Goodman, M.J., and Oranje, A. (2005). *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

For ordering information on this report, write to
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

or call toll-free 1-877-4ED-PUBS or order online at http:/www.edpubs.org.

Content Contact
Arnold Goldstein
202–502–7344
Arnold.Goldstein@ed.gov

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

Executive Summary

The purpose of *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* is to examine aspects of oral reading performance—accuracy, rate, and fluency—that cannot be observed from results of the main NAEP (National Assessment of Educational Progress) reading assessment. The results provided here are intended to inform educators and researchers about these three aspects of fourth-graders' oral reading performance and how they relate to their overall reading ability as measured by the 2002 reading assessment.

This study focuses on one relevant, but sometimes overlooked, aspect of reading performance—oral reading ability. Oral reading performance, measured by the components of accuracy, rate, and fluency, constitutes a cluster of critical literacy proficiencies and functions as a significant indicator of overall reading ability.

The present report is a follow-up study to the 1992 study, *Listening to Children Read Aloud: Data From NAEP's Integrated Reading Performance Record (IRPR) at Grade 4* (Pinnell et al. 1995). Both were commissioned by the National Assessment Governing Board (NAGB). The 1992 study was NAEP's initial attempt at large-scale measurement of oral reading abilities and one of the first ever performed.

Some of the major findings of the 1992 study include the rating of 55 percent of the participants as fluent, with 13 percent rated fluent at the highest level, based on the same fluency scale used in the present study. Another finding showed a significant relationship between oral reading fluency and reading comprehension, as measured by overall reading proficiency on the main NAEP assessment. Furthermore, a majority of the participants (57 percent) were at least 96 percent accurate in their oral reading of the passage used in the study. Moreover, the students' errors seemed related to overall proficiency only when the errors involved a change in the meaning of the oral reading passage. Results for reading rate showed that 61 percent of students read at least 100 words per minute, and, on average, slower readers demonstrated lower reading proficiency. Overall, positive relationships were found among accuracy, rate, and fluency.

NAEP's 2002 data collection on oral reading used much of the same methodology and approach to understanding and reporting oral reading as the 1992 study; however, the results of the two studies are not comparable because different reading passages and administration procedures were used. The students who participated in the 2002 oral reading study also participated in the main NAEP assessment of reading comprehension; therefore, it is possible to examine the relationship between students' oral reading accuracy, rate, and fluency and their reading comprehension (Dole et al. 1991).

The data in this study were collected from a subsample (1,779) of the sample (140,000) of fourth-graders who participated in the NAEP reading assessment during the early spring of 2002. The data were derived from electronic recordings of the participants reading aloud a 198-word excerpt of "The Box in the Barn," one of the passages the students had encountered one week earlier when they sat for the main NAEP assessment. Only three race/ethnicity groups (White, Black, and Hispanic) were represented in the oral reading study in great enough numbers to report results for these students. Differences in student performance are discussed only if they have been determined by *t* tests in combination with false discovery rate procedures for multiple comparisons to be significant at the .05 level.

A nonresponse analysis was conducted because school and student response rates did not meet NCES statistical standard 3-2-5 concerning achieving desirable response rates. The rates are currently set at 85 percent for NAEP. When the rates are between 70 and 85 percent, an extensive analysis is conducted that examines, among other factors, the potential for nonresponse bias at both the school and student levels. A nonresponse bias analysis was completed by computing weighted response rates for various school- and student-level characteristics of interest and by conducting chi-square tests. The school nonresponse investigated in these analyses is cumulative nonresponse to both NAEP and the study. The only variables not significant in the oral reading study are type of location at the school level and gender and year of birth at the student level. All other variables show a differential rate of nonresponse between subgroups. The final rates were not adjusted as a result of the nonresponse bias analysis specifically, but were adjusted as a result of nonresponse. More details on nonresponse bias analysis can be found in appendix A.

Major Findings

Oral Reading Accuracy

In the context of this study, *accuracy* refers to the degree to which a student's oral reading conforms to the letter-sound conventions of printed English (i.e., accuracy measures the child's precision in orally presenting the words in the text). Accuracy is measured as a percentage of words read correctly.

For example, students whose reading was rated 98 percent accurate made between 0 and 4 errors in reading the 198-word passage. Those with 21 or more errors per 198 words were rated as less than 90 percent accurate.

Accuracy and Comprehension

- Approximately three-quarters of the participating fourth-graders read the passage with at least 95 percent accuracy (figure 2-1).
- Students who read with the fewest errors demonstrated greater comprehension, as measured by their higher average reading scores on the main NAEP reading assessment (figure 2-2).
- The average score for students reading with between 95 and 97 percent accuracy (226) was within the *Basic* achievement level (figure 2-2).
- Those students who read with between 90 and 94 percent accuracy had an average score (206) that was not significantly different from the *Basic* achievement-level cut score.
- Students who read the passage with less than 90 percent accuracy had an average score (180) on the main NAEP reading assessment that fell below the *Basic* level (figure 2-2).

Meaning-Change and Non-Meaning-Change Errors

All errors may not be equally disruptive to a reader's attempt to understand a passage. An error may or may not result in a change of meaning to the text. In the oral reading study, variant pronunciations (such as those arising from regional, dialectical, or nonnative speech) were not considered reading errors unless the mispronunciation altered sentence meaning.

- Occurrences of oral reading errors, regardless of their effect on text meaning, were negatively related to comprehension (figure 2-2).
- Approximately 9 out of 10 students read with at least 95 percent accuracy when only meaning-change errors were counted (figure 2-5).
- When only meaning-change errors were considered, students with higher average accuracy rates also had higher average scale scores (figure 2-6).

Accuracy and Self-Correction of Errors

- Nearly one-half of the students self-corrected at least 50 percent of their meaning-change errors, but only about one-quarter of students self-corrected at least 50 percent of non-meaning-change errors (figures 2-7b and 2-7c).
- Generally, the relationship between the proportion of all errors self-corrected and average reading score is positive—the greater the proportion of errors self-corrected (counting all error types), the higher the average score. Students who corrected more than 75 percent of all errors had higher average scores (237) than students who corrected 50 to 75 percent of their errors (230) (figure 2-8).

Oral Reading Rate

Rate refers to the speed at which the student reads aloud. In this study, rate is measured as both the number of words per minute for the entire performance and the number of words in the initial minute of oral reading.

Reading Rate Measures

- The fourth-graders spent an average of 1 minute and 40 seconds reading the 198-word oral reading passage.
- The fourth-graders' average reading rate across the entire passage was 119 words per minute.
- Nearly two-thirds (65 percent) of fourth-graders read the passage with an average rate of at least 105 words per minute for the entire passage (figure 3-1).
- Speed of oral reading measured as words per minute for the entire passage was positively related to comprehension as measured by average score on the main NAEP assessment (figure 3-2).

Reading Rate and Comprehension

- When reading rate is calculated as words per minute for the entire passage, the average score for students who read less than 80 words per minute was 185, and the average score for students who read between 80 and 104 words per minute was 207. The average score for students who read between 105 and 129 words per minute was 225, and the average score for students who read at least 130 words per minute was 244 (figure 3-2).
- Nearly four-fifths (approximately 79 percent) of fourth-graders read the passage at a rate of at least 105 words per minute for the first minute of reading (figure 3-5).
- When reading rate is calculated as the number of words read for the first minute of oral reading, the average score for students who read less than 80 words was 176, and the average score for students who read between 80 and 104 words was 196. The average score for students who read between 105 and 129 words was 216, and the average score for students who read 130 words or more was 238 (figure 3-6).
- Speed of oral reading as measured for the first minute of reading was positively related to comprehension as measured by average score on the main NAEP assessment (figure 3-6).

Oral Reading Fluency

In this study, *fluency* was considered a distinct attribute of oral reading separate from accuracy and rate. Fluency was defined in terms of phrasing, adherence to the author's syntax, and expressiveness and was measured at one of four levels (1–4, with 4 being the measure of highest fluency) on NAEP's Oral Reading Fluency Scale.

Fluency Measures

- The oral reading of approximately 61 percent of fourth-graders was characterized as fluent—that is, in the top two levels of the fluency scale (figures 4-1 and 4-2).
- Ten percent of the students scored at the highest level of the fluency scale, indicating that they read with phrasing that was consistent with the author's syntax and with some degree of expressiveness (figures 4-1 and 4-2).

Fluency and Comprehension

- Fourth-grade students' oral reading fluency showed a positive relationship to their reading comprehension (i.e., more-fluent readers also demonstrated higher comprehension) (figure 4-3).
- Students whose oral reading was rated nonfluent—that is, in the lower two levels of the fluency scale—performed at or below the *Basic* achievement level on the main NAEP reading assessment. The average main NAEP reading score for students rated at fluency level 2 (207) was not significantly different from the *Basic* achievement-level cut score (208), and the average main NAEP reading score for students rated at fluency level 1 (177) fell 31 points below the *Basic* achievement-level cut score (figure 4-3).

The Relation Between Three Measures of Oral Reading Ability—Accuracy, Rate, and Fluency

- Students rated in the two lower levels on the fluency scale were reading with lower than 95 percent accuracy, on average.
- Overall, the data from this study indicate that the three separate oral reading abilities—accuracy, rate, and fluency—are related to each other, and all three are related to reading comprehension.
- Fourth-grade students rated as fluent readers were more likely to be accurate and to read at a faster rate than students rated as nonfluent (figures 4-6 and 4-7).
- The majority of the nonfluent readers were reading at an average rate of fewer than 105 words per minute (figure 4-7).

Accuracy, Rate, and Fluency Results for Subgroups

- Accuracy results for gender and racial/ethnic groups parallel the findings of the main NAEP reading assessment. Those groups of students who performed at higher levels on the main assessment also were more likely to read with greater accuracy and at a faster rate in the oral reading study (figure 2-3 and 2-4).
- When accuracy was measured as a percentage of words read accurately, 37 percent of female students as compared to 32 percent of male students read with at least 98 percent accuracy (figure 2-3). Thirty-eight percent of White students, 23 percent of Black students, and 31 percent of Hispanic students read with at least 98 percent accuracy (figure 2-4).
- Those groups of students who performed at higher levels on the main NAEP assessment were also more likely to read at a faster rate. When rate was measured as words per minute for the entire passage, approximately 44 percent of female fourth-graders read at an average rate of at least 130 words per minute, as did 33 percent of male fourth-graders (figure 3-3). Forty-five percent of White students read at an average rate of at least 130 words per minute, as did 18 percent of Black students and 24 percent of Hispanic students (figure 3-4).
- When rate was measured as words per minute in the first minute of oral reading, 60 percent of female students, as compared to 53 percent of male students, read at a rate of at least 130 words per minute during the initial minute (figure 3-7). Sixty-four percent of White students, 35 percent of Black students, and 45 percent of Hispanic students read at least 130 words in the first minute (figure 3-8).
- Consistent with reading comprehension results in the main assessment, a greater percentage of female students (64 percent) than male students (56 percent) were rated as fluent when fluency levels 3 and 4 are combined (see figure 4-4). A greater percentage of White students (68 percent) were rated as fluent when fluency levels 3 and 4 are combined compared to their Black (40 percent) and Hispanic (46 percent) peers (figure 4-5).

Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 2002 oral reading study and reading assessment. The special study and the assessment were collaborative efforts among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and Pearson Educational Measurement. Most importantly, NAEP is grateful to the students and school staff who made the special study and the assessment possible.

The NAEP 2002 oral reading study and reading assessment were funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. The Acting Commissioner of Education Statistics, Grover J. Whitehurst, and the NCES staff—Arnold Goldstein, Steven Gorman, Carol Johnson, Marilyn Binkley, Andrew Kolstad, and Marilyn Seastrom—worked closely and collegially with the authors to produce this report.

The NAEP project at ETS is directed by Stephen Lazer and John Mazzeo, with assistance from John Barone. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities for the main reading assessment were conducted by Pearson Educational Measurement under the direction of Brad Thayer, Connie Smith, and William Buckles. Development of the software for administering the oral reading study, for recording digitally students' oral reading, and for managing the oral reading scoring database was accomplished by Ordinate under the direction of Jared Bernstein, Isabella Barbier, and Ognjen Todic.

The 2002 oral reading study procedures were developed as part of the 1992 Integrated Reading Performance Record special study. A committee of reading experts was formed to oversee the development of that special study. The members of that committee were Philip Gough, Barbara Kapinus, John Pikulski, Gay Sue Pinnell, and Karen Wixson.

The coding of students' oral reading in 2002 was conducted at ETS with oversight from Mary Daane and Cathy Shaughnessy. Development and scoring activities associated with the main 2002 reading assessment were overseen by Patricia Donahue, with assistance from Robert Finnegan.

The complex statistical and psychometric activities necessary to report results for the NAEP 2002 oral reading study were directed by Andreas Oranje. John Donoghue provided invaluable assistance and advice with designing the study and preparatory psychometric work. The statistical and psychometric activities associated with the main 2002 reading assessment were directed by Catherine McClellan and Jinming Zhang, with assistance from Andreas Oranje, Hue Deng, Kelvin Gregory, and Ying Jin.

The extensive data processing and computer programming activities underlying the statistical and psychometric analyses conducted at ETS are under the direction of David Freund, Edward Kulick, Bruce Kaplan, and Steven Isham. Data analyses presented in this report were managed and conducted by Satwinder Thind and Laura Jerry.

The design and production of this report were overseen by Loretta Casalaina with assistance from Rick Hasney, Susan Mills, and Joseph Kolodey. Quality control and review procedures in the production of this report were supervised by Wendy Grigg. Ming Kuang coordinated the documentation and data checking procedures with assistance from Janice Goodis and Alice Kass. Arlene Weiner coordinated the editorial and proofreading procedures with assistance from Linda Myers, Trish Hamill, Jennifer O'Bryan, and Karen Damiano. The consistency and conventions review process was coordinated by Carmen Payton. The production of the web version of this report was coordinated by Rick Hasney with assistance from Loretta Casalaina.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS.

Table of Contents

List of Tables

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 1. Introduction

For more than 30 years, the National Assessment of Educational Progress (NAEP) has regularly collected, analyzed, and reported valid and reliable information about what American students know and can do in a variety of subject areas, including reading. As authorized by the U.S. Congress, NAEP assesses representative national samples of students at grades 4, 8, and 12. NAEP is administered and overseen by the National Center for Education Statistics (NCES), which is one of three centers within the U.S. Department of Education's Institute of Education Sciences. The content of all NAEP assessments is determined by subject-area frameworks that are developed by the National Assessment Governing Board (NAGB) in a comprehensive process involving a broad spectrum of interested parties, including teachers, curriculum specialists, subject-matter specialists, school administrators, parents, and members of the general public.

In 2002, NAEP conducted a national assessment of fourth-, eighth-, and twelfth-grade students' reading, the results of which were published in *The Nation's Report Card: Reading 2002* (Grigg et al. 2003). The assessment measured reading comprehension by having students read stories, articles, and other types of texts and answer questions about them. A special study, conducted in early spring of 2002, involved a subsample of fourth-graders who participated in the main reading assessment, and focused on a relevant, but sometimes overlooked, aspect of reading development—oral reading ability.

This special reading study was undertaken to discover what listening to fourth-graders reading aloud might teach us about students' oral reading ability and its components—accuracy, rate, and fluency. Fourth-graders were electronically recorded as they read aloud a 198-word excerpt from "The Box in the Barn" (Conner 1988), a reading selection from the main NAEP 2002 reading assessment, which had been administered one week before the oral reading study. All of the participants in the special study had received booklets containing "The Box in the Barn" when they participated in the main reading assessment.

The study reported here is a follow-up to *Listening to Children Read Aloud: Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4,* a 1995 report by Pinnell et al. based on NAEP data collected in 1992. Both studies were commissioned by NAGB. The original study was NAEP's initial attempt at large-scale measurement of oral reading abilities and one of the first ever performed. The major findings of IRPR as reported by Pinnell et al. (1995) were:

- Fifty-five percent of the participants were rated as *fluent*, based on the fluency scale developed for the IRPR to describe aspects of oral reading performance.
- Thirteen percent were rated *fluent* at the highest level.
- A majority of the participants (57 percent) were at least 96 percent accurate in their oral reading of the passage used in the study.
- Students' reading accuracy was associated positively with their comprehension scores.
- Students were less likely to correct their errors when the errors did not change meaning.
- Sixty-one percent of students read at least 100 words per minute.
- On average, slower readers demonstrated lower reading ability as measured by overall reading performance on the main NAEP reading assessment.
- Accuracy and rate were found to be related to fluency—those readers who read fluently were, on average, at least 96 percent accurate. Readers who were rated as *fluent* read the passage at an average rate of at least 126 words per minute.
- Finally, there was a significant relationship between oral reading fluency and reading comprehension, as measured by overall reading performance on the main NAEP assessment.

NAEP's 2002 data collection on oral reading used much of the same methodology and approach to understanding and reporting oral reading as the 1992 study, but, because different reading passages and administration procedures were used, the results of the two studies are not comparable. However, since the students who participated in the 2002 oral reading study also participated in the main NAEP assessment of reading comprehension in 2002, it is possible to examine the relationship between students' oral reading abilities and their reading comprehension. In this and the previous report, differences in student performance are discussed only if they have been determined to be statistically significant at the .05 level, as determined by *t* tests in combination with false discovery rate procedures for multiple comparisons. Details about multiple comparisons can be found in appendix A.

The participants in this study were a subsample of 1,779 fourth-graders from the 140,000 fourth-graders who were sampled for the main NAEP reading assessment. For this subsample of students, NAEP collected two sets of related data: reading comprehension data derived from their performance on the main NAEP reading assessment, and performance data on their oral reading of the excerpt from "The Box in the Barn." With relatively few complex language structures, simple vocabulary, and a familiar topic, this story was one of the easiest passages in the fourth-grade 2002 main NAEP reading assessment.

Rationale

Listening to children read aloud and evaluating their answers to questions about reading material are the two ways teachers can assess students' reading ability. Examining students' answers to questions about texts they have read may offer useful insights into their reading ability and their ability to think about texts. However, listening to students' oral reading performance affords researchers opportunities to examine factors such as accuracy and phrasing, which are not directly observable through their answers to questions. By listening to children read aloud, it is possible to gain important clues about their reading development. Levels of automaticity (ease of word identification or unconscious decoding of words), accuracy, phrasing, apprehension of syntactical complexity, expressiveness, and even functional understanding of punctuation can be detected by carefully listening to and rating oral reading performance (Adams 1990). For these reasons, it is important to analyze and measure oral reading ability.

The approach to reading in the oral reading study is the same one that underpins the NAEP reading framework and that has guided NAEP assessment development for the last decade. This framework, developed through a comprehensive national process involving reading experts and other interested individuals, includes the contexts and purposes of reading as important to readers as they develop understanding of a text (NAGB 2002; National Institute for Child Health and Human Development [NICHHD] 2000). The framework is based on a view of reading as a dynamic, complex interaction among the reader, the text, and the context or situation. With this perspective, NAEP's 2002 reading assessment reflected contemporary understanding that reading abilities develop through an integration of multiple cognitive, affective, and social processes (Dole et al. 1991).

NAEP developed the oral reading study to provide a richer understanding of the literacy development of fourth-grade students within the constraints of a large-scale assessment. The oral reading study was designed to examine several important components of literacy development that are integrated in proficient reading—reading accuracy, reading rate, and reading fluency—and to compare these to overall reading comprehension as measured in the main NAEP reading assessment.

The oral reading study addresses several important assessment goals widely agreed upon in the community of reading educators and others concerned about reading development (Pinnell et al. 1995). The components of oral reading—accuracy, rate, and fluency—constitute a cluster of critical literacy abilities (Snow, Burns, and Griffen 1998). Oral reading functions as a significant indicator of overall reading ability (Fuchs et al. 2001). Educational researchers recognize reading aloud as central both to comprehensive and efficacious reading instruction and to reading assessment (Fuchs et al. 2001; Nathan and Stanovich 1991; NICHHD 2000, April 13, 2000; No Child Left Behind [NCLB] Act of 2001; Rasinski 2000; Reutzel and Hollingsworth 1993; Shinn et al. 1992). This connection between oral and overall reading abilities was noted by the evidence-based report of the National Reading Panel ([NRP] 2000) as summarized in NICHHD (2000, p. 6): "Fluent readers are able to read orally with speed, accuracy, and proper expression. Fluency is one of several critical factors necessary for reading comprehension." Furthermore, the NRP (2000, p. 3-3) found reading efficiency is increased by a range of instructional approaches that encourage repeated oral reading with teacher feedback, suggesting the efficacy of regular classroom assessment of oral reading fluency.

The importance of understanding the role of oral reading achievement is underscored by the emphasis placed on it by such national programs as NCLB 2001 and Reading First, the national initiative established by NCLB. Reading First, which is based on the compilation of evidence about effective diagnostic, classroom, and assessment methods by the NRP, is dedicated to helping all children become successful early readers (NCLB 2001; U.S. Department of Education 2003).

Measures of Oral Reading

The purpose of the study was to gather specific and measurable performance data about reading accuracy, rate, and fluency of fourth-grade readers in the nation. The scorers who participated in the rating and coding of students' oral reading performances were trained not to penalize variant pronunciations, such as those arising from regional, dialectical, or nonnative speech. All scorers were directed to listen to an entire oral rendering, to become familiar with the student's unique speech patterns, before beginning accuracy or fluency coding. Scorers were encouraged to determine whether or not the student knew and correctly read a particular word, given his or her own unique speech characteristics.

Accuracy, in the context of this study, refers to the degree to which a student's oral reading conforms to the letter-sound conventions of printed, standard English; in other words, accuracy measures the student's precision in orally presenting the words in the text. Accuracy results are presented in chapter 2.

Rate refers to the speed at which the student reads aloud. In this study, rate is measured and recorded as both the number of words per minute for the entire performance and the number of words in the initial minute of oral reading. The assessment results of accuracy and rate are presented in chapter 3 of this report.

Fluency, as defined in this study, is a rating of the student's ability to render an appropriately phrased and syntactically coherent delivery of the passage. For this study, the four-level, holistic NAEP Oral Fluency Scale was developed. The lower half of the scale, levels 1 and 2, designates two degrees of nonfluent performance, characterized by word-by-word readings with awkward word groupings and sometimes nonmeaningful syntax. The upper half of the scale, levels 3 and 4, designates two degrees of fluent performance, characterized by preservation of the author's syntax, appropriate phrase groupings, and sometimes expressive interpretation. Chapter 4 presents the results on the NAEP Oral Fluency Scale.

Administration of the Oral Reading Study

Sample

Interviews for the oral reading study were conducted with a nationally representative subsample of the fourth-graders who participated in the main NAEP 2002 reading assessment. All students who participated in the oral reading study had been given the fourth-grade reading passage, "The Box in the Barn," and 12 comprehension questions based on that passage, one of the easiest fourth-grade passages from the main NAEP reading assessment, which was conducted one week prior to the oral portion of the assessment.

The data in this study were collected from a subsample (1,779) of the sample (140,000) of fourth-graders who participated in the NAEP reading assessment during the early spring of 2002. The target student sample size for the oral reading sample was a fraction of the fourth graders targeted for the main NAEP reading study. Even though considerably fewer than 60 students were selected from each school for the oral reading study, further school subsampling was required. For efficiency, the nationally subsampled schools were grouped into 148 geographic clusters, each containing at least five eligible sampled schools. (A cluster could be an individual county, if it met the minimum size requirement, or two or more adjacent counties.) From the 567 counties with at least one eligible grade 4 school, 148 geographic clusters were defined, and 91 were selected with probability proportional to the number of eligible schools. Five of the 91 were selected with certainty because each contained a large number of schools. In each of the remaining 86 sampled clusters, five schools were selected with equal probability. In the five certainty clusters, schools were also subsampled with equal probability, at a rate equal to the product of the cluster probability and the within-cluster probability for noncertainty clusters. Details about the sample design of the main NAEP reading assessment can be found in appendix A.

The 1,779 fourth-graders who participated in the oral reading study represented 84 percent of the students who had been identified to participate in the main NAEP reading with the oral reading study as one of their subtasks. Only three race/ethnicity groups (White, Black, and Hispanic) were represented in the oral reading study in great enough numbers to report results for these students. Of the 16 percent of students who did not participate, 6 percent were students with disabilities (SD) or limited-English-proficient (LEP) students who would not have been able to participate meaningfully in the study. The reasons for nonparticipation of the remaining 10 percent of students cannot be determined. Guardians of the students or the students themselves may refuse participation for any reason. This sample represented 79 percent of identified schools, reflecting both the nonresponse in main NAEP reading and the nonresponse in the oral reading study. The overall response was 66 percent.

A nonresponse bias analysis was conducted to study differential nonresponse across subgroups. This analysis showed that almost all variables, such as gender, race/ethnicity, school affiliation, free/reduced-price school lunch, Title I classifications, and SD/LEP classifications showed some level of differential nonresponse. Details about participation rates and correction for nonresponse bias can be found in appendix A.

Procedure

The 2002 oral reading study involved the computer-assisted collection of digital recordings of students reading aloud. This procedure offered an important technological advance compared with the 1992 procedures. In both 1992 and 2002, data were collected in individual, one-on-one assessment sessions between a student and a trained interviewer. Each collected recordings of students reading aloud. The 2002 administration, however, was enhanced by the use of computer-assisted collection and digital recording.

Each individual session consisted of an introduction by the administrator and an oral reading demonstration by the student. The sessions were digitally recorded and later evaluated by trained scorers. In contrast to the conditions in the 1992 study, the students who participated in the 2002 study were not subject to the distraction of having the administrator taking notes on their performance, providing for a more natural, relaxed environment in which to conduct the assessments. In addition, these digital renderings of students' performance were of much higher quality than the taped recordings used in the earlier study, which greatly improved the ease and efficiency of the rating process.

Introduction of students to the study and initial screening

During the introduction, the student was welcomed and made familiar with the session protocol. The student was then asked to read aloud a brief passage from a story at approximately a second-grade reading level. The intention here was to familiarize students with the digital recording process and to enable the interviewer to determine whether the student should be asked to read aloud from the more difficult assessment passage later in the interview. Students who seemed to have considerable trouble reading aloud from the less difficult passage were thanked for reading the second-grade screening passage and excluded from further participation. If the administrator decided that the student read the excerpt from "The Box in the Barn" with much difficulty (i.e., the student produced a word-by-word delivery, halted between words, was unable to recognize many of the words, took more than 6 minutes to read the passage, or reached a point of frustration and gave up), the session was stopped. However, less than 1 percent of the students selected for participation was excluded for these reasons.

The results reported here reflect data from fourth-graders across the spectrum of abilities, from low- to high-level reading performance. NAEP assessments are subject to nonparticipation when sampled students or their guardians elect to withdraw a student from participation. Federal law mandates complete privacy for all student participants, their families, and schools; therefore, results for individual students or schools are not available for this or any NAEP study.

Preparation for oral reading

"The Box in the Barn," the story from which the oral reading passage was excerpted, is one of the easiest passages on the fourth-grade NAEP reading assessment. The participants in this study had an opportunity to read "The Box in the Barn" during the main assessment. After the screening process for the oral reading study, students were given another opportunity to read the entire story silently. In order to encourage engagement with the content of the story, prior to reading the story silently students were told that they would be asked a few questions about it. When the students indicated they had completed the silent reading task, the supervisor handed them, one at a time, printed copies of three comprehension questions. (The questions can be found in appendix A under "Data Collection.") The process of preparing students for participation in the oral reading study took 30 minutes or less.

This opportunity to think about and discuss the content of the story is an important part of the context in which students were subsequently asked to read a 198-word excerpt of the story out loud. The answers to the three orally presented comprehension questions were not scored or evaluated; rather, they were used as a way to refamiliarize the students with the passage they had encountered in the main NAEP reading assessment, and to give them an opportunity to develop an interest in the passage and the task at hand. Students could read the printed questions while they listened on a headset to a recording of each question as it was read aloud. Students were given time to think about their answers, but if no answers seemed forthcoming, the interviewer could prompt for a response. If a student missed the recorded instructions, the interviewer was permitted to replay them. Students who could not answer the questions were given permission to go back to the story for clarification.

Oral reading demonstrations

Finally, the supervisor showed the students a 198-word passage taken from the beginning of "The Box in the Barn" and asked them to read this out loud. If students asked about the desired reading rate, they were told to read as if they were reading to someone who had never heard the story before. If the students demonstrated difficulty with the task, the interviewer was instructed to say, "Just do the best you can." If students asked for help with pronunciation during the reading, the interviewer was instructed to say, "If you can't figure out a word, you can guess or skip it, and go on." All students who participated in the oral reading demonstration were able to read the entire excerpt aloud.

The results of students' oral reading presented in this report should be interpreted with the understanding that students had an opportunity to become familiar with the story before being asked to read it aloud. Furthermore, the fact that only one passage was used in this study may limit the generalizability of the study's results to other passages. A copy of the story and the 198-word excerpt that students read out loud, and that served as the basis for analyses of students' oral reading ability, are presented in appendix B.

Reporting the Assessment Results

Chapter 2 of this report presents the results of analyses of fourth-graders' oral reading accuracy, including data on the nature of students' errors and self-corrections, and on the relationship between accuracy and the students' performance on the main NAEP reading assessment. Chapter 3 presents results of analyses of students' reading rates, as measured in words per minute across the entire oral reading study excerpt and as words read during the first minute of reading. Reading rates are also discussed in light of average scale scores. Chapter 4 presents measures of fourth-graders' oral reading fluency, including data on the relationships between fluency and accuracy, fluency and rate, and fluency and comprehension (tabulated as average scale scores and NAEP reading achievement levels). Finally, the impact of both accuracy and rate on fluency is discussed.

The average scale score results from the main reading assessment are based on the NAEP 2002 reading scale, which ranges from 0 to 500. In order to calculate students' average scores on the NAEP reading assessment, the analysis begins by determining the percentage of students responding correctly to each multiple-choice question and the percentage of students responding at each score level for each constructed-response question. The analysis entails summarizing the results on separate subscales for each reading context (reading for literary experience and reading for information) and then combining the separate scales to form a single composite reading scale.

Figure 1-1. Policy definitions of the three NAEP achievement levels

Basic (208)	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
Proficient (238)	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
Advanced (268)	This level signifies superior performance.

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

In addition to scale scores, NAEP also reports students' reading performance in terms of achievement levels. These achievement levels are authorized by the NAEP legislation and adopted by NAGB. For each grade assessed in main NAEP—4, 8, and 12—NAGB has adopted three achievement levels: *Basic, Proficient,* and *Advanced.* Figure 1-1 shows NAGB's policy definitions of the three NAEP achievement levels that apply across grades and subject areas. The policy definitions guided the development of the reading achievement levels, as well as the achievement levels established in all other subject areas. Adopting three levels of achievement for each grade signals the importance of looking at more than one standard of performance. The Board believes, however, that the overall achievement goal for students is performance that qualifies at the *Proficient* level or higher as measured by NAEP. The *Basic* level is not the desired goal, but, rather, represents partial mastery that is a step toward *Proficient.* More extensive descriptions of the reading achievement levels are available on the NAGB website (http://nces.ed.gov/nationsreportcard/achlevdev.asp?id=rd).

For reporting purposes, achievement-level cut scores are placed on the reading scale, resulting in four ranges: below *Basic, Basic, Proficient,* and *Advanced*. In chapters 2, 3, and 4, some of the oral reading results for overall accuracy, rate, and fluency are illustrated along with corresponding average scale scores and achievement levels.

NAEP reports overall results for the national sample and for a number of subgroups. This report presents results for two of these categories: gender and race/ethnicity. Data on both were derived from school records. NAEP race/ethnicity group classifications reflect the following mutually exclusive categories: White, Black, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native. Based on statistically determined criteria, at least 62 students in a particular group must participate in order for the results for that group to be considered reliable enough to be reported by NAEP (Johnson and Rust 1992; Allen, Donoghue and Schoeps 2001). For the oral reading portion of the 2002 reading assessment, the race/ethnicity subsample was insufficient to permit reporting of results for Asian/Pacific Islander and American Indian/Alaska Native students.

Achievement Levels

Achievement-level results are performance standards set by the National Assessment Governing Board (NAGB), based on recommendations from panels of educators and members of the public, to provide a context for interpreting student performance on NAEP. These performance standards are used to report what students should know and be able to do at the *Basic, Proficient,* and *Advanced* levels of performance in each subject area and at each grade assessed. For reporting purposes, cut scores are placed on the reading scale, resulting in four achievement-level ranges: below *Basic, Basic, Proficient,* and *Advanced*. The achievement-level results are then reported as percentages of students within each of the four ranges, as well as percentages of students at or above *Basic* and at or above *Proficient*. As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that achievement levels are to be used on a trial basis and should be interpreted with caution. However, NCES and NAGB have affirmed the usefulness of these performance standards for understanding trends in achievement. NAEP achievement levels have been widely used by national and state officials.

Drawing Inferences From NAEP Assessment Results and From the Oral Reading Study

The average scores and percentages presented in this report are estimates based on samples of students rather than on entire populations. Moreover, the collection of questions and the single passage used for the assessment are just samples of the many questions and passages that could have been chosen to assess the skills and abilities described in the NAEP reading framework. Therefore, the results offered are subject to a measure of uncertainty, reflected in the standard error of the estimates—a range of a few points plus or minus the score or percentage—which accounts for potential score or percentage fluctuation due to sampling and measurement error. The estimated standard errors for the estimated scale scores and percentages presented in chapters 2 and 3 are provided in appendix C. All results discussed herein were found to be significantly different at the .05 level.

Readers are cautioned against interpreting NAEP results in a causal sense. Inferences related to the performance of reporting groups defined by gender or race/ethnicity, for example, should take into consideration the many socioeconomic and educational factors that may also impact reading performance.

Additional NAEP data are available in the NAEP Data Tool (http://nces.ed.gov/nationsreportcard/naepdata/) and in restricted-access research databases. Researchers and policy analysts are free to make use of the data (subject to various confidentiality restrictions) as they wish.

Overview of the Report

The oral reading study provides important data for educators and researchers about fourth-graders' oral reading proficiency and how this relates to their overall reading abilities and comprehension. Chapter 2 of this report focuses on one major aspect of oral reading—accuracy—and its relationship with average scale scores. In addition, this chapter examines students' deviations from the text that change meaning and the patterns of self-correction in students' oral reading. Chapter 3 explores students' reading rates (both as words per minute across the entire excerpt and as words read during the first minute of reading) and the relationship between rate and average scale score. Chapter 4 discusses the NAEP fluency scale—that is, how fluency was defined and measured in the oral reading study—and the results of the study. In addition, chapter 4 looks at the interrelationship of fluency, accuracy, and rate and how they are related to students' overall reading performance on the main NAEP 2002 reading assessment. Chapter 5 summarizes overall and subgroup findings of chapters 2, 3, and 4.

The appendixes of this report contain information in addition to the results presented in chapters 2, 3 and 4. Appendix A contains an overview of the development, sampling, administration, coding, data analysis, and significance-testing procedures for the oral reading study. Appendix B presents the text of "The Box in the Barn," the story from which students read aloud for this study. Appendix C contains the data and corresponding standard errors from which the figures presented in this report were drawn.

Chapter 2. Oral Reading Accuracy

Oral reading accuracy figures prominently in discussions of reading development. Readers who have difficulty recognizing and pronouncing individual words often have serious problems deriving meaning from the texts they encounter (Harris and Hodges 1995). When readers pause to decode, they may lose track of ideas unfolding in the sentence or surrounding portion of the text. A child who cannot distinguish "cat" from "cut," or who lingers to differentiate "bag " from "gab," or who does not recognize that "brook" is not "brick," is unlikely to derive much meaning from even the most basic textbooks or stories. While experts probably agree that accuracy should not be the sole purpose of reading instruction, it is widely accepted that higher-level reading skills cannot be developed without a foundation of accurate and efficient word recognition (National Reading Panel [NRP] 2000). Several researchers stress the relationship between fluency and sight word identification (Schwanenflugel et al. 2004; Torgensen, Rashotte, and Alexander 2001).

In the decade since NAEP's first study of fourth-graders' oral reading ability, researchers, educators, and theorists have continued to investigate what constitutes adequate levels of accuracy in reading. Many assessment methods developed for informal or on-the-spot classroom situations use the percentage of words read correctly as a measure of accuracy (Barrantine 1995; Davenport and Lauritzen 2002; Lipson and Wixson 1991).

However, as indicated by alternative expert opinions, defining, quantifying, and measuring accuracy can be a complex endeavor. For example, some informal reading inventories suggest that reading with less than 95 percent word accuracy is insufficient for full comprehension of the text (Harris and Sipay 1985). Kuhn and Stahl (2003) indicate that fluency is supported when students read materials at the independent decoding level (95 percent accuracy) or at the instructional decoding level (90 percent accuracy). Rayner et al. (2001) discuss the importance of word-reading skills for both accuracy and fluency. Other reading experts propose that not all errors are necessarily disruptive to the individual reader's process or comprehension. Instead, errors in reading words may be caused by prior knowledge and by experiences and cognitive processing that the reader draws on during the act of interpreting the words on the page. These experts encourage reading teachers to explore the nature of a student's oral reading errors before determining that student's ability to read and comprehend (Bloome and Dail 1997; Clay 1993; Leslie and Taft 1985).

Overall Accuracy Results

To characterize the accuracy with which students read the passage out loud, three specific types of errors were captured in the coding: substitutions, omissions, and insertions. Only whole-word omissions were coded as omissions (e.g., "he wished" for "he secretly wished"). Partial-word omissions were coded as substitutions (e.g., "in" for "inside"). If contiguous words were omitted, each word in the set was counted as an error (e.g., "Dad wouldn't be back for two hours" instead of "Dad wouldn't be back for at least two hours" would count as two omissions). Only whole-word insertions were coded as insertions (e.g., "wanting his very own puppy" for "wanting his own puppy"). Inserts of prefixes or suffixes were coded as substitutions (e.g., "return" for "turn" or "unhappy" for "happy"). Coders received training to help them avoid penalizing regional, dialectical, and nonnative pronunciations as phonemic errors (see appendix A for a more detailed description of the coding procedures that were used in this study).

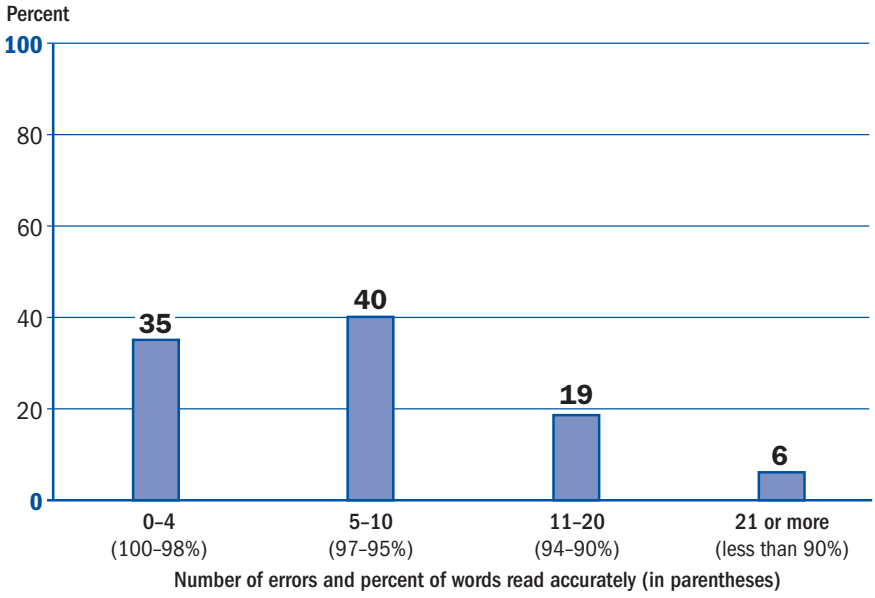
The percentage of students reading with different degrees of accuracy is displayed in figure 2-1. Students who read with 4 or fewer errors were reading with at least 98 percent accuracy, students with 5 to 10 errors were reading with approximately 95 to 97 percent accuracy, and so forth. As shown in the figure, approximately three-quarters of students read the excerpt from " The Box in the Barn" (one of the easiest fourth-grade passages from the main NAEP reading assessment), with at least 95 percent accuracy, and approximately one-third (35 percent) read the excerpt with at least 98 percent accuracy. Six percent of the nation's fourth-graders read this passage with less than 90 percent accuracy.

The results of the 2002 oral reading and the 1992 IRPR studies are not directly comparable because different oral reading passages, administration procedures, and results categories were used. Nevertheless, the two studies show roughly parallel findings. In the 1992 IRPR study, more than one-half (approximately 57 percent) of fourth-graders read a sample passage with at least 96 percent accuracy. Further, less than one-quarter (23 percent) of the fourth-graders read with less than 94 percent accuracy.

The relationship between students' oral reading accuracy and their reading comprehension as measured on the NAEP reading assessment is displayed in figure 2-2. Significance testing revealed a statistically significant difference between the average scores of students in each accuracy category, indicating a negative relationship between oral reading errors and reading comprehension (see appendix A for details on significance test procedures). Thus, the greater the number of oral reading errors made by students in the study, the lower their average scores on the comprehension assessment. These findings parallel the conclusions of the NRP (2000) that reading comprehension depends in part upon a base of accurate word recognition.

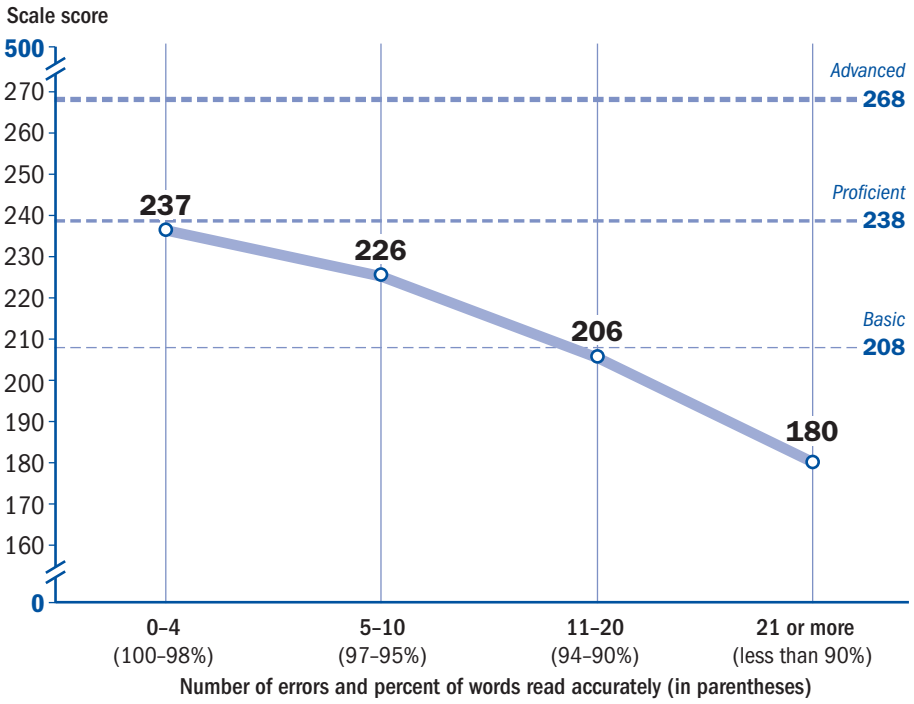
Another way to examine the relationship between accuracy and comprehension is to study how students with different degrees of accuracy performed on the NAEP reading test in relation to the achievement levels, *Basic*, *Proficient*, and *Advanced*. The average score for students reading with at least 98 percent accuracy (237) was not significantly different from the *Proficient* achievement-level cut score (238). The average score for students reading with between 95 and 97 percent accuracy (226) was within the *Basic* achievement level. Those students who read with between 90 and 94 percent accuracy had an average score (206) that was not significantly different from the *Basic* achievement-level cut score, and those who read with less than 90 percent accuracy had an average score (180) that fell 28 points below the *Basic* achievement-level cut score (208).

Figure 2-1. Percentage of students, by degree of reading accuracy, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

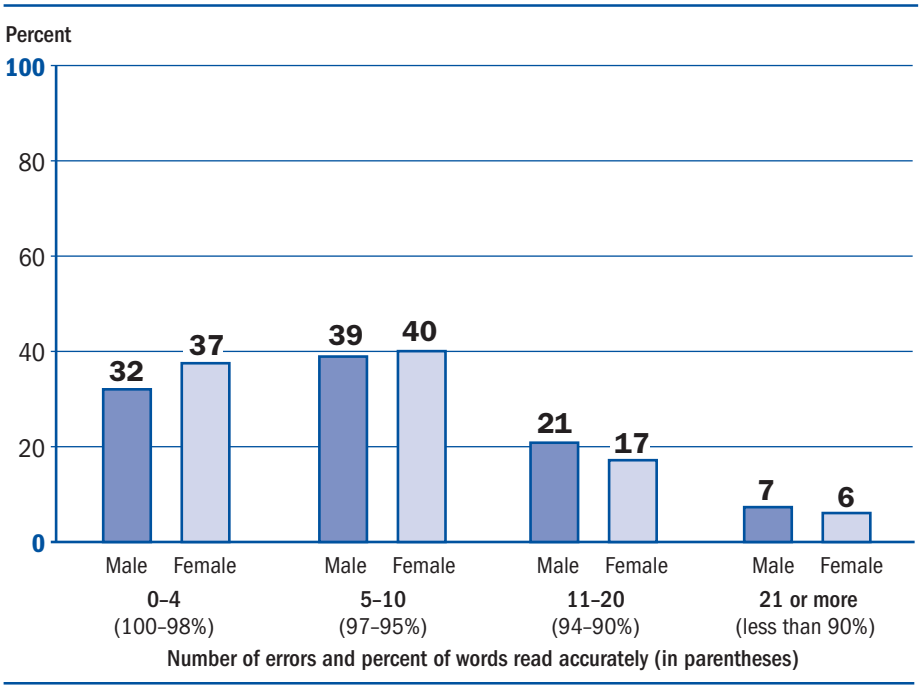
Figure 2-2. Average reading scale scores in relation to the achievement levels, by degree of reading accuracy, grade 4: 2002



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

Figures 2-3 and 2-4 present accuracy results by gender and racial/ethnic subgroups. Differences in students' performance on the 2002 reading assessment between demographic subgroups are discussed only if they have been determined to be statistically significant. The reader should bear in mind that the estimated average scale score for a subgroup of students does not reflect the entire range of performance within that group. Differences in subgroup performance cannot be ascribed solely to students' subgroup identification. Average student performance is affected by the interaction of a common set of educational, cultural, and social factors not discussed in this report or addressed by NAEP assessments. Note that only three racial/ethnic subgroups (White, Black, and Hispanic) were represented in the oral reading study in great enough numbers to report results for these students.

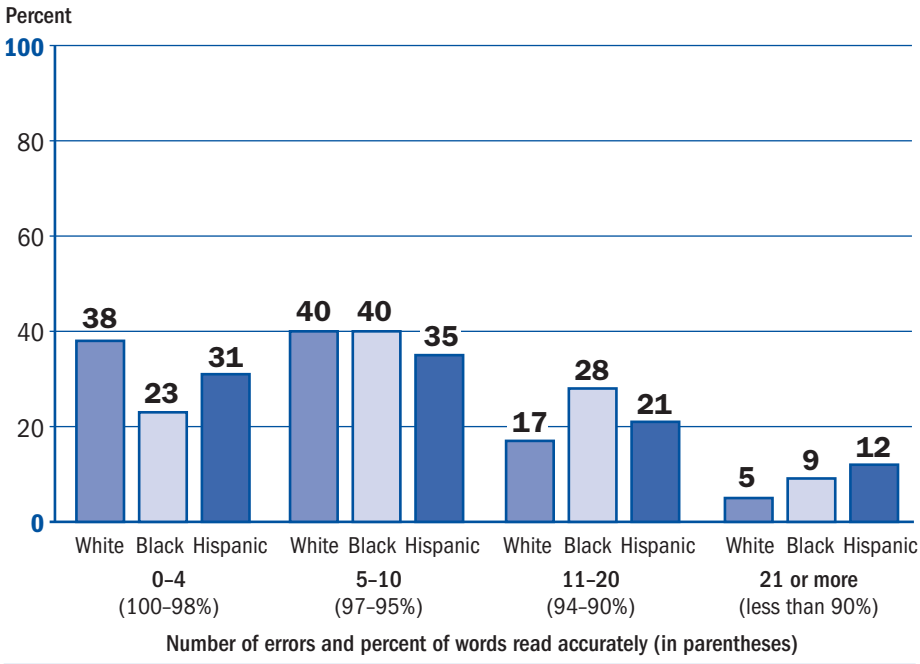
Figure 2-3. Percentage of students, by gender and degree of reading accuracy, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Consistent with the average score results from the main reading assessment reported in the 2002 reading report card (Grigg et al. 2003), female fourth-graders were generally more accurate in their oral reading than their male counterparts. Compared with 37 percent of female students reading with at least 98 percent accuracy, 32 percent of male students read with that level of accuracy. A higher percentage of White students (38 percent) than Black students (23 percent) read with at least 98 percent accuracy, but the apparent difference between White students and Hispanic students (31 percent) who read with at least 98 percent accuracy was not found to be statistically significant.

Figure 2-4. Percentage of students, by race/ethnicity and degree of reading accuracy, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Meaning-Change and Non-Meaning-Change Errors

When reading aloud from "The Box in the Barn," students sometimes made substitutions such as "noise" for "cries" which reflects the student's understanding of context. Other substitutions such as "waiting" for "wanting" may result from the student's misreading of the text. All errors may not be equally disruptive to a reader's attempt to understand a passage. An error may or may not result in a change of meaning to the text. For example, substituting the word "pony" for "horse" would probably be less disruptive to the text's overall meaning than substituting the word "house" for "horse." The errors made by students in the oral reading study were evaluated for their potential to disrupt students' understanding of the passage. Substitution of "house" for "horse" would have been classified as a meaning-change error, while "pony" for "horse" would have been classified as a non-meaning-change error.

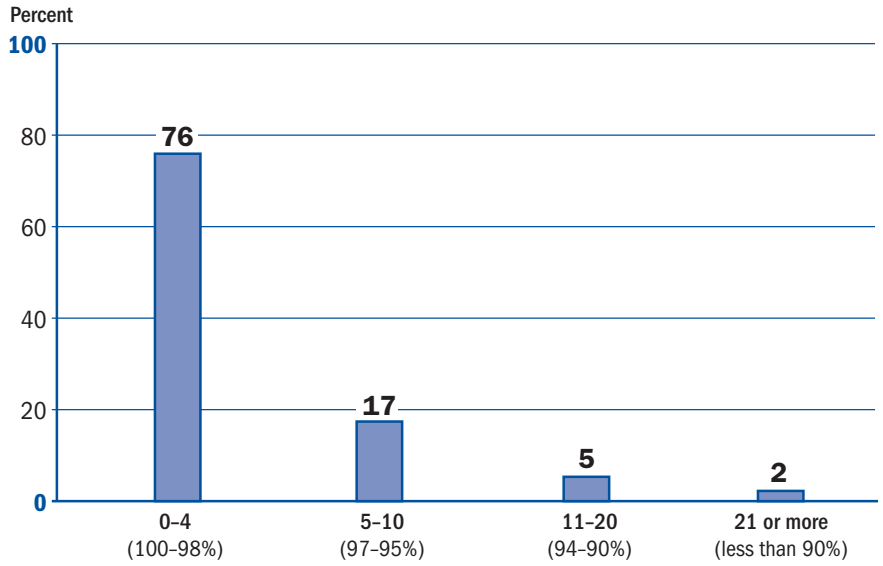
Figure 2-5 presents oral reading accuracy results based on meaning-change errors only. Here, those errors that were not considered to have caused a change in the meaning of the text were left out of the calculation. As a result, a greater percentage of students in this calculation appear to be reading with greater accuracy. Approximately three-fourths (76 percent) read the passage with four or fewer meaning-change errors—that is, they read at least 98 percent of the words either exactly as in the printed text or with only non-meaning-change errors. Just under one-fifth of the students (17 percent) read with between 5 and 10 meaning-change errors (95–97 percent accuracy). Finally, approximately 7 percent of the participants read with 11 or more meaning-change errors (5 percent of the students read with 90–94 percent accuracy, and 2 percent of the students read with less than 90 percent accuracy).

As mentioned in the section on accuracy, the results of the 2002 oral reading study are not directly comparable to the results of the 1992 IRPR. However, this study's findings on meaning-change errors are reflective of those of the earlier study. The IRPR found more than three-quarters (approximately 78 percent) of the fourth-graders read with 97 percent or higher accuracy when only meaning-change errors were considered, less than one-fifth (approximately 19 percent) of the students read with at least 94 percent accuracy when only meaning-change errors were considered, and 2 percent of the students read with less than 94 percent accuracy when only meaning-change errors were considered.

The relationship between meaning-change errors and comprehension is displayed in figure 2-6. Note that the number of students who read with less than 90 percent accuracy when only meaning-change errors were counted was too small to permit a reliable estimate of their average scores. Similar to the relationship between all types of reading errors (regardless of how they may have disrupted the text's meaning) discussed in the previous section, the number of meaning-change errors made by students reading aloud showed a significant negative relationship with reading comprehension. On average, the greater the number of meaning-change errors made by students, the lower their comprehension scores. The average score for students with between 11 and 20 meaning-change errors was 180, the average score for students with between 5 and 10 meaning-change errors was 206, and the average score for students with 4 or fewer errors was 231. When accuracy is figured based on all errors, the average score for students with 21 or more errors was 180, the average score for students with between 11 and 20 errors was 206, the average score for students with between 5 and 10 errors was 226, and the average score for students with 4 or fewer errors was 237. This finding is not surprising, given the number of empirical studies showing a relationship between comprehension and oral reading errors as documented earlier (Calfee and Piaotkowski 1981; Herman 1985; Snow, Burns, and Griffin 1998; Stanovich 1986).

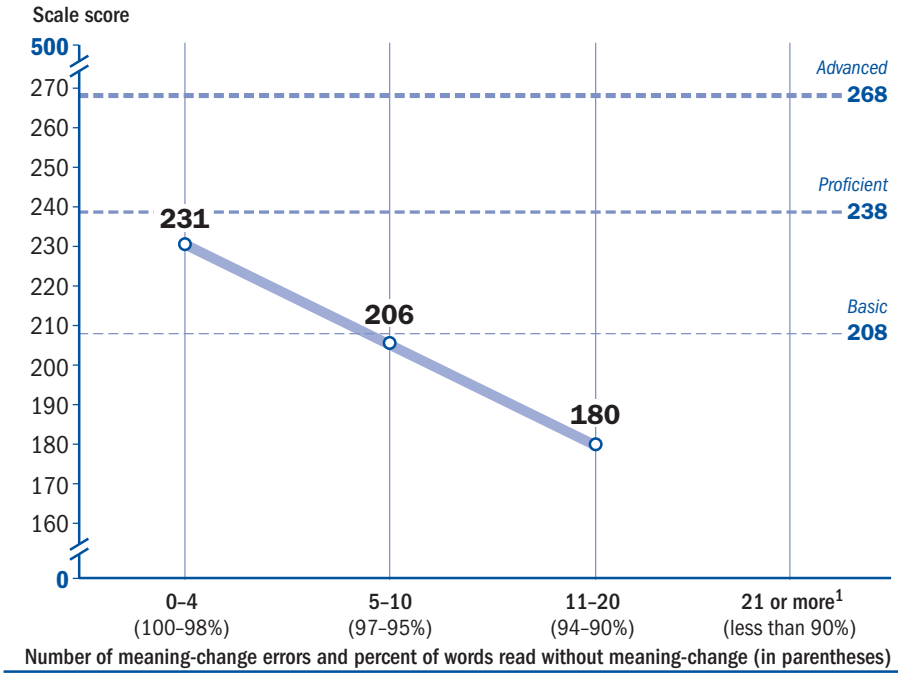
When only meaning-change errors were counted, the average score for students reading with at least 98 percent accuracy (231) fell within the *Basic* achievement level. The average score for students reading with between 95 and 97 percent accuracy (206) was not significantly different from the *Basic* achievement-level cut score (208). Those students who read with between 90 and 94 percent accuracy had an average score (180) that fell 28 points below the *Basic* achievement-level cut score (208).

Figure 2-5. Percentage of students, by degree of reading accuracy when only counting meaning-change errors, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 2-6. Average reading scale scores in relation to the achievement levels, by degree of reading accuracy when only counting meaning-change errors, grade 4: 2002



[1] Sample size was insufficient to permit a reliable estimate for students with 21 or more errors that resulted in a change of meaning.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

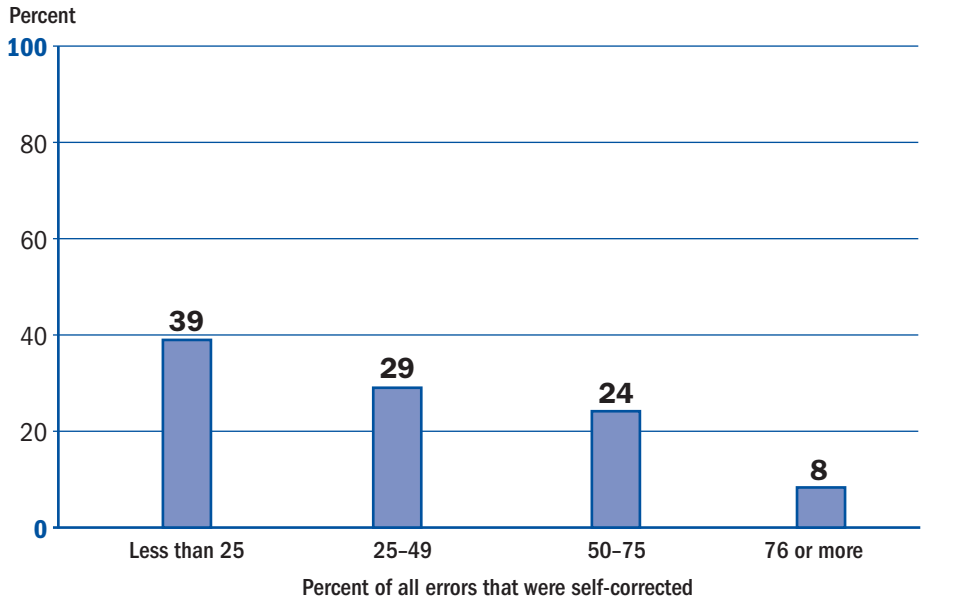
Self-Corrections

When readers recognize that they have made an error in reading a word in the text, they may choose to correct their mistake or move on and continue reading. Making such decisions is part of being a strategic reader. If readers are closely monitoring their understanding of the text, they may recognize that a particular error disrupts the meaning of the text, and may pause to correct that mistake before moving on. With less disruptive mistakes (for example, those that may not change the meaning of the text), readers may choose to continue reading without pausing, since they may not sense any disruption in the coherency or meaning of the text. The National Reading Panel Report (NRP) says of this phenomenon, "Fluency represents a level of expertise beyond word recognition accuracy, and reading comprehension may be aided by fluency" (NRP 2000).

The oral reading of fourth-graders in the oral reading study was evaluated for evidence of self-correction. A self-correction was identified when any substitution, omission, or insertion was subsequently corrected, so that the student's final production was an exact match with the text. Figures 2-7a, 2-7b, and 2-7c present the proportion of self-corrected errors observed in students' oral reading. These data are examined in three ways: the proportion of all errors that were self-corrected, the proportion of just meaning-change errors that were self-corrected, and the proportion of just non-meaning-change errors that were self-corrected.

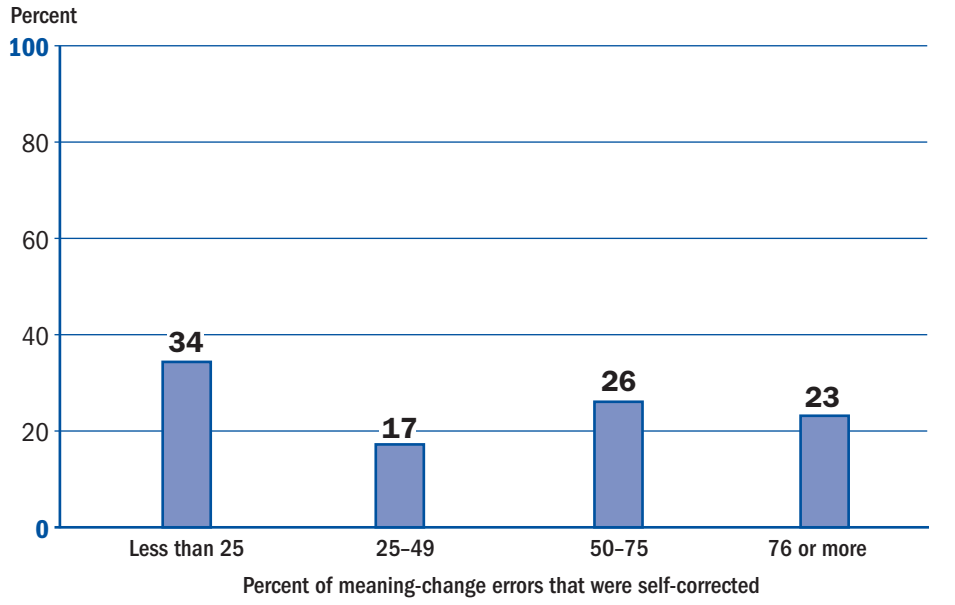
As shown in figure 2-7a, the majority of students (approximately 68 percent) self-corrected less than half of their oral reading errors. Only 8 percent of students self-corrected more than three-fourths of the errors they made. The proportion of self-corrected errors appears to be somewhat dependent on whether or not the errors resulted in a meaning change. Figure 2-7b shows that nearly one-half (approximately 49 percent) of the students corrected 50 percent or more of the errors they made that had caused a change in meaning. The picture is different in figure 2-7c, which shows that only about one-quarter (approximately 26 percent) of the students corrected 50 percent or more of the errors they made that did not cause a change in meaning.

Figure 2-7a. Percentage of students, by percentage of all errors that were self-corrected, grade 4: 2002



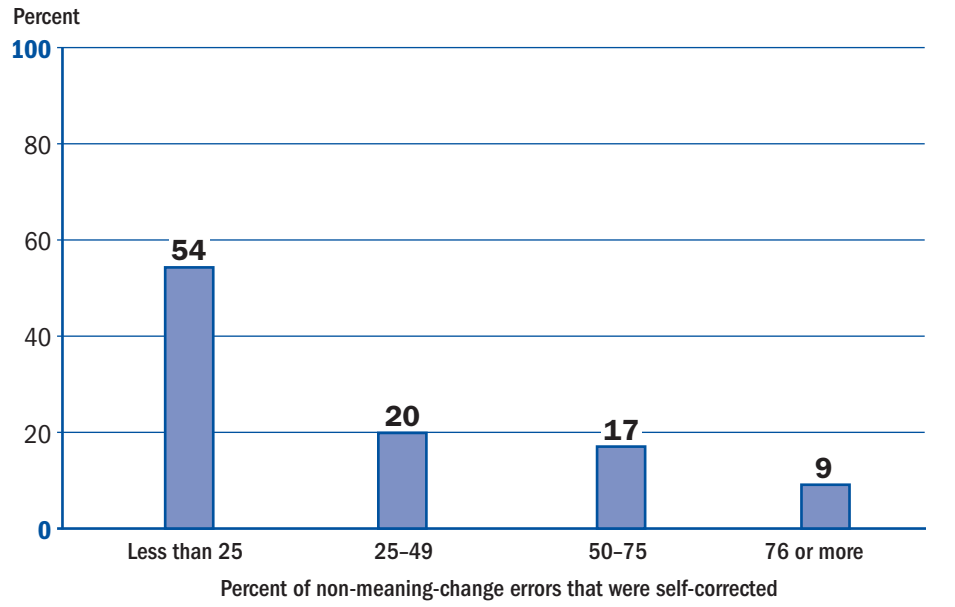
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 2-7b. Percentage of students, by percentage of meaning-change errors that were self-corrected, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 2-7c. Percentage of students, by percentage of non-meaning-change errors that were self-corrected, grade 4: 2002



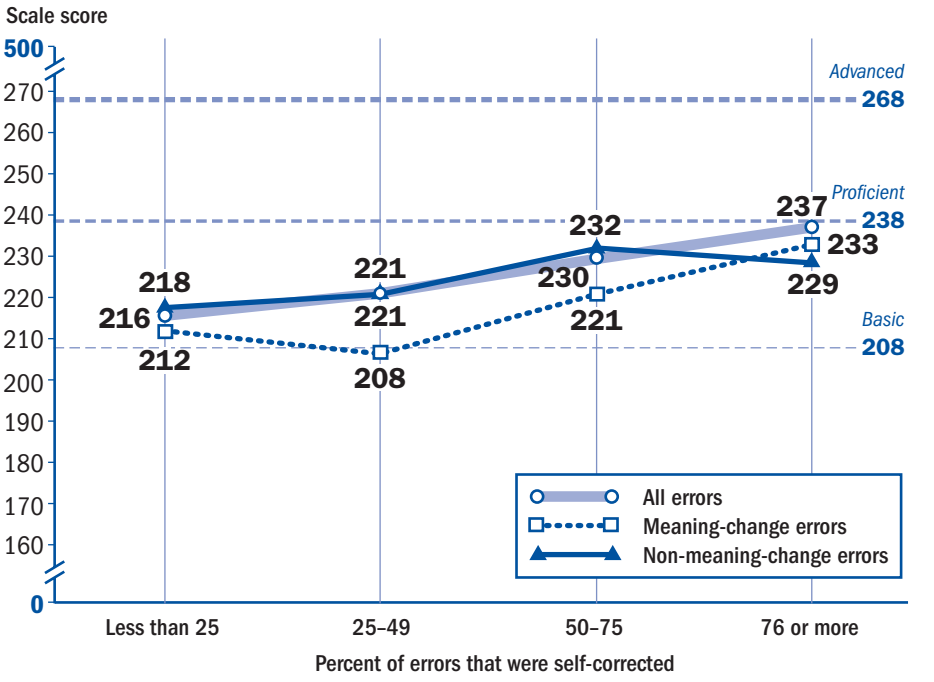
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

The relationship between the proportion of errors that students self-corrected and their average reading scores is displayed in figure 2-8. The figure shows relationships between the proportion of self-correction of each of the categories of error—all errors, meaning-change errors, and non-meaning-change errors—and average scale scores. Generally, the relationship between the proportion of all self-corrected errors and average reading scores is positive—the greater the proportion of self-corrected errors (counting all error types), the higher the average score. Students who corrected more than 75 percent of all their errors had an average score (237) that was not significantly different from the *Proficient* achievement-level cut score (238), and students who corrected 50 to 75 percent of their errors scored within the *Basic* achievement level.

The relationship between self-correction of meaning-change errors and average score is similar. Students who self-corrected more than 75 percent of meaning-change errors had an average score that was higher than students who self-corrected a smaller proportion of meaning-change errors. These results agree with other research studies showing positive relationships between comprehension and phonemic awareness (Griffith and Olson 1992; Stanovich 1993; Yopp 1992).

Students who self-corrected 50 to 75 percent of non-meaning-change errors had an average score that was higher than students who self-corrected a smaller proportion of non-meaning-change errors. However, no statistically significant difference was detected between the average score of students who corrected 50 to 75 percent of non-meaning-change errors and that of students who corrected more than 75 percent of non-meaning-change errors.

Figure 2-8. Average reading scale scores in relation to the achievement levels, by the type and percentage of errors that were self-corrected, grade 4: 2002



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

Chapter 3. Oral Reading Rate

Maintaining a fairly rapid and moderately steady pace of oral reading can be essential to the process of making connections between ideas in a text. If reading proceeds too slowly or mechanically, these connections may become difficult or impossible to make (Crawford, Stieber, and Tindal 2001; Rasinski 2000; Samuels 1994). This aspect of oral reading has been shown to be closely connected to overall reading ability. Recognizing that reading rates may differ depending on the time interval designated for rate calculation and that nervousness or fatigue may affect oral reading at different stages of the student's performance, two measures of rate were derived from this study. These were: a rate in words per minute, based on the full duration of students' presentations as measured by the audio software used for the online scoring of the oral reading presentations, and a rate based on the number of words read by students in the first minute of oral reading. The second measure, number of words read in the first minute, was calculated by the recording software based on the first 60 seconds of the digitized recordings of students' oral reading.

Students were not instructed to read the passage quickly. Rather, they were asked to read the story, one of the easiest fourth-grade passages from the main NAEP reading assessment, as if they were reading it to someone who had never heard it before. Consequently, readers of this report are urged not to consider the reading rates in this study as representative of students' fastest rates; instead, these reading rates may be considered a natural pace for most of the students—a rate they might assume when attempting to present their best oral reading performance. Students with the most severe oral reading difficulties (less than 1 percent of the subsample taken from the fourth-grade participants in the 2002 main NAEP reading assessment) were screened out of participation in the oral reading study.

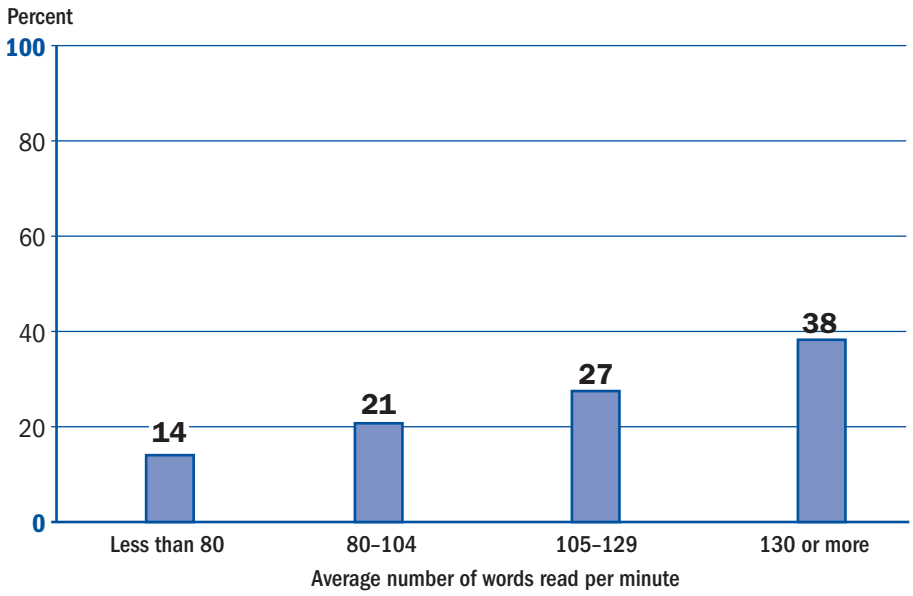
Students spent an average of 1 minute and 40 seconds reading the entire 198-word oral reading study passage. Their average rate across the entire passage was 119 words per minute.

Figure 3-1 displays the percentage of students who read the entire 198-word oral reading study passage at various rates, where rate is the average number of words read per minute. The categories of oral reading rate were based in the research literature. The range of oral reading rates for fourth-graders reading a fairly difficult passage has been reported to be about 74 –128 words per minute with a mean of 101 words per minute (Pinnell et al. 1995). The 1992 Integrated Reading Performance Record (IRPR) study reported oral reading rates with intervals of less than 80, 80–104, 105–129, and 130 or more words per minute (Pinnell et al. 1995). Hence, those categories were used in the present study.

The majority of fourth-graders (approximately 65 percent) read this 198-word segment at a rate of at least 105 words per minute. In the IRPR (1992) study more than one-half (approximately 57 percent) of the students read orally at a speed of at least 105 or more words per minute. While the results of the 1992 and 2002 studies are not scientifically comparable, they do offer corresponding depictions of the oral reading rates of fourth-grade students.

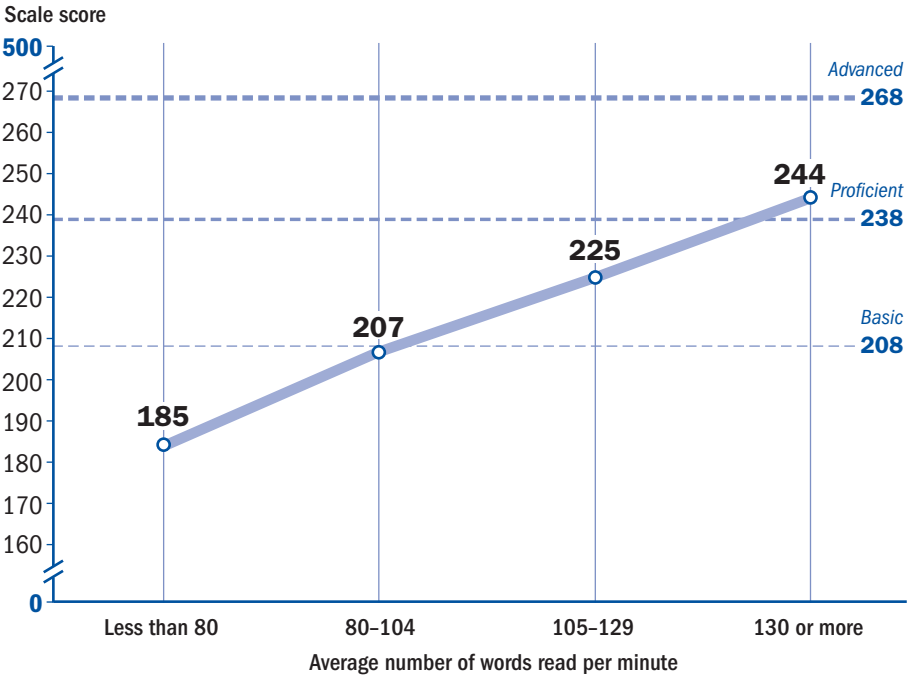
The relationship between average reading rate and average reading score is displayed in figure 3-2. The relationship is a positive one—faster reading rates were associated with higher comprehension scores. The reader is reminded that this finding does not suggest a causal relationship between oral reading rate and comprehension. Students with better comprehension skills are likely to read the passage faster because they have less difficulty reading individual words, and thus are able to move more quickly through the passage. The faster readers may have developed skills in automatic decoding of words and reading phrases or sentences that are necessary for focusing on overall text meaning. Putting together words into phrases and sentences more quickly may facilitate connecting ideas across the text—an important component of comprehension (Crawford, Stieber, and Tindal 2001; Rasinski 2000; Samuels 1994).

Figure 3-1. Percentage of students, by average number of words read per minute, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 3-2. Average reading scale scores in relation to the achievement levels, by average number of words read per minute, grade 4: 2002

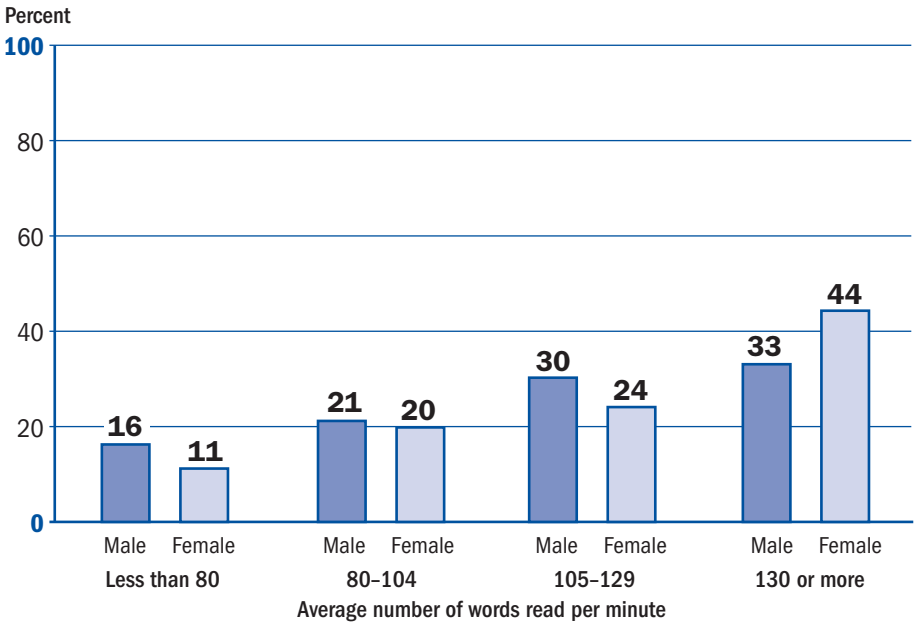


NOTE: The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

The task of the beginning reader is to move from the early phases of "sounding out" words to the more skilled phase in which word recognition occurs almost instantaneously. This developmental change allows the word recognition process to occur fluently, automatically, and rapidly enough to allow for the abstraction of meaning from text. Without efficient (automatic) word recognition skills, comprehension is impaired even when the underlying comprehension processes are well developed (Fletcher and Lyon 1998). As shown in figure 3-2, the average score (185) for students whose reading rate was less than 80 words per minute and the average score (207) for students whose reading rate was between 80 and 104 words per minute for the entire passage fell below the *Basic* achievement level; the average score (225) for students who read between 105 and 129 words per minute was above the *Basic* achievement level; and the average score (244) for students who read at least 130 words per minute was above *Proficient*.

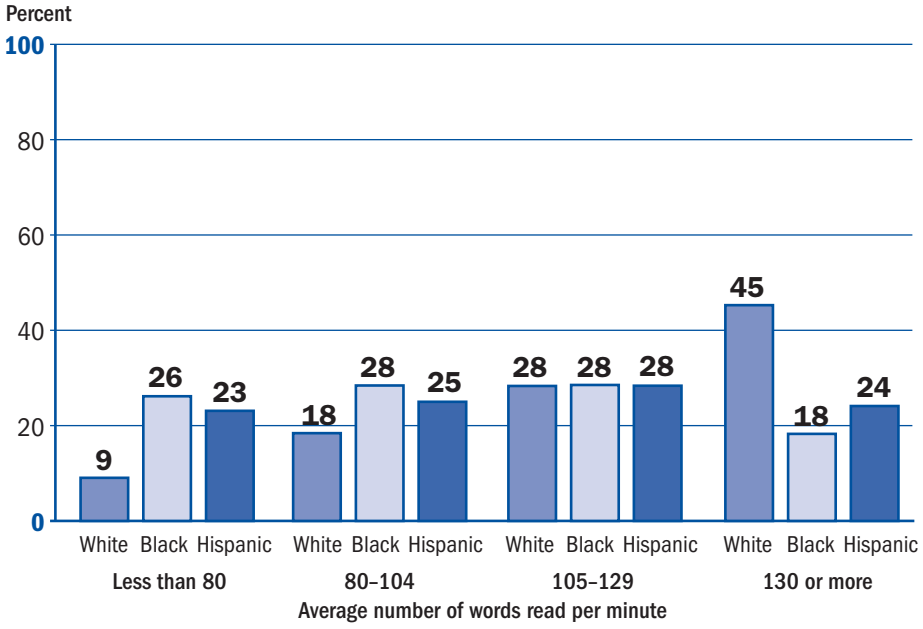
Figures 3-3 and 3-4 present reading rate results by gender and race/ethnicity subgroups, respectively. As shown in figure 3-3, female students tended, on average, to read the entire passage at a faster rate than their male counterparts. Approximately 44 percent of female fourth-graders and 33 percent of male fourth-graders read the entire passage at an average rate of at least 130 words per minute. Differences in average reading rate are also evident across groups of students categorized by race/ethnicity. Figure 3-4 shows that 45 percent of White students read aloud at an average rate of 130 words per minute or more, compared to 18 percent of Black students and 24 percent of Hispanic students.

Figure 3-3. Percentage of students, by gender and average number of words read per minute, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

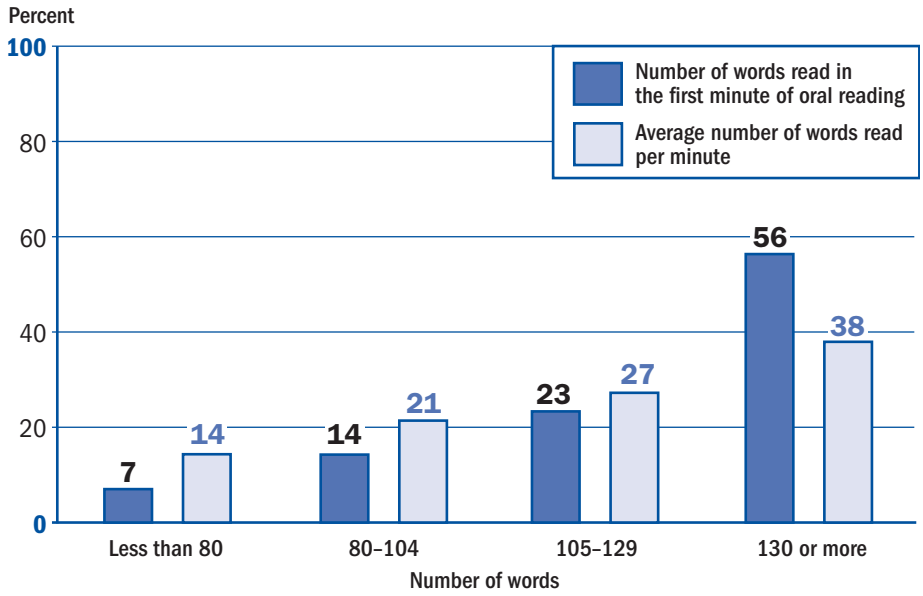
Figure 3-4. Percentage of students, by race/ethnicity and average number of words read per minute, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Another measure of oral reading rate is the number of words students were able to read within the first minute of their oral reading (Crawford, Stieber, and Tindal 2001; Pinnell et al. 1995; Rasinski 2000; Samuels 1994). Figure 3-5 presents students' oral reading rates in this manner. As shown in the figure, 23 percent of the students read between 105 and 129 words and approximately 56 percent read 130 or more words (approximately 79 percent of the students read at least 105 words) in the initial minute of their oral reading. When rate was measured across the entire passage, 27 percent of the students read at an average rate of 105 to 129 words per minute, and 38 percent of the students read at least 130 words per minute (approximately 65 percent of the students read at least 105 words per minute). Thus, the data for words read in the first minute seem to reflect higher percentages of students reading at faster rates than the data for words per minute across the entire passage.

Figure 3-5. Percentage of students, by number of words read in the first minute of oral reading and average number of words read per minute, grade 4: 2002

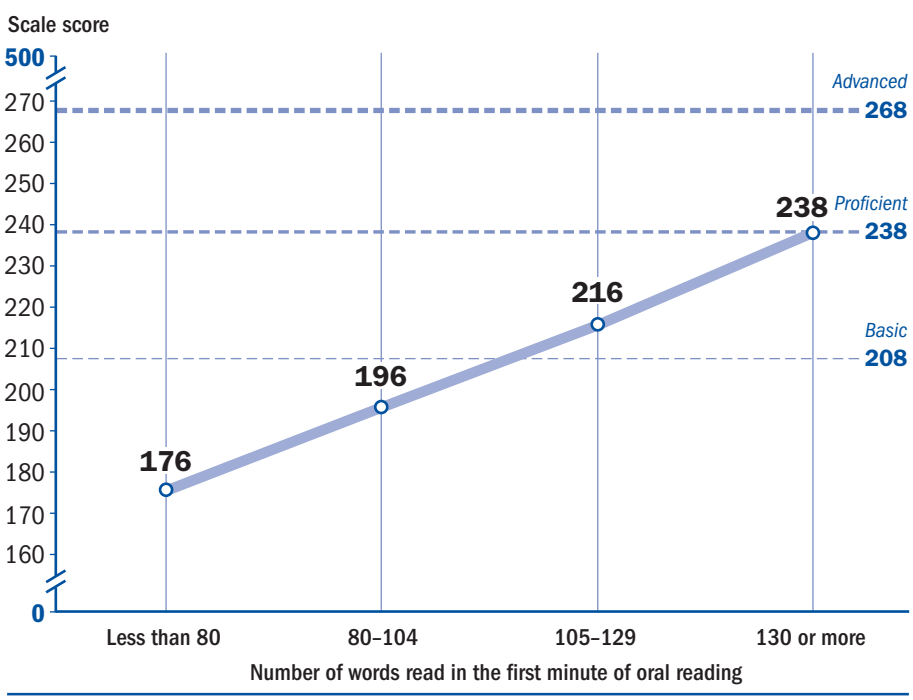


NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Regardless of the reason, it is evident that measures of oral reading rate may produce somewhat different results depending on the method used to calculate the rate. By comparing students' reading rates for the entire passage and for the first minute of reading, the findings of this study suggest that measures of rate taken over very short durations may yield an impression that students' are able to read faster than profiles based on measures taken over longer time periods. Therefore, the reader is urged to consider the importance and impact that the duration of reading demonstrations may have on rate measurements when considering the results of this and other research on students' reading ability.

The relationship between the number of words read in the first minute and students' average reading scores is presented in figure 3-6. Even though different measures of calculating reading rate produced somewhat different results, the association with reading comprehension remained positive. The more words students read during the first minute of their oral reading, the higher their average reading scores. The average score for students who read less than 80 words per minute for the first minute of oral reading (176) and the average score for students who read between 80 and 104 words during the first minute (196) fell below the *Basic* achievement level. Students who read between 105 and 129 words during the first minute had an average score of 216, which fell within the *Basic* achievement level, and students who read 130 words or more during the first minute had an average score of 238, which fell at the *Proficient* achievement level.

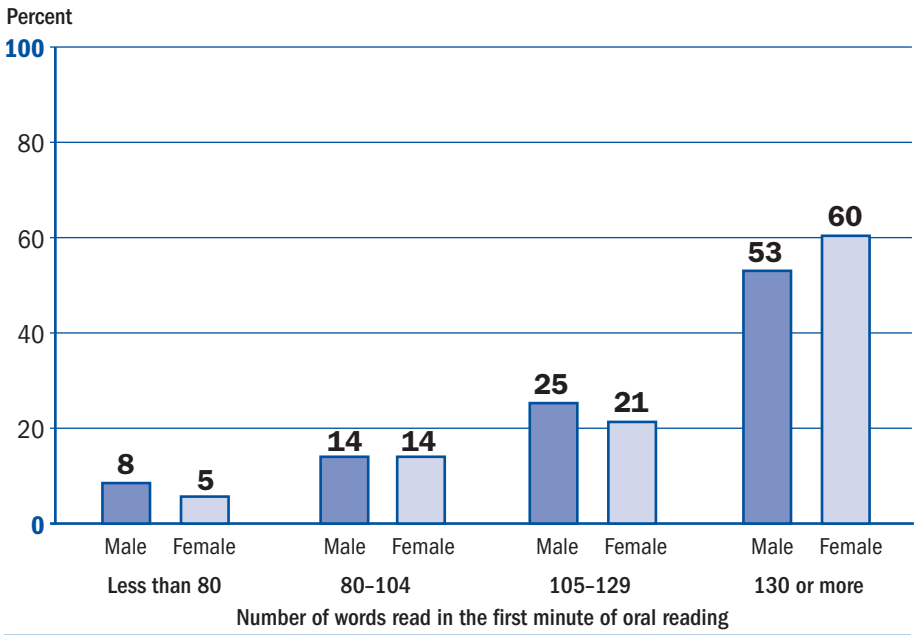
Figure 3-6. Average reading scale scores in relation to the achievement levels, by number of words read in the first minute of oral reading, grade 4: 2002



NOTE: The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

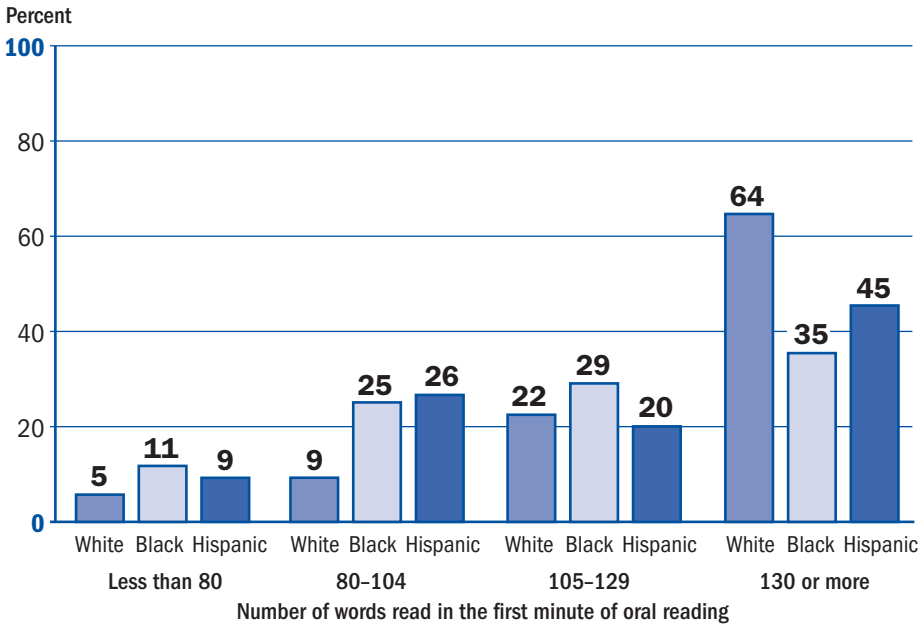
Figures 3-7 and 3-8 present the distribution of number of words read in the first minute by gender and racial/ethnic subgroups, respectively. Sixty percent of female students read 130 or more words in the first minute of oral reading, and 53 percent of male students read 130 or more words in the first minute. Furthermore, 44 percent of female students and 33 percent of male students read 130 or more words per minute across the entire passage, as shown in figure 3-3. Figure 3-8 shows differences in the number of words read in the initial minute across the three racial/ethnic subgroups of students. A higher percentage of White students than Black students and Hispanic students read at an average rate of at least 130 words during the first minute. Sixty-four percent of White students, 35 percent of Black students, and 45 percent of Hispanic students read at least 130 words in the first minute.

Figure 3-7. Percentage of students, by gender and number of words read in the first minute of oral reading, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 3-8. Percentage of students, by race/ethnicity and number of words read in the first minute of oral reading, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Chapter 4. Reading Fluency

Generations of teachers have considered that good silent readers exhibit fluent oral reading abilities. Skilled readers not only recognize and read words quickly, but also deliver a smooth oral reading performance that reflects their understanding of the text they are reading. Research supports the notion that oral reading fluency is strongly connected to students' overall reading ability (Allington 1983; Fuchs et al. 2001; Good, Kame'enui, and Simmons 2001; Reutzel and Hollingsworth 1993; Shinn et al. 1992; Stayter and Allington 1991). Yet, while rate and accuracy of oral reading are relatively straightforward characteristics both to observe and measure, it has proven more difficult to capture and measure the fluency, or "ease," with which children read texts. It is precisely this attribute of reading, however, that appears to have a strong connection to students' comprehension of the texts they read.

Describing Oral Reading Fluency

Despite continuing research in oral fluency and its association with reading development, there is no widely accepted definition of reading fluency. Reading experts use the term "fluency" to express various characteristics of oral reading (Zuttell and Rasinski 1991). Some associate fluency with the "automaticity" of word recognition that allows readers to focus their attention on comprehending ideas, rather than on decoding words and phrases (Samuels 1994). Other researchers believe that measures of phrasing and expressiveness should be part of the definition of fluency (Bowers and Young 1995). Despite the differences in the way researchers discuss fluency, most concur that it is closely associated with students' ability to understand the passages they encounter.

In their attempts to understand a text, readers rely on their prior experiences with the organization and language patterns of text they have encountered. They need to recognize individual words and their meanings, but must go beyond simply recognizing words to understanding phrases and ideas within the larger text. As readers develop the ability to recognize phrases and words, they free attention that can be devoted to the interpretation of the words and to the entire text they are reading (Schreiber 1980; Schreiber 1987; Snow, Burns, and Griffin 1998).

Studies in the field of oral reading have also shown that when readers adequately understand the texts they are reading, they recognize the phrasal and syntactical structures employed by the author and will attempt to replicate those structures in their own oral delivery (Schreiber 1980; Schreiber 1987; Snow, Burns, and Griffin 1998). Furthermore, an understanding of the larger connections between parts of the text (e.g., among the events, characters, setting, and motivations of a narrative text; or among descriptions, concepts, and analyses of an informative text) is related to fluency in that such awarenesses enable the reader to use appropriate emphasis and expressiveness (both aspects of fluency) in oral reading (Juel 1991; LaBerge and Samuels 1974; NICHHD 2000).

The Oral Reading Fluency Scale

For the purpose of describing and evaluating "fluency," the oral reading study employed a four-level oral reading fluency scale that was first developed for the 1992 study. The oral reading study fluency scale is presented in figure 4-1. It was applied holistically, rather than analytically: trained raters were asked to categorize individual students' oral reading by the level description that best categorized their overall performance.

Figure 4-1. NAEP oral reading fluency scale, grade 4: 2002

Fluent	Level 4	Reads primarily in larger, meaningful phrase groups. Although some regressions, repetitions, and deviations from text may be present, these do not appear to detract from the overall structure of the story. Preservation of the author's syntax is consistent. Some or most of the story is read with expressive interpretation.
	Level 3	Reads primarily in three- or four-word phrase groups. Some small groupings may be present. However, the majority of phrasing seems appropriate and preserves the syntax of the author. Little or no expressive interpretation is present.
Nonfluent	Level 2	Reads primarily in two-word phrases with some three- or four-word groupings. Some word-by-word reading may be present. Word groupings may seem awkward and unrelated to larger context of sentence or passage.
	Level 1	Reads primarily word-by-word. Occasional two-word or three-word phrases may occur—but these are infrequent and/or they do not preserve meaningful syntax.

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

This fluency scale focused on several key elements of oral reading. First, the apparent grouping of words, or phrasing, produced by students was central to describing their fluency. For the purposes of scoring, this phrasing manifested itself through the intonation, stress, and pauses exhibited by oral readers, and in the perceived rise and fall of pitch, or simply by the hesitation or pause between phrase endings and phrase beginnings as students read the passage aloud.

A second key element of the fluency scale involved listening for whether readers adhered to the author's syntax and sentence structure. Recognizing the author's syntax is critical to smooth oral delivery of texts because identical groups of words can acquire varying syntactical patterns when readers apply different intonations, stress placements, or pause insertions. Adhering to the author's syntax during oral reading requires that the reader be aware of the ideas expressed in the text. Alternatively, delivering a phrase or sentence in a syntactical structure that differs from the one intended by the author may indicate that the reader has not comprehended or has momentarily lost track of the "thread" of meaning in a passage.

The last element of the fluency scale was expressiveness in the oral reading presentation. While fourth-graders were not expected to offer a consistently expressive oral presentation, their overall fluency rate was influenced by their ability to convey expressiveness naturally through the passage. The presence of some expressiveness on the part of the oral reader—the interjection of feeling, anticipation, or even a level of characterization—clearly marked the performance of more fluent readers, while the lack of emotional engagement with the text marked less fluent readings of the passage.

A deliberate decision was made not to include accuracy as part of the fluency scale. Accuracy was considered as a separate component from fluency. An analysis of students' accuracy in reading the text was conducted separately, and the data are presented in chapter 2. Word recognition was not considered in the fluency ratings because all readers—even the most fluent—commit some errors when reading from a printed text. Often these errors—substitutions, insertions, or omissions—do not indicate that a reader misunderstands the text, and do not cause such a misunderstanding of the text (Goodman and Goodman 1994). It is possible for readers to attain high levels of comprehension even as their own language experiences and background knowledge may cause them to deviate slightly from the printed word. Research has shown that despite occasional verbal "slips," readers manage to capture the intended meaning of texts. Of course, as these errors increase and as readers more widely diverge from the text they are reading, they are less likely to grasp the underlying meaning of the text (Snow, Burns, and Griffin 1998; Vellutino 1979; Vellutino 1991).

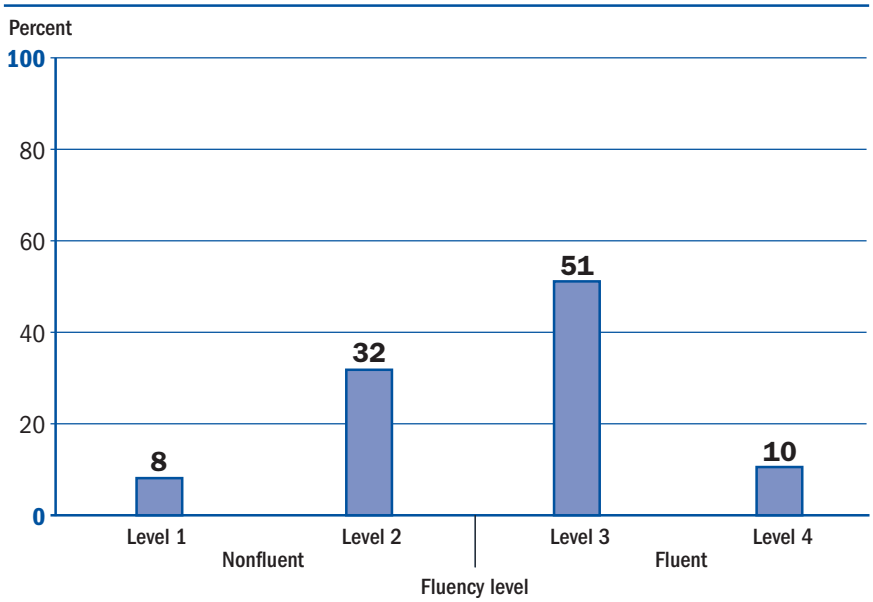
Fourth-Graders' Fluency With the Oral Reading Study Passage

As shown in figure 4-2, when reading from "The Box in the Barn," one of the easiest fourth-grade passages from the main NAEP assessment, approximately 61 percent of fourth-graders attained a level of 3 or 4 on the fluency scale, putting them in the "fluent" range, while the remainder of the students, approximately 40 percent, attained a "nonfluent" rating of a 1 or 2 on the fluency scale. Generally, students who were rated at the upper half of the scale displayed competent oral reading abilities; they read in meaningful phrase groups and preserved the author's syntax, indicating their understanding of the unfolding events and characterization in the passage. Some of the most fluent participants even read with appropriate expressiveness. Students who performed at the lower level of the fluency scale displayed their difficulty with the text by reading primarily in single- or two-word phrases, demonstrating an inability to relate sentences to the global context of the passage. Generally, the struggle these less fluent readers confronted as they worked their way through the individual words on the page appeared to inhibit their overall ability to comprehend the story they were reading (figure 4-1). Readers are reminded that students with the most severe oral reading difficulties (less than 1 percent of the subsample taken from the fourth-grade participants in the 2002 main NAEP reading assessment) were screened out of participation in the oral reading study.

In general, raters looked for oral fluency that conveyed textual meaning through appropriate phrasing and smooth oral delivery. A rating of 3, attained by 51 percent of the fourth-graders, was considered to be within the fluent range of reading, indicating reading primarily in three- or four-word phrases with little or no expressive interpretation. Only 10 percent of students actually read the story with the expressiveness necessary for a rating of 4. Thus, while more than half of the readers could be considered fluent, only a small percentage of the nation's fourth-graders exhibited the consistent adherence to the author's syntax and at least minimal expression characteristic of the level 4 readers described in the fluency scale.

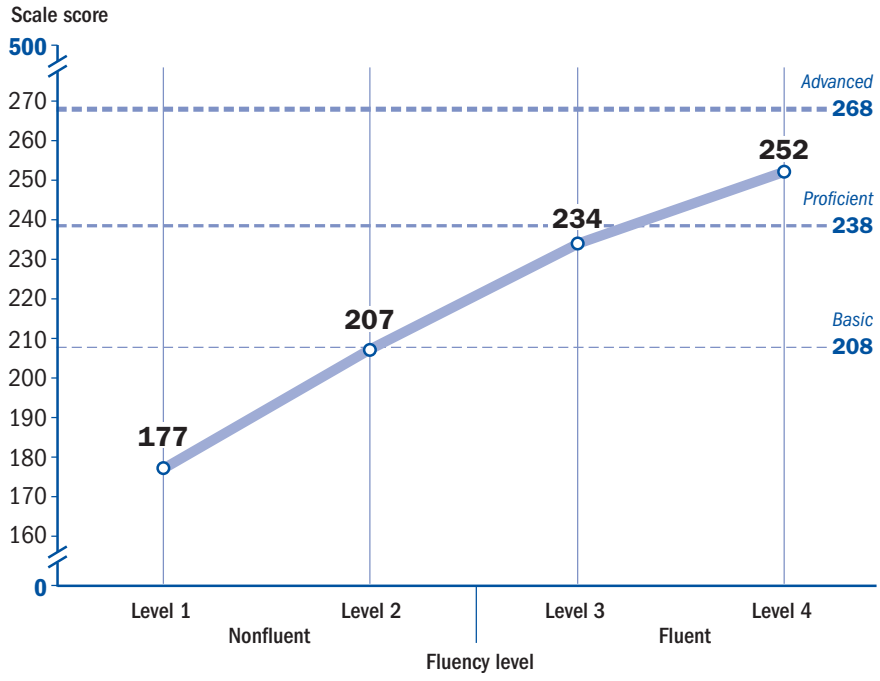
An examination of students' fluency ratings and their average scores on the main NAEP reading assessment shows a strong correlation between fluency and overall reading ability. Figure 4-3 displays the relationship between fluency scale ratings and students' average scores on the NAEP reading assessment. Students at the lowest fluency ratings had the lowest average scores; as fluency rates increased, so did the average scale scores of fourth-grade readers. While the 1992 IRPR and this report cannot be considered comparable because they used different administration procedures and different reading passages, the two studies produced analogous findings on fluency and overall reading ability. In both studies, fluency levels are positively related to average reading scores; as fluency ratings increased so did average scale scores. Hence the findings of these two NAEP studies corroborate those of other researchers who have observed and documented a strong connection between fluency and reading achievement; Reutzel and Hollingsworth 1993; NICHHD 2000).

Figure 4-2. Percentage of students, by NAEP reading fluency scale level, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 4-3. Average reading scale scores in relation to the achievement levels, by NAEP reading fluency scale level, grade 4: 2002



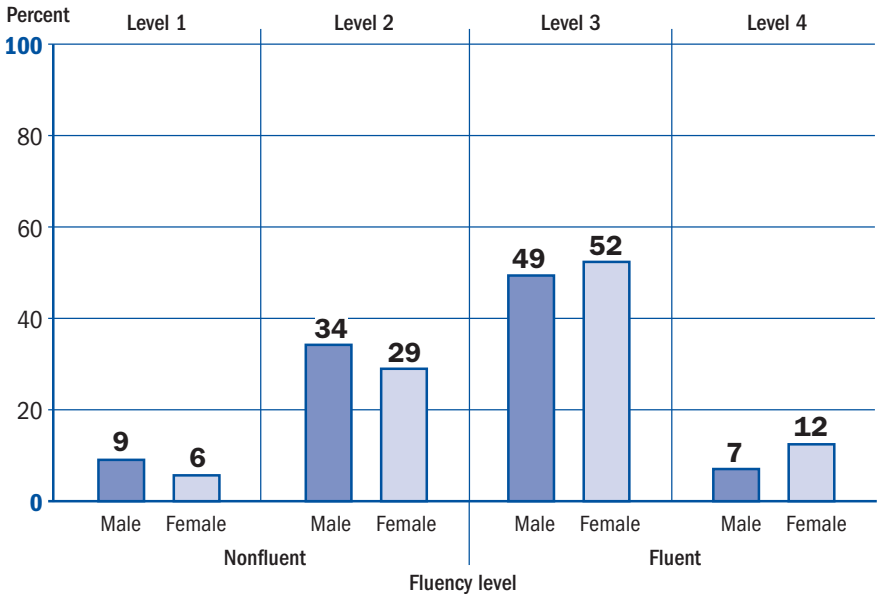
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

The figure also maps the NAEP achievement levels—*Basic*, *Proficient*, and *Advanced*—onto the scale scores for the NAEP assessment. As described in chapter 1 of this report, the NAEP achievement levels are intended to communicate three levels of performance, with the *Proficient* level or above as the desired goal for all students. This additional information shows that students who were categorized as level 4 readers (the highest fluency category in the NAEP fluency scale) had an average score that placed them in the *Proficient* achievement-level range. Level 3 readers had an average score that placed them within the *Basic* achievement-level range. Those students who were rated as "nonfluent" (levels 2 and 1 on the NAEP fluency scale) had average scores that were below *Basic*. The average score of level 2 readers on the fluency scale (207) was one point below the *Basic* cut point of 208.

Figures 4-4 and 4-5 present fluency scale results for subgroups of students defined by gender and by race/ethnicity, respectively. Consistent with results from the NAEP 2002 and previous NAEP reading assessments, which show female students outperforming male students, on average, in reading comprehension, female students were more likely than their male counterparts to be rated as "fluent" (levels 3 and 4 combined). Nearly two-thirds (approximately 64 percent) of the female participants were rated in fluency levels 3 and 4 combined, as compared to approximately 56 percent of male participants.

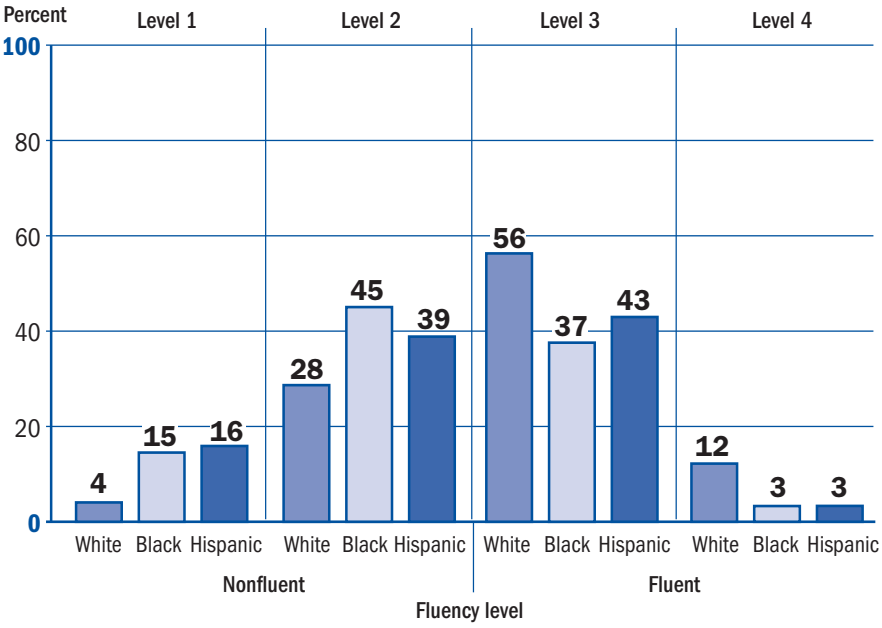
When the data are analyzed by race/ethnicity, a higher percentage of White students than Black and Hispanic students were rated as "fluent" (levels 3 and 4). No statistically significant difference was found between the likelihood of Hispanic students and Black students to be rated as "fluent." Approximately 68 percent of White students, 40 percent of Black students, and 46 percent of Hispanic students attained fluency ratings in the top half of the fluency scale.

Figure 4-4. Percentage of students, by gender and NAEP reading fluency scale level, grade 4: 2002



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Figure 4-5. Percentage of students, by race/ethnicity and NAEP reading fluency scale level, grade 4: 2002



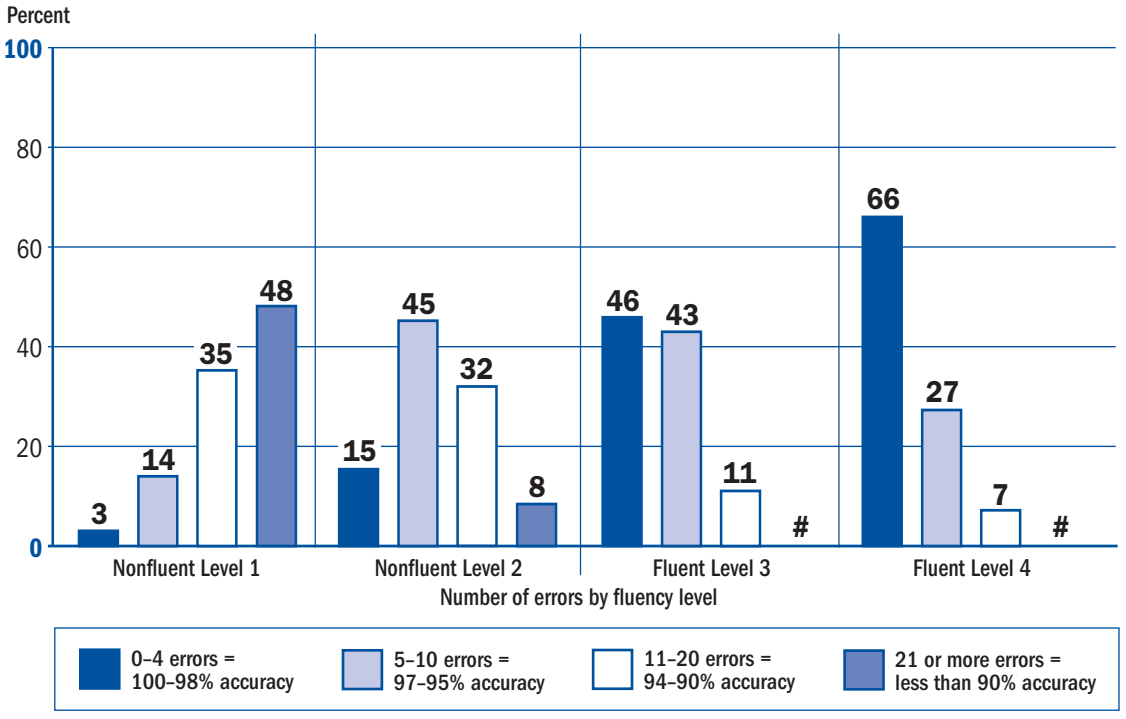
NOTE: Detail may not sum to totals because of rounding. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Fourth-Graders' Oral Reading Fluency and Accuracy

Figure 4-6 displays the accuracy—percentage of words read correctly—at each of the four levels of fluency. The degree of accuracy of fourth-graders' oral reading was based on the total number of errors (both meaning-change and non-meaning-change) in oral reading performance. Approximately 89 percent of the students at level 3 and approximately 93 percent of the students at level 4 read the 198-word passage with 10 or fewer errors (an accuracy rate of at least 95 percent). However, even at level 2, more than one-half (approximately 60 percent) of the students read the passage with 10 or fewer errors.

Figure 4-6 also shows that at fluency level 1, more than four-fifths (approximately 83 percent) of the students read the excerpt with 11 or more errors (an accuracy rate of less than 95 percent), and nearly one-half (48 percent) of the students at this level read the excerpt with 21 or more errors (an accuracy rate of less than 90 percent). Higher fluency levels, as well as higher average scale scores, were associated with lower numbers of errors when reading the assigned text. Results pertaining to fluency and overall reading achievement are similar to those for accuracy and overall reading achievement presented in chapter 2 and for rate and overall reading achievement presented in chapter 3.

Figure 4-6. Percentage of students, by degree of reading accuracy and NAEP reading fluency scale, grade 4: 2002

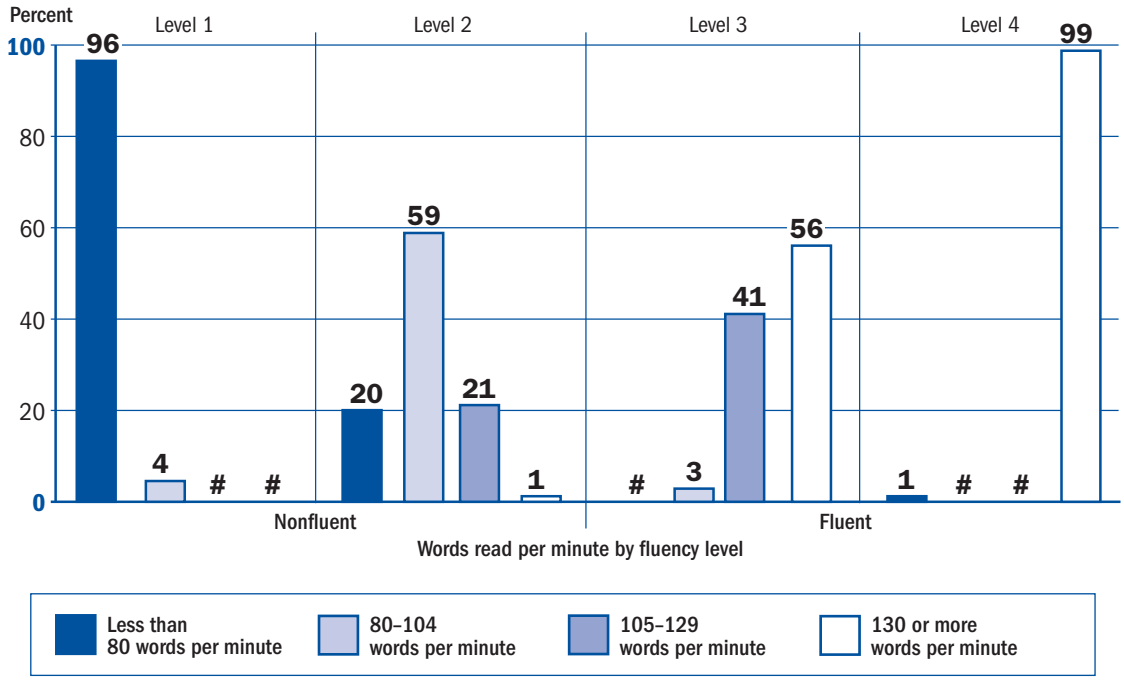


The estimate rounds to zero.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Fourth-Graders' Oral Reading Fluency and Rate

The relationship between oral reading fluency and rate, in terms of words read per minute, is displayed in figure 4-7. These data indicate that the most fluent readers also tended to demonstrate faster rates of oral reading. The overall reading rate for fourth-graders was 119 words per minute across the entire passage. However, among the most fluent readers (level 4), 99 percent read at a rate of 130 or more words per minute. At level 3, 56 percent read at this rate. The majority of level 2 readers (59 percent) were reading at a rate of between 80 and 104 words per minute. Nearly all (96 percent) of the readers in level 1, the lowest fluency category, read the passage at a rate that was less than 80 words per minute. Here again, it is evident that students rated at this lowest fluency level were struggling with reading this passage orally.

Figure 4-7. Percentage of students, by reading rate and NAEP reading fluency scale level, grade 4: 2002



The estimate rounds to zero.
NOTE: Detail may not sum to totals because of rounding. Reading rate is defined by number of words read per minute. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Chapter 5. Summary of Oral Reading Study Results

This chapter summarizes findings about fourth-grade students' oral reading from a special study that was part of the 2002 National Assessment of Educational Progress (NAEP). The data derived from the subsample of students who participated in the 2002 NAEP reading assessment suggest that the three separate components of oral reading ability—accuracy, rate, and fluency—are very much related to each other—and all three have a connection to reading comprehension as measured by the main NAEP reading assessment. "Fluent" readers in this study were likely to read higher percentages of words accurately, to read the passage at a faster rate, and to have scored higher, on average, on the NAEP reading assessment than "nonfluent" readers. Most of the students could read the study passage fluently, with a fairly high degree of accuracy, and at a rate of at least 105 words per minute. However, a group of students whose average scale score and labored oral reading performance suggested they were struggling also comprised the group that demonstrated, on average, the lowest performance on measures of accuracy, speed, and fluency.

Oral Reading Accuracy

One major finding of the oral reading study is that fourth-graders were able to read the excerpt from "The Box in the Barn," one of the easiest fourth-grade passages from the main NAEP reading assessment, with a fairly high degree of accuracy. A general consensus among researchers indicates that a 95 percent level of accuracy in word recognition is necessary for independently understanding a given text (Harris and Sipay 1985, pp. 193–203; Kuhn and Stahl 2003). The oral reading study shows that approximately 75 percent of students demonstrated at least 95 percent accuracy of word-reading skills when reading aloud from the passage used in this study. Students with the most severe oral reading difficulties (less than 1 percent of the subsample of participants in the 2002 NAEP reading assessment) were screened out of participation in the oral reading study.

For the most part, the fourth-grade participants in this study were well able to decode the excerpt that was presented to them. When all oral reading errors (both those that changed the meaning of the text and those that preserved it) were counted, more than one-third (35 percent) of the students read the passage with at least 98 percent accuracy, and two-fifths (40 percent) read with between 95 and 97 percent accuracy; thus a total of approximately three-quarters (75 percent) of the students read this passage at the recommended accuracy level of 95 percent or higher. Further, the 19 percent of students whose average score (206) was not significantly different from the *Basic* achievement-level cut score (208) read with 90 to 94 percent accuracy on average. The six percent of the students who read with less than 90 percent accuracy had an average scale score (180) that fell 20 points below the *Basic* achievement-level cut score (208). Thus, it seems that, on average, about 75 percent of the students who participated in the oral reading study did not have major problems decoding the oral reading excerpt used in this study. The results of the IRPR (Pinnell et al. 1995), though not directly comparable, were similar; more than half of the fourth-grade participants read with at least 96 percent accuracy.

A similar picture of fourth-graders' accuracy emerges when specifically examining the errors they made that changed the meaning of the text. Approximately nine out of ten students were able to read with at least 95 percent accuracy when only meaning-change errors were counted. Nevertheless, the frequency of meaning-change errors demonstrated a negative relationship to comprehension, like the frequency of all types of errors (regardless of their effect on text meaning). When only meaning-change errors were counted, the average main NAEP reading score for students reading with at least 98 percent accuracy (231) was within the above *Basic* range (above 208) and 7 points below the *Proficient* achievement-level cut score (238). The average main NAEP reading score for students reading with between 95 and 97 percent accuracy (206) was not significantly different from the *Basic* achievement-level cut score (208). Those students who read with between 90 and 94 percent accuracy had an average main NAEP reading score (180) that fell 28 points below the *Basic* achievement-level cut score (208).

Students' Self-Correction of Errors in Oral Reading

Students' self-correction of oral reading errors seemed to be somewhat dependent upon whether or not those errors resulted in a meaning change. Although nearly one-half of the students self-corrected at least 50 percent of their meaning-change errors, only about one-quarter of students self-corrected at least 50 percent of their errors that did not cause a meaning change. Generally, the relationship between the percentage of all errors the students corrected and their average reading scores is positive—the greater the proportion of all self-corrected errors (counting all error types), the higher the average score. Students who corrected more than 75 percent of all their errors had higher average scores (237) than students who corrected 50–75 percent of their errors (230).

Oral Reading Rate

One facet of this study's findings about the rate of students' oral reading may bear particular relevance for parents, teachers, researchers, and others interested in the measurement of reading abilities. In this study, the picture of fourth-graders' oral reading rate varied depending upon the way rate was defined and measured. When rate was defined as words per minute for the entire length of the excerpt, there was less variation in the percentage of students in each rate category (i.e., 14 percent of students read at less than 80 words per minute; 21 percent read 80–104 words per minute; 27 percent read 105–129 words per minute; and 38 percent read at least 130 words per minute). However, when rate was measured as words per minute for just the first minute of reading, the percentage of students in the fastest rate category was twice the percentage in the second fastest rate category. Specifically, 7 percent of students read at less than 80 words per minute; 14 percent read from 80–104 words per minute; 23 percent read at 105–129 words per minute, but 56 percent read 130 or more words per minute for the first minute).

This disparity in oral reading rate results depending on the unit used to calculate rate presents a caveat: any assessment of reading rate may be partially reflective of the method used to measure that rate. Therefore, parents and educators should interpret research findings and assessment data very carefully, taking into account the methodologies used to derive the data.

The average score for students who read less than 80 words per minute (185) fell below the *Basic* achievement-level cut score, and the average score for students who read less than 105 words per minute for the entire passage (207) was not significantly different from the *Basic* achievement-level cut score (208). The average score for students who read between 105 and 129 words per minute (225) was within the above *Basic* achievement level, and the average score for students who read at least 130 words per minute (244) fell within the above *Proficient* achievement level. The average score for students who read less than 80 words per minute for the first minute of oral reading (176) and the average score for students who read between 80 and 104 words during the first minute (196) fell below the *Basic* achievement-level cut score. Students who read between 105 and 129 words during the first minute had an average score of 216, within the above *Basic* achievement level, and students who read 130 words or more during the first minute had an average score of 238, within the above *Proficient* achievement level. These data indicate a positive relationship between reading comprehension and reading rate, whether measured for the entire passage or for the first minute of reading.

Most of the fourth-graders in this study read at an expected rate. Nearly two-thirds (approximately 65 percent) of the fourth-graders read the passage with an average rate of at least 105 words per minute for the entire passage, and nearly four-fifths (approximately 79 percent) read the passage with a rate of at least 105 words for the first minute of reading. While the oral reading study results cannot be compared directly to the results of the IRPR (Pinnell et al. 1995), the earlier report showed findings that parallel the results of this study: in 1992 more than one-half of the students (approximately 57 percent) read orally at speeds of at least 105 words per minute

for the entire passage. Based on the results of the 2002 oral reading study, it can be inferred that, when reading the excerpt from "The Box in the Barn," neither decoding nor oral reading rate is a major problem for about 75 percent of the fourth-graders who participated in the oral reading study.

The data on reading rate show a positive relationship between speed of oral reading and reading comprehension. Clearly, no causal relationship can be established from these data; however, it would seem likely that the more competent readers, with larger reading vocabularies, well-developed word-reading skills, and an ability to comprehend grade-appropriate texts, could read through a text at a faster rate than students with less skill in these areas. Nearly two-thirds of the students who participated in the oral reading study read the excerpt with an average rate of at least 105 words per minute. In examining the words read in just the first minute of oral reading, the reading rate appears to be somewhat faster—approximately 79 percent were able to read at least 105 words in the initial minute. Although the results of the 1992 IRPR are not scientifically comparable to the present findings, the earlier research also showed a correspondence between reading rate and comprehension.

Oral Reading Fluency

The term "fluency" in relation to oral reading has been defined in different ways by different researchers. Generally, however, fluency is considered to be an effortless, smooth, and coherent oral production of a given passage. For the purpose of this study, fluency was defined in terms of phrasing, adherence to the author's syntax, and expressiveness. In coding students' oral reading in the study, it was considered as a distinct attribute of oral reading—separate from accuracy and rate. Nevertheless, the results in this chapter point to a strong relationship between fluency, accuracy, and rate—with all three additionally showing a strong relationship to reading comprehension, as measured by the NAEP 2002 reading assessment.

Based on the NAEP fluency scale developed for this study, approximately 61 percent of fourth-graders read the study passage in a manner that was judged to be "fluent" (i.e., in the top two levels of the four-level fluency scale). These students read most, if not all, of the passage in phrase groups that adhered to the author's syntax—indicating that they were paying attention to the meaning and language structures of the text. The 10 percent of students who were rated in the top level also incorporated a degree of expressiveness in their reading.

Students' ratings of fluency had a positive relationship to comprehension—higher fluency ratings were associated with higher average reading scores. Students who were rated as "nonfluent" (levels 1 and 2 on the fluency scale) had average reading scores that placed them below the *Basic* achievement level.

The results of the oral reading study show a strong positive relationship between oral reading fluency and reading comprehension, as measured by average scale score and as described by NAEP reading achievement levels. Students who were rated at the highest fluency level (4) scored above the *Proficient* level, with an average scale score of 252. Students who were rated at the second highest fluency level (3) fell well above the *Basic* achievement-level cut score (208) and below the *Proficient* achievement-level cut score (238), with an average score of 234. In contrast, students who were rated in both of the nonfluent levels scored on average below the *Basic* level, which is set at 208. Students at fluency level 2 scored 207 on average, and students at fluency level 1 scored 177 on average. These findings on the correspondence between fluency and comprehension corroborate those of the 1992 IRPR, which also showed that the students who had higher fluency ratings also scored higher in comprehension (Pinnell et al. 1995).

Subgroup Performance

The accuracy, rate, and fluency results for gender and racial/ethnic subgroups parallel the results of the main NAEP reading assessment. The groups of students who demonstrated higher performance levels on the main assessment also tended to read with greater accuracy. Female fourth-graders were generally more accurate in their reading than their male counterparts. Compared to 37 percent of female students reading with at least 98 percent accuracy, 32 percent of male students read with that level of accuracy. A higher percentage of White students than Black students read with at least 98 percent accuracy, but the apparent difference between White students and Hispanic students who read with at least 98 percent accuracy was not found to be statistically significant. Thirty-eight percent of White students, 23 percent of Black students, and 31 percent of Hispanic students read the passage with 98 percent accuracy. The difference in the percentages of White and Black students was statistically significant, but the difference between the percentages of White and Hispanic students was not.

Differences in average reading rate are also evident across subgroups of students categorized by gender and race/ethnicity. Forty-four percent of female students and 33 percent of male students read the entire passage at a rate of 130 or more words per minute, and 60 percent of female students and 53 percent of male students read at a rate of 130 or more words per minute for the first minute of oral reading. Of the White students, 45 percent read aloud at an average rate of at least 130 words per minute across the entire passage, as did 18 percent of Black students and 24 percent of Hispanic students. Sixty-four percent of White students, 35 percent of Black students, and 45 percent of Hispanic students read at a rate of 130 or more words per minute for the first minute of oral reading.

A greater percentage of female students (67 percent) than male students (55 percent) was rated as "fluent," as was a greater percentage of White students (68 percent) compared with their Black (40 percent) and Hispanic (46 percent) peers.

References

Adams, M. J. (1990). *Beginning to Read: Thinking and Learning About Print.* Cambridge, MA: MIT Press.

Allen, N.L., Donoghue, J.R., and Schoeps, T.L. (2001). *The NAEP Technical Report* (NCES 2001-509). U.S. Department of Education, Office of Educational Research and Improvement. Washington, DC: National Center for Education Statistics. Retrieved November 11, 2004, from http://nces.ed.gov/nationsreportcard/pdf/main1998/2001509b.pdf.

Allington, R.L. (1983). Fluency: The Neglected Reading Goal in Reading Instruction. *The Reading Teacher, 36*(6): 556–561.

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

Barrentine, S.J. (Ed.). (1995). *Reading Assessment: Principles and Practices for Elementary Teachers: A Collection of Articles from The Reading Teacher.* Newark, DE: International Reading Association.

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

Bloome, D., and Dail, A.R.K. (1997, December). Toward (Re)Defining Miscue Analysis: Reading as a Social and Cultural Process. *Language Arts, 74*(8): 610–617.

Bowers, P.G., and Young, A. (1995). Individual Differences and Text Determinants of Reading Fluency and Expressiveness. *Journal of Experimental Child Psychology, 60*(3): 428–454.

Calfee, R.C., and Piaotkowski, D.C. (1981). The Reading Diary: Acquisition of Decoding. *Reading Research Quarterly, 16,* 346–373.

Campbell, J.R., Kapinus, B.A., and Beatty, A.S. (1995). *Interviewing Children About Their Literacy Experiences.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

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

Clay, M.M. (1993). *An Observation Survey of Early Literacy Achievement.* Auckland, New Zealand: Heinemann Education.

Conner, B.E. (1988, July/August). The Box in the Barn. *Highlights for Children,* pp. 18–22.

Crawford, L., Stieber, S., and Tindal, G. (2001). Using Oral Reading Rate to Predict Student Performance on Statewide Assessment Tests. *Educational Assessment, 7*(4): 303–321.

Davenport, M.R., and Lauritzen, C. (2002). Inviting Reflection on Reading through Over the Shoulder Miscue Analysis. *Language Arts, 80*(2): 109–118.

Dole, J. A., Duffy, G.G., Roehler, L.R. and Pearson, P.D., (1991). Moving From the Old to the New: Research on Reading Comprehension Instruction. *Review of Educational Research, 61*(2): 239–264.

Fletcher, J.M., and Lyon, G.R. (1998). Reading: A Research-Based Approach. In W.M. Evers (Ed.), *What's Gone Wrong in America's Classrooms* (pp. 50–77). Stanford, CA: Hoover Institution Press.

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

Fuchs, L.S., Fuchs, D., Hosp, M.D., and Jenkins, J.R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5*(3): 239–256.

Good, R.H., Kame'enui, E.J., and Simmons, D.C, (2001). The Importance of Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes. *Scientific Studies of Reading, 5*(3): 257–288.

Goodman, Y.M., and Goodman, K.S. (1994). To Err is Human: Learning about Language Processes by Analyzing Miscues. In R. B. Ruddell, M. R. Ruddell, and H. Singer (Eds.), *Theoretical Models and Processes of Reading*. Newark, DE: International Reading Association.

Gough, P.B. (1972). One Second of Reading. In J. F. Kavanaugh and I. G. Mattingly (Eds.), *Language by Ear and Eye* (pp. 331–358). Cambridge, MA: MIT Press.

Griffey, Q.L., Zigmund, N., and Leinhardt, G. (1988). The Effects of Self-Questioning and Story Structure Training on the Reading Comprehension of Poor Readers. *Learning Disabilities Research, 4*(1): 45–51.

Griffith, P., and Olson, M.W. (1992). Phonemic Awareness Helps Beginning Readers Break the Code. *Reading Teacher, 45*(7): 516–523.

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

Harris, A.J., and Sipay, E.R. (1985). *How to Increase Reading Ability*: *A Guide to Developmental Methods*. New York: Longman.

Harris, T.L., and Hodges, R.E. (1995). *The Literacy Dictionary*. Newark, DE: International Reading Association.

Herman, P.A. (1985). The Effect of Repeated Readings on Reading Rate, Speech Pauses, and Word Recognition Accuracy. *Reading Research Quarterly, 20*, 553–565.

Johnson, E.G., and Rust, K.F. (1992). Population Inferences and Variance Estimation from NAEP Data. *Journal of Education Statistics, 17*: 174–190.

Juel, C. (1991). Beginning Reading. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (Eds.), *Handbook of Reading Research* (pp. 759–788). New York: Longman.

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

Kish, L. (1995). *Survey Sampling*. New York: John Wiley and Sons, Inc.

Kuhn, M.R., Stahl, S.A. (2003). Fluency: A Review of Developmental and Remedial Practices. *Journal of Educational Psychology, 95*(1): 3–21.

LaBerge, D., and Samuels, S. (1974). Toward a Theory of Automatic Information Processing in Reading. *Cognitive Psychology, 6:* 293–323.

Leslie, L., and Taft, M.L. (1985). The Effects of Prior Knowledge and Oral Reading Accuracy on Miscues and Comprehension. *Journal of Reading Behavior, 17*(2): 163–179.

Lipson, M.Y., and Wixson, K.K. (1991). *Assessment and Instruction of Reading Disability: An Interactive Approach*. New York: Harper Collins.

Miller, R.G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlag.

Nathan, R.G., and Stanovich, K.E. (1991). The Causes and Consequences of Differences in Reading Fluency. *Theory Into Practice, 30*(3): pp. 176–84.

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

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

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

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

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

National Center for Education Statistics (2002). *2002 NCES Statistical Standards*. Retrieved November 5, 2004, from http://nces.ed.gov/statprog/stat_standards.asp.

National Institute for Child Health and Human Development. (2000). *Report of the National Reading Panel*. Washington, DC: Author.

National Reading Panel. (2000). Fluency. In *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Bethesda, MD: National Institutes of Health, National Institute of Child Health and Human Development.

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

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

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

Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B., and Beatty, A.S. (1995). *Listening to Children Read Aloud: Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4* (NCES 95–726). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Posner, M.I., and Snyder, R.R. (1975). Attention and Cognitive Control. In R.L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 58–85). Hillsdale, NJ: Erlbaum.

Rasinski, T.V. (2000). Speed Does Matter. *Reading Teacher, 54*(2): pp. 146–151.

Rayner, K., Foorman, B.R., Perfetti, C.A., Pesetsky, D., and Seidenberg, M.S. (2001). How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interest, 2*(2): 31–74.

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

Reutzel, D.R., and Hollingsworth, P.M. (1993). Effects of Fluency Training on Second Graders' Reading Comprehension. *Journal of Educational Research, 86*(6): 325–331.

Samuels, S.J. (1994). Toward a Theory of Automatic Information Processing in Reading, Revisited. In R.B. Ruddell, M.R. Ruddell, and H. Singer (Eds.). *Theoretical Models and Processes of Reading* (pp. 816–837). Newark, DE: International Reading Association.

Samuels, S J., LaBerge, D., and Bremer, C. (1978). Units of Word Recognition: Evidence for Developmental Changes. *Journal of Verbal Learning and Verbal Behavior, 17:* 715–720.

Schreiber, P.A. (1980). On the Acquisition of Reading Fluency. *Journal of Reading Behavior, 12*: 177–186.

Schreiber, P.A. (1987). Prosody and Structure in Children's Syntactic Processing. In R. Horowitz and S. J. Samuels (Eds.), *Comprehending Oral and Written Language*. New York: Academic Press.

Schwanenflugel, P.J., Hamilton, A.M., Kuhn, M.R., Wisenbaker, J.M., and Stahl, S.A. (2004). Becoming a Fluent Reader: Reading Skill and Prosodic Features in the Oral Reading of Young Readers. *Journal of Educational Psychology, 96*(1): 119–129.

Shinn, M.R., Good, R.H., Knutson, N., Tilly, W.D., and Collins, V. (1992). Curriculum-Based Measurement of Oral Reading Fluency: A Confirmatory Analysis of Its Relation to Reading. *School Psychology Review, 21*(3): pp. 459–479.

Snow, C.E., Burns, M.S., and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Stanovich, K.E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly, 21:* 360–406.

Stanovich, K.E. (1993). Romance and Reality Distinguished Educator Series. *Reading Teacher, 47*(4): 280–291.

Stayter, F.Z., and Allington, R.I. (1991). Fluency and the Understanding of Texts. *Theory Into Practice, 30*(3): 143–148.

Torgensen, J.K., Rashotte, C.A., and Alexander, A. (2001). Principles of Fluency Instruction in Reading: Relationships with Established Empirical Outcomes. In M. Wolf (Ed.), *Dyslexia, Fluency, and the Brain* (pp. 333–355). Parkton, MD: York Press.

U.S. Department of Education. (n.d.). *Reading First*. Retrieved December 31, 2003, from http://www.ed.gov/programs/readingfirst/index.html.

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

Vellutino, F.R. (1979). *Dyslexia: Theory and Research*. Cambridge, MA: MIT Press.

Vellutino, F.R. (1991). Introduction to Three Studies on Reading Acquisition: Convergent Findings on Theoretical Foundations of Code-Oriented Versus Whole-Language Approaches to Reading Instruction. *Journal of Educational Psychology, 83:* 437–443.

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

Williamson, R. (2001, April). A Snail's Tale. *Highlights for Children*, p. 2.

Yopp, H.K. (1992). Developing Phonemic Awareness in Young Children. *Reading Teacher, 45*(9): 696–703.

Zuttell, J., and Rasinski, R. (Eds.). (1991). Fluency in Oral Reading [Special issue]. *Theory Into Practice, 30*.

Appendix A. Overview of Procedures Used for the NAEP 2002 Oral Reading Study

This appendix provides information regarding the development, administration, scoring, and analysis of data from the National Assessment of Educational Progress (NAEP) 2002 oral reading study, which was conducted as a special study during the same time period as the administration of the NAEP 2002 reading assessment. Since the results of this special study are related to results from the 2002 reading assessment (referred to subsequently as the "main" reading assessment), some technical information regarding the main assessment is also provided. For a more complete description of the procedures and methods used for the main assessment, see appendix A of *The Nation's Report Card: Reading 2002* (Grigg et al. 2003).

The NAEP 2002 Oral Reading Study and Reading Assessment

The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and test specifications. In addition, NAGB periodically directs the National Center for Education Statistics (NCES) to conduct special studies in conjunction with the national assessment. These special studies often address issues related to academic achievement in various subject areas that are not addressed in the main NAEP assessments. NAGB directed NCES to conduct an oral reading study in 2002 to be based on methodology used in 1992 (Pinnel et al. 1995). The oral reading study was conducted to address an important aspect of reading development that is not captured in the main assessment. The reading assessment focuses specifically on students' comprehension of written passages, and the oral reading study was designed to examine the oral reading abilities of fourth-grade students, and to relate these abilities to reading comprehension, as measured by the main reading assessment.

The design of the NAEP 2002 reading assessment followed the guidelines first developed for the 1992 reading assessment (NAGB 2002). The framework reflects the expert opinions of educators and researchers about reading. Its purpose is to present an overview of the most essential outcomes of students' reading education. The development of this framework and the specifications that guided the development of the assessment, which is administered at intervals to fourth-, eighth-, and twelfth-grade students, involved the critical input of hundreds of individuals across the country, including representatives of national education organizations, teachers, parents, policymakers, business leaders, and the interested general public. The framework development process was managed by the Council of Chief State School Officers (CCSSO) for NAGB.

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

The "contexts for reading" dimension of the NAEP reading framework provides guidance for the types of texts to be included in the assessment. Although many commonalities exist among the different reading contexts, they do lead to real differences in what readers do. For example, when *reading for literary experience*, readers make complex, abstract summaries and identify major themes. They describe the interactions of various literary elements (e.g., setting, plot, characters, and theme). When *reading for information*, readers critically judge the form and content of the text and explain their judgments. They also look for specific pieces of information. When *reading to perform a task* (not assessed at grade 4), readers search quickly for specific pieces of information.

The "aspects of reading" dimension of the NAEP reading framework provides guidance for the types of comprehension questions to be included in the assessment. The four aspects are 1) *forming a general understanding*, 2) *developing interpretation*, 3) *making reader/text connections*, and 4) *examining content and structure*. These four aspects represent different ways in which readers develop understanding of a text. In *forming a general understanding*, readers must consider the text as a whole and provide a global understanding of it. As readers engage in *developing interpretation*, they must extend initial impressions in order to develop a more complete understanding of what was read. This involves linking information across parts

of a text or focusing on specific information. When *making reader/text connections*, the reader must connect information in the text with knowledge and experience. This might include applying ideas in the text to the real world. Finally, *examining content and structure* requires critically evaluating, comparing and contrasting, and understanding the effect of different text features and authorial devices.

The NAEP 2002 reading framework, in addition to describing the nature of the main reading assessment, provided a basis for administration, scoring, and reporting of a special study of oral reading abilities. Specifically, it called for the 2002 oral reading study to replicate procedures used to examine and report students' oral reading abilities in the 1992 Integrated Reading Performance Record (IRPR) by Pinnell et al. (1995). The 1992 IRPR was, in fact, a much broader special study that also included interviews with students about their literacy habits and attitudes, as well their classroom experiences related to reading instruction. The results of the IRPR are available in two reports—one that focused on the oral reading part of the study (Pinnell et al. 1995), and one that focused on the literacy interviews conducted with the fourth-grade participants (Campbell, Kapinus, and Beatty 1995). The 2002 oral reading study focused entirely on oral reading and did not replicate any other portions of the 1992 special study.

Foremost among the rationales set forth by NAGB for conducting the oral reading study was the recognition that oral reading is used frequently in classrooms to measure reading development. By observing students reading aloud, teachers are often able to examine elements of students' reading development that are not easily captured in paper-and-pencil assessments. Teachers commonly recognize that students who are able to read grade-appropriate materials fluently, accurately, and with adequate speed are developing the skills they need to comprehend texts. Like classroom teachers, NAGB recognized that combining observations of students' oral reading with written tests of reading comprehension creates a more complete profile of students' overall reading development.

Although the oral reading study replicated most of the procedures used in the 1992 IRPR oral reading component, several enhancements to procedures and the use of different reading materials preclude the possibility of comparing results between the two studies. Among the several enhancements, the most notable were the use of a programmed administration delivered via a computer laptop and the digital recording of students' responses. In 1992, the study used cassette tape recorders to record students' oral reading, and the study was administered by trained interviewers who conducted the interview sessions with the participants, following a written administration manual. The other major difference between the 1992 and 2002 studies was the use of different reading materials. Because the reading passage that had been read aloud by students in the 1992 study had been released to the public when the reports were published, it was not possible to use the same passage in 2002. Both the 1992 passage and the 2002 passage were selected for the study, in part, because they were considered to be less difficult than the other passages in the fourth-grade NAEP reading assessment.

The Oral Reading Study Design

Each student who participated in the oral reading study had already participated in the main NAEP reading assessment. Thus, the oral reading study sample of students was a subsample of students who participated in the main reading assessment. The special study was conducted in individual sessions between the participant and a trained administrator, who used a laptop computer to deliver instructions and record students' oral reading. After screening and additional preparatory activities, students were asked to read aloud an excerpt of one of the texts they had read as part of the main assessment. The following sections provide more detailed descriptions of the major components of the oral reading study design.

Sample Selection

The oral reading study design called for a nationally representative sample of fourth-graders with approximately 1,800 participants. These students were selected from among those taking certain booklets administered as part of the main NAEP 2002 reading assessment. The selection procedures for the oral reading study involved multistage, multiphase sampling of schools and students.

Sample Selection for Main NAEP 2002 Assessment

The grade 4 2002 main NAEP assessments tested public and private school students. Samples were selected based on a two-stage design: selection of schools and selection of students within schools. The first-stage sample of schools was selected with probability proportional to a measure of size based on estimated enrollment at grade 4. Each participat-

ing school provided a list of fourth-graders from which a systematic sample of students was drawn. Depending on the school's size, one or more sessions of 60 students were sampled. Half of the selected students were assigned a reading assessment booklet and the remainder a writing booklet.

The public and private school sample designs differed with respect to sample size requirements and stratification. For public schools, representative samples were drawn within each state and the District of Columbia, as well as from separate lists of Bureau of Indian Affairs (BIA) schools and Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS). Each sample was designed to produce aggregate estimates with approximately equal precision. The target sample in each state was 6,300 fourth-grade students. With a general target of 60 sampled students per school, roughly 100 participating schools were needed per state. Special procedures to reduce overall burden were used for states with many small schools, and for states having small numbers of grade-eligible schools.

Prior to sample selection, public schools were hierarchically stratified by district status (districts with more than 20 percent of their state's students were in a separate stratum), urbanization, and minority class. "Minority class" ensured representation of specific race/ethnicity groups. Within minority strata, schools were sorted by state achievement data for states where it was available. Where state achievement data were not available, schools were sorted by median household income of the zip code area where the school was located. Achievement data were supplied by the states themselves. Median income data were obtained from the 1990 Census. Other stratification variables were obtained from the National Center for Education Statistics' Common Core of Data (CCD), which is available online at http://www.nces.ed.gov/ccd.

For private schools, target student sample sizes were set for four separate reporting groups: Roman Catholic (6,000 students), Lutheran (1,500 students), Conservative Christian (1,500 students), and Other Private (3,000 students). Within these explicit strata, the private schools were implicitly stratified by census division, urbanization, and percent minority (Black/Hispanic/American Indian). Implicit strata were collapsed extensively to ensure that the expected number of schools within each implicit stratum was reasonably large.

Participation in state NAEP was not mandatory in 2002. Since the aggregate of the individual state samples was to be used as the public school sample for the national study, some provision needed to be made to ensure representation from a state even if that state declined to participate in state NAEP. Subsamples of schools were drawn from the state samples to use for the national sample under these circumstances. These subsamples were drawn for every state to cover all contingencies. They provided a suitable starting point for selecting the public school portion of the oral reading sample.

The process of drawing a national subsample for use in NAEP involved computing appropriate probabilities of school selection using a national target sample size assigned proportionally to each jurisdiction (as if no state NAEP samples had been drawn) and then dividing these probabilities by the full-sample and private-school NAEP probabilities to obtain conditional probabilities of selection for subsampling. School samples were drawn using the conditional probabilities. The resultant unconditional probabilities of selection for the subsample of schools are equal to the appropriate values for a stand-alone national sample. The goal was 35,500 assessed students at grade four.

Sample Selection for the Oral Reading Study

The target student sample size for the oral reading sample was a fraction of the fourth-graders targeted for the main NAEP reading study. Even though considerably fewer than 60 students were selected from each school for the oral reading study, further school subsampling was required. For efficiency, the nationally subsampled schools were grouped into 148 geographic clusters, each containing at least 5 eligible sampled schools. (A cluster could be an individual county, if it met the minimum size requirement, or 2 or more adjacent counties.) From the 567 counties with at least one eligible grade 4 school, 148 geographic clusters were defined and 91 were selected with probability proportional to the number of eligible schools. Five of the 91 were selected with certainty because each contained a large number of schools. In each of the remaining 86 sampled clusters, 5 schools were selected with equal probability. In the five certainty clusters, schools were also subsampled with equal probability, at a rate equal to the product of the cluster probability and within-cluster probability for noncertainty clusters.

The oral reading study design targeted assessed or absent students using any one of eight reading booklets. Since the booklets are assigned randomly, the set of students assigned these booklets constitutes a valid random sample of students capable of taking the NAEP assessment. In most schools, all such students were recruited to participate in the oral reading study. Usually, this produced a caseload of about seven students per school.

The sampling frame for public schools was the 1999–2000 Common Core of Data, and for nonpublic schools was the 1999-2000 Private School Survey. Out of 148 clusters created, 91 were selected—of which 5 were selected with certainty. Each selected school that participated in the study and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences between the student sample and the respective population from which they were drawn. Sampling weights account for disproportionate representation due to oversampling of students who attend schools with high concentrations of Black and/or Hispanic students. Among other uses, sampling weights also account for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.

Table A-1 provides a summary of the school and student participation rates for main NAEP and the oral reading study. Participation rates are presented for public and nonpublic schools, both individually and combined. Three different rates are presented. The first rate is a weighted percentage of schools participating in the study. This rate is based only on the schools that were initially selected for the study. The numerator of this rate is the estimated number of schools that participated in the study. The denominator is the estimated number of the initially selected schools.

The second school participation rate is a school-centered weighted percentage of schools participating in the assessment before substitution of demographically similar schools. This rate is based only on the schools that were initially selected for the study. The numerator of this rate is the estimated number of schools represented by the initially selected schools that participated in the study. The denominator is the estimated number of schools represented by the initially selected schools that had eligible students enrolled.

The third school participation rate is a school-centered weighted participation rate after substitution. The numerator is the estimated number of schools represented by the participating schools, whether originally selected or selected as a substitute for a school that did not participate. The denominator is the estimated number of schools represented by the initially selected schools that had eligible students enrolled.

If school participation is associated with the size of the school, the student-centered and school-centered school participation rates will differ. In the oral reading study the student-centered rate is higher than the school-centered rate. This indicates that larger schools participated at a higher rate than smaller schools.

Also presented in table A-1 are weighted student participation rates. The numerator of this rate is the estimated number of students who are represented by the students assessed (in either an initial session or a makeup session). The denominator of this rate is the estimated number of students represented by the eligible sampled students in participating schools.

A nonresponse analysis was conducted because school and student response rates did not meet NCES statistical standard 3-2-5 concerning achieving desirable response rates. The rates are currently set at 85 percent for NAEP. When the rates are between 70 and 85 percent, an extensive analysis is conducted that examines, among other factors, the potential for nonresponse bias at both the school and student levels. A nonresponse bias analysis was completed by computing weighted response rates for various school- and student-level characteristics of interest and by conducting chi-square tests. The school nonresponse investigated in these analyses is cumulative nonresponse to both NAEP and the study. At the school level, the characteristics of interest are census region, public or private affiliation, reporting subgroup for private schools, type of location, estimated grade enrollment, and percentage of students in each racial/ethnic category. At the student level, the characteristics of interest are the students' gender, race/ethnicity, relative age, year of birth, free lunch eligibility, whether or not the student receives Title I assistance, and whether or not the student is classified as a student with disability or a limited-English-proficient student. The only variables not significant in the oral reading study are type of location at the school level and gender and year of birth at the student level. All other variables show a differential rate of nonresponse between subgroups. The final rates were adjusted as a result of nonresponse. The observed sample is classified in demographic groups, and students sampling weights

Table A-1. Grade 4 oral reading study school and student participation rates, by type of school: 2002

School type	Overall participation rate		Weighted school participation				Weighted student participation	
	Before substitution	After substitution	Oral Reading conditional on main NAEP participation	Percent before substitution	Percent after substitution	Number of schools participating after substitution	Weighted percent student participation	Number of students assessed
Main NAEP								
Combined national	79	80	†	84	85	5,518	94	140,487
Public	80	80	†	85	85	5,067	94	133,805
Nonpublic	71	77	†	74	81	451	95	5,578
Oral Reading								
Combined national	66	66	85	78	79	360	84	1,779
Public	66	66	85	79	79	309	83	1,613
Nonpublic	61	68	81	68	76	51	90	166

† Not applicable.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

are adjusted to ascertain that the representation of these groups in the sample matches the sampling frame. More details can be found in the section on Weighting and Variance Estimation.

With respect to school nonresponse, schools in the South seem to have a substantially lower nonresponse compared to schools in other regions. Also, nonpublic schools seem to have a larger nonresponse than public schools. With respect to student nonresponse, students in private schools seem to show a lower nonresponse than students in public schools. Black and Hispanic students appear to show a higher nonresponse than White students. Students older than the modal age seem to show higher nonresponse than students at or below the modal age. Students who are eligible for free or reduced-price lunch appear to show a higher nonresponse than either students who are not eligible or students for whom no information is available. Students who receive Title I assistance appear to have a higher nonresponse than students who do not receive Title I. Also, students identified as limited-English-proficient and those identified with disabilities appear to show a higher nonresponse than students who are not identified as belonging to either of these categories.

Materials

The oral reading study used a single literary text that had been one of the easiest fourth-grade passages in the main NAEP reading assessment to assess reading for literary experience. The story "The Box in the Barn," which is presented in its entirety in appendix B, was selected because it presented a narrative structure and main character with which most fourth-graders would be at least somewhat familiar. In addition, it was considered to be among the least difficult passages in the 2002 fourth-grade NAEP reading assessment. A 198-word excerpt of the text was selected from the passage for the oral reading task. This particular portion was selected because it represented a scene from the story with a natural beginning and ending—giving the oral reading task a sense of coherence and completeness. The excerpt that students were asked to read aloud is shaded in the text of the entire story presented in appendix B.

Data Collection

The oral reading study was administered by trained interviewers who followed procedural directions in a three-step interview guide. In the first step, interviewers showed students a copy of Williamson's (2001) "A Snail's Tail," used as a screening passage, and asked them to read it aloud. The purpose of the

introductory passage was to ascertain the ease with which the student handled the relatively easy text, which was considered to be at approximately a second-grade level. If the supervisor decided that the student read with much difficulty (i.e., the student produced a word-by-word delivery, halted between words, was unable to recognize many of the words, took more than six minutes to read the passage, or reached a point of frustration and gave up), the interviewer was instructed to stop the student at the end of a sentence by saying, "That's the end of the test. Thank you for helping us today." The interviewer then gave the student a small thank-you gift and walked him or her back to the classroom. If, however, the supervisor decided that the student's performance was above the acceptable threshold, the student was cleared for the next step of the procedure.

Step two in the interview involved showing the student a copy of "The Box in the Barn" and asking him/her to read it silently. When the student finished with the silent reading, the interviewer showed him/her three questions about the passage. The students looked at the questions, one at a time, while simultaneously hearing a recording of each question over a headset. The following are the three questions the students heard:

1) Why did Jason think everyone would be angry with him when they found the puppy missing?
2) Describe how Jason might have felt if the box had been empty when it was opened at the party. Explain why he might have felt that way.
3) From when Jason got up in the morning until he went to bed that night, his feelings changed as different things happened. Describe three different feelings that Jason had and explain what made him have those feelings.

The interviewers were instructed to allow students time to think about their answers before prompting them for their answers. After the student replied, the interviewer asked, "Is there anything else?" If the student didn't answer or said "I don't know," the interviewer offered him/her a chance to go back to the story. If he/she still did not answer, the interviewer moved the student on to the next question. After the student heard and had ample opportunity to respond to each question, the interviewer moved to the third portion of the oral reading study.

The students' answers to the taped comprehension questions were not rated for assessment purposes. The questions were asked in order to support the subsequent oral reading activity by affording students an opportunity to refresh their understanding of the story's plot and characterization.

In the third step of the interview, the student was asked to read aloud the 198-word passage from "The Box in the Barn." If the student inquired about how fast or slow to read, the interviewer said "Read the story out loud as if you're reading it to someone who's never heard it before." If the student seemed to be having difficulty, the interviewer urged, "Just do the best you can." If the student asked for help in pronouncing a word, the interviewer responded, "If you can't figure out a word, you can guess or skip it, and go on."

On average, students spent about 1 minute and 40 seconds reading the excerpt from "The Box in the Barn." The session ended when the student had finished his or her oral reading of the excerpt, and the interviewer took back the copy of the story and said, "That's the end of the test. Thank you for helping us today. We have a thank-you gift for you." After receiving the gift, the student was escorted back to his/her classroom. The total time required to administer the oral reading assessment to each student was less than 30 minutes.

Scoring and Analysis of Students' Oral Reading

The analysis of students' oral reading was conducted by a team of 20 experienced educators trained as coders. Eighteen of the coders were currently employed elementary or middle-school language arts teachers, one was a speech therapist for elementary and middle-school children, and one was a retired reading specialist who maintained a private tutorial practice. The coders were native speakers of American English of various ethnic backgrounds, including four African Americans. All of the teachers were employed in either inner-city schools or suburban schools in the vicinity of major cities, where they had taught students from diverse ethnic and linguistic backgrounds.

The use of digital recordings facilitated reliable and valid scoring procedures. The coders worked at individual computer terminals equipped with headsets through which they listened to individual students' oral reading. The scoring system was designed to distribute student recordings randomly to individual coders so that scoring was distributed evenly across coders. Each coder entered the coding into an individual computer terminal. Two trainers were able to monitor the coding continuously and to

check for acceptable interrater reliabilities. At least 25 percent of the recordings, selected randomly, were recoded by a second individual from the team of 20 coders.

The three aspects of oral reading—accuracy, rate, and fluency—were analyzed separately in different coding series. In each series, coders were first introduced to the construct and provided instruction on the specific criteria to be applied. For instance, coders were trained not to categorize variant pronunciations (such as those arising from regional, dialectical, or nonnative speech) as reading errors, unless the mispronunciation altered sentence meaning. After each practice session, trainers reviewed the practice codings and discussed any errors with the entire team. A second practice session comprising more challenging coding decisions was then conducted. Practice sessions and follow-up reviews lasted about 7.5 hours. After additional discussion of the second practice set, individual coders were examined in a qualification coding session to determine if they understood and correctly applied the coding criteria. Finally, coders listened in pairs and worked together to discuss attributes of the students' oral reading and the application of coding criteria. Periodically, the trainers interrupted the coders' work to reinforce their understanding of coding criteria in mini-training sessions. The mini-training and paired-scoring sessions occurred recursively across each coding series and throughout the entire coding period, which lasted four weeks.

Training the Coders

Preparations for coding began about one month before the team of coders started their work. When Ordinate delivered the fourth-graders' oral reading files to ETS, three NAEP administrators reviewed the digitized oral reading samples and chose anchor recordings (four benchmark samples for each of the four fluency levels) and training samples. Other training materials included guidelines for marking fluency and a notation system developed by NAEP administrators for marking accuracy determinations.

The training materials included background information on NAEP's mission in general and on the 1992 IRPR study and the 2002 oral reading assessment in particular, typescripts of "The Box in the Barn" and "The Hungry Spider" passages, coding guidelines, including descriptions of the four fluency levels, and guidelines for marking and recording errors.

During the first three days of fluency coding, the coders became familiar with the fluency guide and listened to anchor recordings for each fluency level. After extensive discussion and review of the anchors, coders were quizzed on their ability to assign appropriate fluency scores. They listened to and rated two practice sets (with 17 examples in each set) of students reading the excerpt from "The Box in the Barn." In the online scoring lab where coding activities were conducted, four computers were available for coders to periodically review the anchors to prevent coder drift.

Live coding began during the third day of fluency training. The reliability interface developed by Ordinate allowed the two on-site administrators to check on the coding judgments of each of the 20 coders. When coders were found to be making off-the-mark decisions, they were asked to listen to the samples they had misrated and referred to the anchor sets for standards centering.

Accuracy training for "The Box in the Barn" took three days. The materials used for accuracy training included guidelines showing sample reading errors and a notation system for recording the errors. Coders were also given instructions on how to enter judgments about whether the reading errors impacted sentence meaning in an online interface developed by Ordinate for electronically tracking and compiling the oral reading data. This accuracy interface allowed administrators to monitor for reliability by noting instances where paired scores showed different numbers of words read correctly. When discrepant coding was noted, coders were directed to review the specific accuracy guidelines accounting for their differing correct word counts and to review the accuracy anchor samples.

Fluency

Fluency coding was based on the fluency scale presented and described in chapter 3 of this report. This fluency scale was developed for use in the 1992 study, and was used again in 2002 without any modification. To facilitate training of coders and to anchor the four fluency levels, exemplar digital recordings of individual students' oral reading were selected from the oral reading study sample. For each of the four fluency levels, 10 to 15 recordings of different students were identified as exemplary of fluency at the given level.

The actual fluency coding, after training and qualification were completed, was accomplished by

Table A-2. Interrater reliability rates for coders' fluency decisions, grade 4: 2002

Variable pair	Reliability rates
Percent exact agreement between first and second coder	81
Percent adjacent agreement between first and second coder	100
Percent classification agreement between first and second coder	92
Intraclass correlation	.82

NOTE: Of a total 1779 recordings, at least 25 percent (546) were second-coded.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

having coders listen to each recording twice. The first exposure allowed coders to become familiar with the student's unique vocal characteristics. While listening to the recording a second time, coders made their decisions regarding the level of fluency exhibited by the student's oral reading. This was a holistic coding of fluency, which considered the overall performance of the student and was guided by the specific descriptions provided in the fluency scale, as well as the exemplar (or anchor) recordings at each fluency level.

To determine whether fluency ratings were reliable, 546 of the oral readings were randomly chosen and recoded by a second coder. The reliability rates for the coders' fluency decisions are shown in table A-2. For the fluency rating of 546 oral readings, coders achieved an exact agreement rate of 81 percent and an adjacent agreement of 100 percent. The percent agreement on classification, coding either nonfluent (fluency level 1 or 2) or fluent (fluency level 3 or 4), was 92 percent.

Accuracy

Accuracy coding required a multistep process for each student in the oral reading study sample. First, coders listened without interruption to the individual recordings of students' oral reading. As with the fluency coding, this first exposure to the student's reading allowed the coder to become familiar with the student's unique vocal characteristics. Because specific determinations had to be made regarding students' production of words, coders were specially trained to be sensitive to regional and ethnic variations in oral reading style and individual word pronunciation.

After the coder's initial exposure to the student's oral reading, the coder listened once more from the beginning of the recording for the purpose of documenting the total number of words the student read correctly and the number of oral reading errors. Notations were made on a typescript of the oral reading passage, and were subsequently recorded in a coding template. Each word of the oral reading passage was numbered, and for each word read in error, coders entered the corresponding numeral into the template, followed by specific notations that described the specific type of error. Each error was counted, whether or not it was subsequently self-corrected by the student. However, successful self-corrections were documented and become part of the database for that individual student.

Three specific types of errors were documented in the accuracy coding—substitutions, omissions, and insertions. The following general guidelines were followed in determining when a student's production of a word or group of words represented an error. For each type of error, more detailed explanations and examples were included in the training instructions and the coding guidelines to facilitate agreement among coders and to ensure that each coder knew how to code each type of error.

Substitution—Only whole-word substitutions were noted. Partial attempts were not considered substitutions. The addition or deletion of prefixes or suffixes to text words was considered a whole-word substitution for the text word. When one or more words were substituted for a group of contiguous text words, the substitution was recorded as one complex substitution, but each of the words in the contiguous grouping of text words was counted as an error. For instance, if a student pronounced, "Jason was the boy," where the text said "Jason saw the box," the last three words in the student's sentence were coded as errors.

For each substitution, an additional coding decision was made as to the grapho-phonemic correspondence between the text word and the word substituted by the student. Subsequent to the coding and analyses of students' oral reading, however, it was determined that the overall accuracy demonstrated by students did not permit a meaningful detailed analysis of a single type of error—

substitutions. Consequently, results based on the grapho-phonemic coding procedure are not presented in this report.

Omissions—Only whole-word omissions were coded as omissions (e.g., "he wished" for "he secretly wished"). Partial word omissions were coded as substitutions (e.g., "in" for "inside"). If contiguous words were omitted, each word in the set was counted as an error (e.g., Dad wouldn't be back for two hours" instead of "Dad wouldn't be back for at least two hours" would count as two omissions).

Insertions—Only whole-word insertions were coded as insertions (e.g., "wanting his very own puppy" for "wanting his own puppy"). Inserts of prefixes or suffixes were coded as substitutions (e.g., "return" for "turn," or "unhappy" for "happy").

For each of the three types of error described above, two additional coding decisions and notations were documented—self-correction and meaning change. The general guidelines for these coding decisions are provided below.

Self-correction—If after making one of the three errors, students subsequently reread the word or words correctly, it was documented as a self-corrected error. The entire word or groups of words had to have been read correctly after the initial error for it to be counted as a self-correction.

Meaning change—Whether or not the original error had been self-corrected, the coders determined if the initial error resulted in a change of meaning to the text. This decision was based on a very general consideration of the main idea of the sentence and its relation to the entire passage. It was not based simply on grammatical considerations.

Table A-3. Interrater reliability rates for coding of all errors, meaning change, and self-correction, grade 4: 2002

Reliability metric	Errors	Meaning change	Self-correction
Percent exact agreement	98	99	99
Cohen's kappa	.71	.44	.58

NOTE: Of a total 1779 recordings, at least 25 percent (460) were second-coded. Cohen's kappa penalizes for "easiness of classification" (i.e., for an overwhelming agreement on or yes or no). Therefore, this statistic may not be highly informative for word-by-word scoring.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

After the appropriate notations were made on the typescript, each observed error was recorded in the scoring system's coding template. The numeral of the text word in which the error had occurred was entered into the template, followed by a set of notations documenting the type of error, whether or not it was self-corrected, whether or not it resulted in meaning change, and, for substitution errors, the degree of phonemic correspondence.

The interrater reliability rates of the coders' decisions for all types of errors are shown in table A-3. For the 460 oral readings that were second-coded for accuracy, coders achieved an average of 98 percent exact agreement. In making decisions on whether errors changed meaning, coders achieved an exact agreement of 99 percent. The exact agreement rate was 99 percent for decisions on whether errors were self-corrected. The high exact-agreement percentages in the oral reading study result from the low error rate; the majority of students did not make errors in reading the majority of words in the script. Thus, the accuracy-coding task was relatively uncomplicated. This lack of complexity is reflected in Cohen's kappa. A low kappa generally indicates a straightforward scoring task.

Rate

The rate, or speed, of students' oral reading was calculated electronically by the scoring system developed for the study. Rate was calculated in two different ways. First, the total number of words read by students across the entire passage was divided by the amount of time it took for students to complete their reading. In most cases, the total number of words equaled 198—the total number of words in the oral reading passage. As a second calculation of oral reading rate, the number of words read in the first minute of a student's oral reading was calculated.

Main Reading Assessment Analysis

The average reading scores presented in this report were derived from analyses of the students' performance on the 2002 main reading assessment. A detailed description of the analysis procedures is provided in the main report of the 2002 reading results (Grigg et al. 2003). A scale ranging from 0 to 500 was created to report performance for each reading purpose assessed at grade 4—literary and information. The scales summarize student performance across all three types of questions in the assessment (multiple-choice, short constructed-response, and extended constructed-response). The

composite scale (used in this report for reporting student performance on the main reading assessment) is a weighted average of the two separate scales for each reading purpose. The weight for each reading purpose is proportional to the relative importance assigned to the reading purpose by the specifications developed through the consensus planning process and given in the framework.

Item Response Theory (IRT) was used to estimate average reading scores for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity, even when students receive different sets of items collected as part of the main assessment. (The main NAEP reading assessment uses matrix sampling procedures in which representative samples of students take various portions of the entire pool of assessment questions. Individual students are required to take only a small portion, but the aggregate results across the entire assessment allow for broad reporting of reading abilities for the targeted population.) One desirable feature of IRT is that it locates items and students on this common scale. In contrast to classical test theory, IRT does not rely solely on the total number of correct item responses, but uses the particular patterns of student responses to items in determining the student's location on the scale.

Weighting and Variance Estimation

A complex sampling design was used to select the students who were assessed. Sampling for the study was conditional on selection to receive certain booklets in the main NAEP reading sample. The properties of a sample selected through such a design could be very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data. One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. These weights included adjustments for school and student nonresponse.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student reading ability: the uncertainty due to sampling only a relatively small number of students, and the uncertainty due to sampling only a portion of the cognitive domain of interest. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

NAEP uses a jackknife replication procedure to estimate uncertainty due to sampling only a relatively small number of students. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. An example of such information is gender or race/ethnicity, but also information collected about students' oral reading ability, such as accuracy and reading rate. To estimate uncertainty associated with estimates of reading proficiency, from the main NAEP reading assessment, a component based on NAEP's marginal estimation methodology and imputation is added to the jackknife component.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources—inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate and, because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

Drawing Inferences From the Results

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

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

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

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

$$256 \pm 1.96 \times 1.2$$

$$256 \pm 2.4$$

$$(253.6, 258.4)$$

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

Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether the evidence, based on the data from the groups in the sample, is enough to conclude that the averages or percentages are actually different for those groups in the population. If the difference is statistically significant, the report describes the group averages or percentages as being different (e.g., one group performed higher or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population. NAEP only reports subgroup results if the students within a subgroup are adequately distributed across primary sampling units (PSU) to allow for reasonably accurate estimation of standard errors. NAEP only publishes those statistics that have standard error estimates based on five or more degrees of freedom. In other words, subgroup members must be represented across at least five PSUs to allow for reasonably accurate standard errors. NAEP limits comparisons involving extreme percentages. When percentages are close to 0 or 100, their distributions differ greatly from *t*- or *z*-distributions. For this reason, hypothesis tests of the sort used by NAEP are not appropriate in these cases. Under these conditions, no test is made. NAEP warns the user to interpret estimates with caution if the coefficient of variation of the denominator of a ratio estimate, such as a weighted percent or a weighted scale score average, is more than 20 percent. In that case the standard errors associated with the ratio estimate could be severely biased or unstable, and the symbol "!" is placed next to the standard error.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample is obtained.

This estimate of the degree of uncertainty, called the "standard error of the difference" between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

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

The standard error of the difference can be used, just as the standard error for an individual group average or percentage, to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the .05 level.

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

Group	Average Scale Score	Standard Error
A	218	0.9
B	216	1.1

The difference between the estimates of the average scale scores of groups A and B is two points (218 – 216). The estimated standard error of this difference is

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

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

$$2 \pm 1.96 \times 1.4$$

$$2 \pm 2.7$$

$$(-0.7, 4.7)$$

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

This procedure is appropriate to use when it is reasonable to assume that the groups being compared have been independently sampled for the assessment. This is the approach used for NAEP reports when comparisons involving independent groups are made. The assumption of independence is violated to some degree when comparing group results (e.g., comparing results for males and females), since these samples of students have been drawn from the same schools. When the groups being compared do not share students (as is the case, for example, comparing males and females), the impact of this violation of the independence assumption on the outcome of the statistical tests is assumed to be small, and NAEP, by convention, has, for computational convenience, routinely applied the procedures already described to those cases as well.

When making comparisons of results for groups that share a considerable proportion of students in common, it is not appropriate to ignore such dependencies. In such cases, NAEP has used procedures appropriate to comparing dependent groups. When the dependence in group results is due to the overlap in samples (e.g., when a subgroup is being compared to a total group), a simple modification of the usual standard error of the difference formula can be used. The formula for such cases is

$$SE_{Total-Subgroup} = \sqrt{(SE^2_{Total} + SE^2_{Subgroup} - 2pSE^2_{Subgroup})}$$

where p is the proportion of the total group contained in the subgroup. This is a special form of the common formula for standard error of dependent samples. The standard formula can be found, for example, in Kish (1995).

Conducting Multiple Tests

The procedures in the previous section and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, there are times when many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., .05), adjustments (called "multiple comparison procedures"; Miller 1981) must be made to the

methods described in the previous section. One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure was used to control the certainty level (Benjamani and Hochberg 1995).

Unlike the other multiple comparison procedures that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons. Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures (Williams, Jones, and Tukey 1999). A detailed description of the FDR procedure will appear in the technical documentation section of the NAEP website at http://nces.ed.gov/nationsreportcard.

To illustrate how the FDR procedure is used, consider the comparisons of the average reading scale scores for three groups presented in table A-4. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference. (Rounding of the data occurs after the test is done.)

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

If only two groups were compared, there would be a significant difference between the average scale scores if the significance level were less than 5 percent. However, because the difference in average scale scores for all three groups is of interest, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons.

Table A-4. Example of False Discovery Rate comparisons of average reading scale scores for different groups of students

Students	Average scale score	Standard error
Group 1	201	2.6
Group 2	210	1.8
Group 3	204	2.1

	Difference	Differences in averages	Standard error of difference	Percent confidence[1]
Group 1–Group 2	8.5	3.16	2.69	1
Group 1–Group 3	2.7	3.34	0.81	42
Group 2–Group 3	5.8	2.77	2.10	4

[1]The percent confidence is 2(1-F(x)) where F(x) is the cumulative distribution of the *t*-distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Using the FDR procedure to take into account that all comparisons are of interest, the percents of confidence in the example are ordered from largest to smallest: 42, 4, and 1. In the FDR procedure, 42 percent confidence for the group 1 to 3 comparison would be compared to 5 percent, 4 percent for the group 2 to 3 comparison would be compared to .05 × (3 – 1)/3 = .033 = 3 percent, and 1 percent for the group 1 to 2 comparison would be compared to .05 × (3 – 2)/3 = .017 = 2 percent. The procedure stops with the first contrast found to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference between the average scale scores of group 1 and group 2 students is significant; for all of the other groups, average scale scores are not significantly different from one another.

NAEP Reporting Groups

Results are provided in this report for groups of students defined by shared characteristics—gender and race/ethnicity. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs). In 2002, the first-stage sampling units were schools (public and nonpublic) in the selection of the combined sample. Further details about the procedure for determining minimum sample size appear in the technical documentation of the NAEP website at http://nces.ed.gov/nationsreportcard/. However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the two subpopulations for which results are presented in this report are presented below.

Gender

Results are reported separately for male students and female students based on school records.

Race/Ethnicity

In all NAEP assessments, data about student race/ethnicity are collected from two sources: school records and student self-reports. Previously, NAEP has used student self-reported race as the primary race/ethnicity reporting variable. In 2002, it was decided to change the student race/ethnicity variable highlighted in NAEP reports. Starting in 2002, school-recorded race became the race/ethnicity variable presented in NAEP reports. The mutually exclusive racial/ethnic categories were White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other.

Cautions in Interpretations

As previously stated, the NAEP reading scale makes it possible to examine relationships between students' performance and various background factors measured by NAEP. However, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not reflect the influence of unmeasured variables. The results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

A caution is also warranted for some small population group estimates. The effects of exclusion-rate changes for small subgroups may be more marked for small groups than they are for the whole population. The standard errors are often quite large around the score estimates for small groups, which in turn means the standard error around the gain is also large.

Appendix B. Oral Reading Passage and Questions to Prompt Students' Recall of the Passage

This appendix gives the complete text of "The Box in the Barn," the story used in the main NAEP 2002 reading assessment and the oral reading study. Participants received the whole text to read silently. They were then asked three questions, also included here, to prompt their recall of the story. Digital recordings of the students reading orally from the 198-word shaded portion of the story provided the raw data for the study results.

The Box in the Barn

By Barbara Eckfeld Conner

Jason heard his mom calling him. Instead of answering her, he slipped deeper into the tall weeds behind his house. He closed his eyes, thinking of what he had done.

He had gotten up that morning in a good mood. Raspberry pancakes were on the table when he walked into the kitchen rubbing his eyes and yawning.

"After breakfast, Jason, I want you to go into town with me," Mom said quietly. "It's your sister's birthday, and we need to shop for her gifts."

Jason was eager to go, even if the gifts weren't for him. Buying presents was always fun.

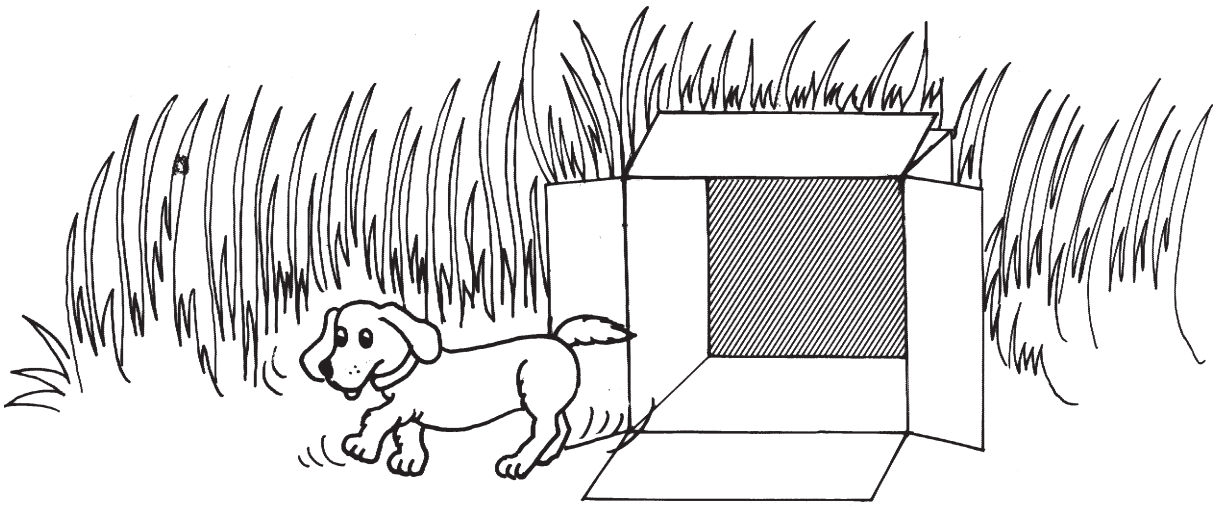
As they drove to town, Jason couldn't help but ask the question that had been on his mind since yesterday when Aunt Nancy came. "What's in the big box that Dad took to the barn, Mom? Is it something Aunt Nancy bought for Megan's birthday?"

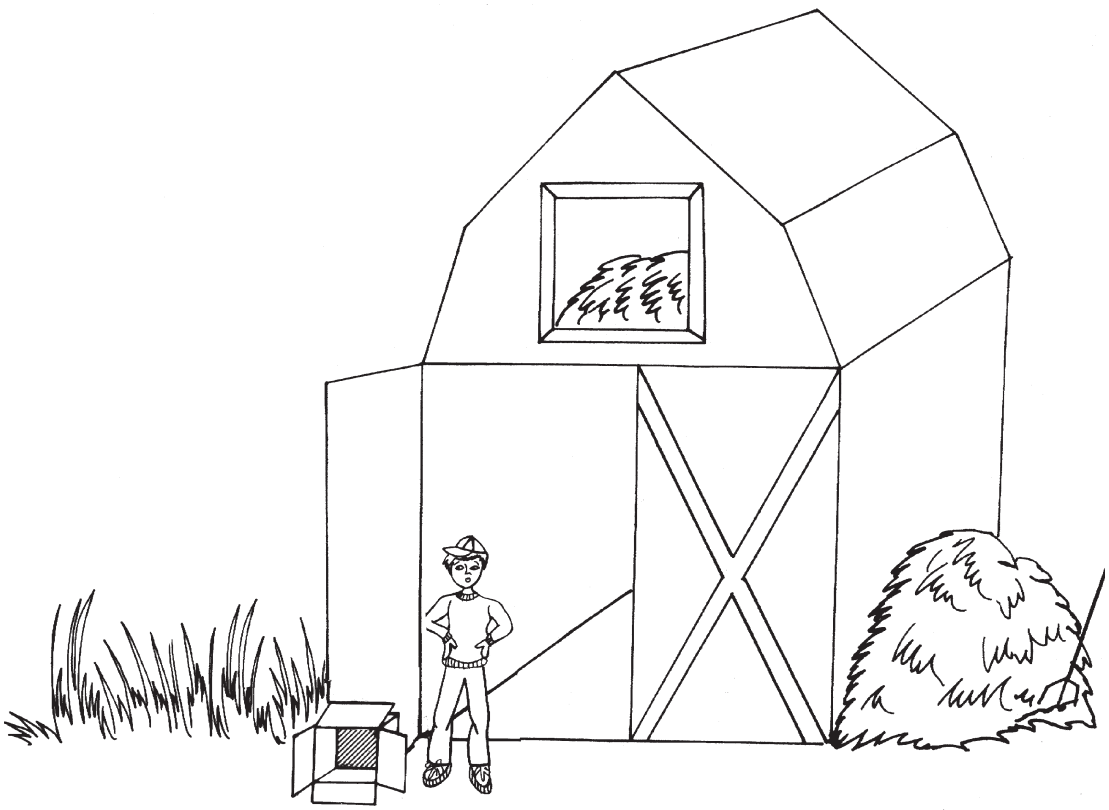
"It's a surprise, Jason, and I don't want you going near that barn today. Do you hear me?"

Jason sat staring at the road ahead. He knew that nothing would change her mind. Only now he was more curious than ever!

Back home, Megan ran out to meet Jason, her eyes wide and excited. "Jason, Jason, I'm six years old!" she cried, jumping up and down.

"I know, I know." Jason gave her a big hug.





Soon the house was buzzing with excitement. Megan sat on the stool watching while Mom and Aunt Nancy prepared the birthday dinner. Dad wouldn't be back for at least two hours. Jason wandered outside trying to think of something to do, but his thoughts kept returning to the box in the barn.

He started walking toward the barn, not at all sure what he'd do when he got there. He was hoping for just a glimpse of the box. Instead he heard a strange noise coming from inside the barn. He wished he could just turn back to the house, but his legs carried him into the barn. Jason saw the box. It was sitting between two bales of hay. He could hear loud wailing cries. Leaning over, Jason carefully lifted the lid. There was the most cuddly puppy he had ever seen!

"You must be pretty scared, huh, fellow?" Jason said quietly as he held the wiggly dog. "Megan's going to love you!" He secretly wished the puppy was for him. After all, Mom and Dad knew that he had been wanting his own puppy. Probably Aunt Nancy didn't know that, and anyway Megan would be happy.

The shaded text above is the 198-word portion that students in the 2002 oral reading study read aloud.

Soon Jason was playing happily with the puppy, and he forgot that he wasn't supposed to be in the barn. Taffy, their big brown horse, stuck his head in the window as if to say, "What's going on?" Jason jumped, remembering that he wasn't supposed to be there. The puppy ran off as fast as it could out of the barn and into the field.

Jason stumbled out of the barn looking wildly for any trace of the puppy. "Come on puppy! Oh, please come here!" he called, his eyes welling up with tears.

Now here he was, two hours later, hiding in the weeds. He'd looked everywhere, but the puppy was gone. He had ruined his sister's birthday.

"Jason! It's time for dinner!" Mom called even louder now. Just when he was determined to stay forever in the tall weeds, he heard his sister's voice.

"Jason! It's time for my party, Jason!" Megan yelled excitedly.

Jason rubbed his swollen eyes, trying to look normal. He couldn't ruin everything for her. "I'm here, Megan," he called.

"Are you OK?" she asked with genuine concern.

"Sure. Let's hurry." Jason grabbed her hand as they ran back.

As soon as they reached the house, the party began. Jason tried to pretend that everything was fine. When it was time to open Megan's birthday gifts, he sat in the big easy chair, hoping no one would notice him. Finally the last present was open.

"I'll be right back," Dad said.

Jason knew Dad was going to the barn. Megan would probably never forgive him for losing her birthday puppy. Everyone, even Aunt Nancy, would be angry when they found out the puppy was gone.

"Jason! Come here!" It was Dad calling from the front yard.

Jason slowly got out of the chair. It was hard to move, but Megan grabbed his hand and said, "Come on, Jason! Let's see what Dad wants."

Jason followed Megan out the door. Mom and Aunt Nancy followed close behind.

There was Dad standing with the box next to him in the grass. "Jason, I want you to open this box and see what's inside."

Jason looked up and saw that Dad was smiling. He turned and saw that Mom, Aunt Nancy, and Megan were smiling, too. What would he say to them when there was nothing in the box? But as Jason looked down, expecting to see nothing at all, he jumped back in surprise. The puppy looked up at him, with sleepy eyes.

"Wow!" said Jason, bewildered.

"The puppy's for you, Son," his father said.

"I thought you'd like a gift, too, even if it isn't your birthday," said Aunt Nancy, laughing.

Megan started clapping. "Isn't he wonderful, Jason?" The puppy jumped up, ready to play. Jason and Megan spent the rest of the day with the puppy.

Later, when he was getting ready for bed, Jason turned to his father and said, "You know, Dad, I feel bad about something I did today."

Dad waited patiently as Jason explained what had happened. "And I still can't figure out how my puppy got back into his box!" he added.

"Well, Son, on my way home I saw your puppy running along the side of the road. I figured he had gotten out of his box somehow.... You must have felt terrible during the party," Dad continued. "I get the feeling you've learned a lot today." He pulled back the covers on Jason's bed.

Jason looked down at his new puppy, who was sleeping soundly in a basket by the bed. "Dad, I think I'll call him Buddy."

Dad smiled and tucked the covers snugly around Jason.

Questions to Prompt Students' Recall of the Passage

When the student finished reading "The Box in the Barn" silently, the interviewer showed him/her three questions about the passage. The students looked at the questions, one at a time, simultaneously hearing a recording of each question over a headset. The following are the three questions the students heard:

1) Why did Jason think everyone would be angry with him when they found the puppy missing?
2) Describe how Jason might have felt if the box had been empty when it was opened at the party. Explain why he might have felt that way.
3) From when Jason got up in the morning until he went to bed that night, his feelings changed as different things happened. Describe three different feelings that Jason had and explain what made him have those feelings.

Appendix C. Standard Errors

The comparisons presented in this report are based on statistical tests that consider the magnitude of the differences between group averages or percentages and the standard errors of those statistics. This appendix contains the standard errors for the estimated averages and percentages in all the figures throughout this report. Because the oral reading study scores and percentages are based on samples rather than the entire population(s), the results are subject to a measure of uncertainty reflected in the standard errors of the estimates. 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.

Table C-1. Data for figure 2-1: Percentage of students, by degree of reading accuracy, grade 4: 2002

Number of errors [Percent of words read accurately]	Percent of students
0–4 [100–98]	35 (1.4)
5–10 [97–95]	40 (1.4)
11–20 [94–90]	19 (1.0)
21 or more [less than 90]	6 (0.7)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-2. Data for figure 2-2: Average reading scale scores, by degree of reading accuracy, grade 4: 2002

Number of errors [Percent of words read accurately]	Average scale scores
0–4 [100–98]	237 (1.4)
5–10 [97–95]	226 (1.3)
11–20 [94–90]	206 (2.2)
21 or more [less than 90]	180 (3.8)

NOTE: Standard errors of the estimated scale scores appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-3. Data for figure 2-3: Percentage of students, by gender and degree of reading accuracy, grade 4: 2002

Number of errors [Percent of words read accurately]	Percent of students	
	Male	Female
0–4 [100–98]	32 (1.7)	37 (1.9)
5–10 [97–95]	39 (1.6)	40 (1.8)
11–20 [94–90]	21 (1.4)	17 (1.3)
21 or more [less than 90]	7 (1.0)	6 (‡)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated scale scores appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-4. Data for figure 2-4: Percentage of students, by race/ethnicity and degree of reading accuracy, grade 4: 2002

Number of errors [Percent of words read accurately]	Percent of students		
	White	Black	Hispanic
0–4 [100–98]	38 (1.7)	23 (2.6)	31 (4.3)
5–10 [97–95]	40 (1.8)	40 (2.1)	35 (3.1)
11–20 [94–90]	17 (1.3)	28 (2.3)	21 (‡)
21 or more [less than 90]	5 (‡)	9 (‡)	12 (‡)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated scale scores appear in parentheses. Detail may not sum to totals because of rounding. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-5. Data for figure 2-5: Percentage of students, by degree of reading accuracy when only counting meaning-change errors, grade 4: 2002

Number of meaning-change errors [Percent of words read without meaning change]	Percent of students
0–4 [100–98]	76 (1.2)
5–10 [97–95]	17 (0.7)
11–20 [94–90]	5 (0.8)
21 or more [less than 90]	2 (0.3)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-6. Data for figure 2-6: Average reading scale scores in relation to the achievement levels, by degree of reading accuracy when only counting meaning-change errors, grade 4: 2002

Number of meaning-change errors [Percent of words read without meaning change]	Average scale scores
0–4 [100–98]	231 (1.1)
5–10 [97–95]	206 (2.3)
11–20 [94–90]	180 (3.0)
21 or more [less than 90]	‡

‡ Reporting standards not met. Sample size was insufficient to permit a reliable estimate for students with 21 or more errors that resulted in a change of meaning.
NOTE: Standard errors of the estimated scale scores appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-7. Data for figure 2-7a: Percentage of students, by percentage of all errors that were self-corrected, grade 4: 2002

Percent of all errors that were self-corrected	Percent of students
Less than 25	39 (1.4)
25–49	29 (1.1)
50–75	24 (1.3)
76 or more	8 (0.7)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-8. Data for figure 2-7b: Percentage of students, by percentage of meaning-change errors that were self-corrected, grade 4: 2002

Percent of meaning-change errors that were self-corrected	Percent of students
Less than 25	34 (1.6)
25–49	17 (1.1)
50–75	26 (1.3)
76 or more	23 (1.2)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-9. Data for figure 2-7c: Percentage of students, by percentage of non-meaning-change errors that were self-corrected, grade 4: 2002

Percent of non-meaning-change errors that were self-corrected	Percent of students
Less than 25	54 (1.5)
25–49	20 (1.0)
50–75	17 (1.0)
76 or more	9 (0.8)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-10. Data for figure 2-8: Average reading scale scores in relation to the achievement levels, by the type and percentage of errors that were self-corrected, grade 4: 2002

	Average scale scores		
Percent of errors that were self-corrected	All errors	Meaning-change errors	Non-meaning-change errors
Less than 25	216 (1.6)	212 (1.6)	218 (1.5)
25–49	221 (1.7)	208 (2.9)	221 (2.2)
50–75	230 (1.8)	221 (1.9)	232 (2.3)
76 or more	237 (2.9)	233 (1.7)	229 (3.6)

NOTE: Standard errors of the estimated scale scores appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-11. Data for figure 3-1: Percentage of students, by average number of words read per minute, grade 4: 2002

Average number of words read per minute	Percent of students
Less than 80	14 (1.0)
80–104	21 (1.1)
105–129	27 (1.3)
130 or more	38 (1.5)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-12. Data for figure 3-2: Average reading scale scores in relation to the achievement levels, by average number of words read per minute, grade 4: 2002

Average number of words read per minute	Average scale scores
Less than 80	185 (1.9)
80–104	207 (1.9)
105–129	225 (1.5)
130 or more	244 (1.2)

NOTE: Standard errors of the estimated percentages appear in parentheses. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-13. Data for figure 3-3: Percentage of students, by gender and average number of words read per minute, grade 4: 2002

Average number of words read per minute	Percent of students	
	Male	Female
Less than 80	16 (1.6)	11 (1.1)
80–104	21 (1.3)	20 (1.5)
105–129	30 (1.8)	24 (1.6)
130 or more	33 (1.9)	44 (1.9)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-14. Data for figure 3-4: Percentage of students, by race/ethnicity and average number of words read per minute, grade 4: 2002

Average number of words read per minute	Percent of students		
	White	Black	Hispanic
Less than 80	9 (1.2)	26 (2.5)	23 (‡)
80–104	18 (1.5)	28 (2.3)	25 (‡)
105–129	28 (1.7)	28 (2.4)	28 (3.5)
130 or more	45 (1.9)	18 (‡)	24 (‡)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-15. Data for figure 3-5: Percentage of students, by number of words read in the first minute and average number of words read per minute of oral reading, grade 4: 2002

Number of words	Percent of students	
	Reading number of words in the first minute	Reading number of words per minute, on average
Less than 80	7 (0.7)	14 (1.0)
80–104	14 (1.1)	21 (1.1)
105–129	23 (1.2)	27 (1.3)
130 or more	56 (1.8)	38 (1.5)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-16. Data for figure 3-6: Average reading scale scores in relation to the achievement levels, by number of words read in the first minute of oral reading, grade 4: 2002

Number of words read in the first minute of oral reading	Average scale scores
Less than 80	176 (3.8)
80–104	196 (2.0)
105–129	216 (1.8)
130 or more	238 (1.0)

NOTE: Standard errors of the estimated percentages appear in parentheses. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-17. Data for figure 3-7: Percentage of students, by gender and number of words read in the first minute of oral reading, grade 4: 2002

Number of words read in the first minute of oral reading	Percent of students	
	Male	Female
Less than 80	8 (1.2)	5 (‡)
80–104	14 (1.3)	14 (1.6)
105–129	25 (1.6)	21 (1.6)
130 or more	53 (2.1)	60 (2.2)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-18. Data for figure 3-8: Percentage of students, by race/ethnicity and number of words read in the first minute of oral reading, grade 4: 2002

Number of words read in the first minute of oral reading	Percent of students		
	White	Black	Hispanic
Less than 80	5 (‡)	11 (‡)	9 (‡)
80–104	9 (1.1)	25 (1.8)	26 (‡)
105–129	22 (1.5)	29 (2.8)	20 (‡)
130 or more	64 (2.1)	35 (3.0)	45 (4.4)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. The oral reading study passage comprises 198 words. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-19. Data for figure 4-2: Percentage of students, by NAEP reading fluency scale level, grade 4: 2002

Fluency scale	Percent of students
Level 1	8 (0.9)
Level 2	32 (1.4)
Level 3	51 (1.7)
Level 4	10 (0.9)

NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-20. Data for figure 4-3: Average reading scale scores in relation to the achievement levels, by NAEP reading fluency scale level, grade 4: 2002

Fluency scale	Average scale scores
Level 1	177 (3.4)
Level 2	207 (1.7)
Level 3	234 (1.0)
Level 4	252 (2.9)

NOTE: Standard errors of the estimated scale scores appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-21. Data for figure 4-4: Percentage of students, by gender and NAEP reading fluency scale level, grade 4: 2002

	Percent of students	
Fluency scale	**Male**	**Female**
Level 1	9 (1.4)	6 (‡)
Level 2	34 (2.0)	29 (1.8)
Level 3	49 (2.2)	52 (2.1)
Level 4	7 (1.0)	12 (1.3)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-22. Data for figure 4-5: Percentage of students, by race/ethnicity and NAEP reading fluency scale level, grade 4: 2002

	Percent of students		
Fluency scale	**White**	**Black**	**Hispanic**
Level 1	4 (‡)	15 (‡)	16 (‡)
Level 2	28 (1.8)	45 (2.8)	39 (4.3)
Level 3	56 (2.2)	37 (2.9)	43 (4.5)
Level 4	12 (1.3)	3 (‡)	3 (‡)

‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. Data for Asian/Pacific Islander and American Indian/Alaska Native categories of race/ethnicity are not given because sample sizes for them were insufficient.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-23. Data for figure 4-6: Percentage of students, by degree of reading accuracy and NAEP reading fluency scale level, grade 4: 2002

	Percent of students			
Fluency scale	**0–4 errors [100–98%]**	**5–10 errors [97–95%]**	**11–20 errors [94–90%]**	**21 or more errors [less than 90%]**
Level 1	3 (‡)	14 (‡)	35 (‡)	48 (5.0)
Level 2	15 (1.8)	45 (2.2)	32 (2.1)	8 (‡)
Level 3	46 (2.0)	43 (2.1)	11 (1.1)	# (‡)
Level 4	66 (3.2)	27 (‡)	7 (‡)	# (‡)

The estimate rounds to zero.
‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. Percent of words read accurately appear in brackets.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

Table C-24. Data for figure 4-7: Percentage of students, by reading rate and NAEP reading fluency scale level, grade 4: 2002

	Percent of students			
Fluency scale	Less than 80	80–104	105–129	130 or more
Level 1	96 (2.2)	4 (‡)	# (‡)	# (‡)
Level 2	20 (1.5)	59 (2.3)	21 (1.8)	1 (‡)
Level 3	# (‡)	3 (‡)	41 (1.8)	56 (1.9)
Level 4	1 (‡)	# (‡)	# (‡)	99 (1.2)

The estimate rounds to zero.
‡ Reporting standards not met. Standard error estimates cannot be accurately determined.
NOTE: Standard errors of the estimated percentages appear in parentheses. Detail may not sum to totals because of rounding. Reading rate is defined by number of words read per minute. The oral reading study passage comprises 198 words.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Oral Reading Study.

THIS PAGE INTENTIONALLY LEFT BLANK.

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

Official Business Only
Penalty for Private Use, $300

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