DOCUMENT RESUME

ED 378 548 CS 011 968

AUTHOR Langer, Judith A.; And Others
TITLE Reading Assessment Redesigned: Authentic Texts and Innovative Instruments in NAEP's 1992 Survey.
INSTITUTION Educational Testing Service, Princeton, NJ. Center for the Assessment of Educational Progress.; National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO ISBN-0-88685-152-1; NAEP-23-FR-07; NCES-95-727
PUB DATE Jan 95
NOTE 199p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS *Evaluation Methods; Grade 4; Grade 8; Grade 12; Intermediate Grades; *Reading Achievement; *Reading Comprehension; Reading Research; Secondary Education; Sex Differences; *Student Evaluation
IDENTIFIERS Alternative Assessment; *National Assessment of Educational Progress; Reading Uses; *Text Factors

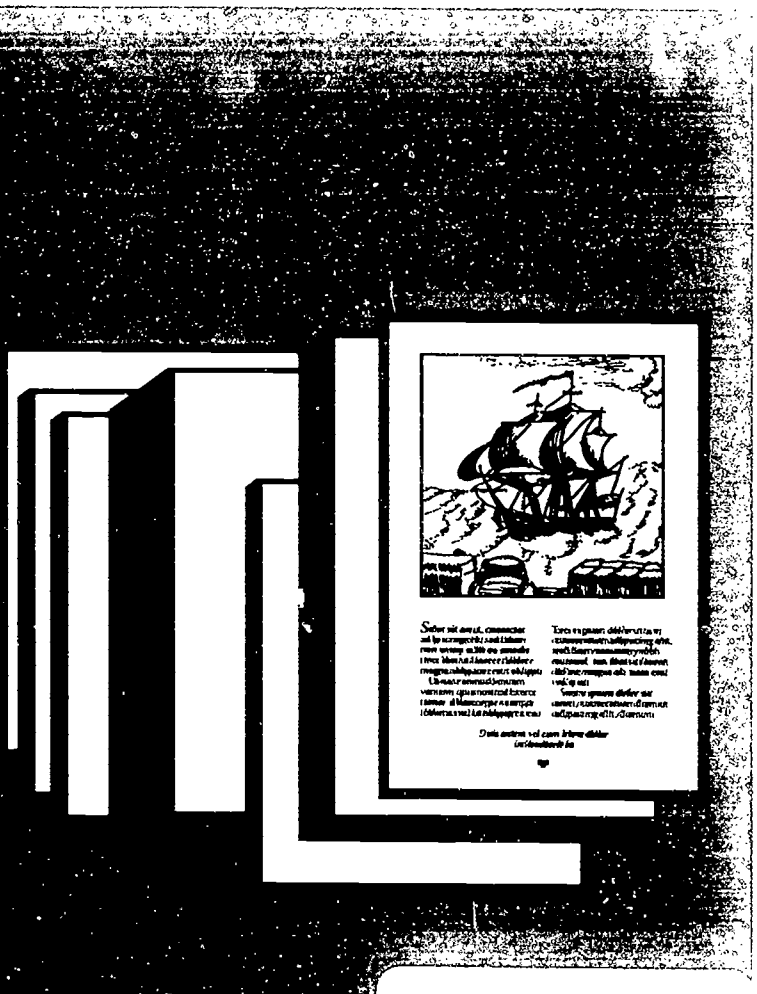
ABSTRACT

Highlighting the important innovations embodied in the 1992 National Assessment of Educational Progress' (NAEP) Reading Report Card, this report provides information on how the NAEP's large-scale reading assessment is evolving in response to changing perceptions of reading development and assessment procedures. Included in the report is an overview of the theoretical framework underlying the assessment, a description of and presentation of reading materials used in the assessment, a discussion of students' performance on constructed-response questions, and a presentation of example questions. Major findings discussed in the report include: (1) at grades 4, 8, and 12, students' average performance was highest on multiple choice questions, somewhat lower on short constructed-response questions, and lowest on extended-response questions; (2) the advantage of female students over male students in reading achievement was more evident for the short constructed-response questions than for multiple-choice questions, and the most evident for extended-response questions; and (3) when demonstrating comprehension of texts that they had selected from a compendium of seven short stories, eighth and twelfth graders demonstrated relative success in answering the constructed-response questions. Also included in the report are results of students' performance in reading for different purposes. Finally, two special studies conducted in 1992 are highlighted in the report--a literary selection task and a comparison of oral and written responses to comprehension questions. Contains 31 tables and five figures of data. A procedural appendix is attached. (RS)

NATIONAL CENTER FOR EDUCATION STATISTICS

Reading Assessment Redesigned

Authentic Texts and Innovative Instruments in NAEP's 1992 Survey



THE NATION'S REPORT CARD

naep

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

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

What is The Nation's Report Card?

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

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

In 1988, Congress created the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The board is responsible for selecting the subject areas to be assessed, which may include adding to those specified by Congress; identifying appropriate achievement goals for each age and grade; developing assessment objectives; developing test specifications; designing the assessment methodology; developing guidelines and standards for data analysis and for reporting and disseminating results; developing standards and procedures for interstate, regional, and national comparisons; improving the form and use of the National Assessment; and ensuring that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.

The National Assessment Governing Board

Honorable William T. Randall, Chair
Commissioner of Education
State Department of Education
Denver, Colorado

Mary R. Blanton, Vice-Chair
Attorney
Blanton & Blanton
Salisbury, North Carolina

Honorable Evan Bayh
Governor of Indiana
Indianapolis, Indiana

Patsy Cavazos
Principal
W.G. Love-Accelerated Elementary School
Houston, Texas

Honorable Naomi K. Cohen
Former Representative
State of Connecticut
Hartford, Connecticut

Charlotte A. Crabtree
Professor of Education
University of California
Los Angeles, California

Catherine Davidson
Secondary Education Director
Central Kitsap School District
Silverdale, Washington

James Ellingson
4th Grade Teacher
Probstfield Elementary School
Moorehead, Minnesota

Chester E. Finn, Jr.
Founding Partner & Sr. Scholar
The Edison Project
Washington, DC

Michael J. Guerra
Executive Director
National Catholic Education Association
Secondary School Department
Washington, DC

William (Jerry) Hume
Chairman
Basic American, Inc.
San Francisco, California

Jan B. Loveless
Educational Consultant
Jan B. Loveless & Associates
Midland, Michigan

Marilyn McConachie
Local School Board Member
Glenview High Schools
Glenview, Illinois

Honorable Stephen E. Merrill
Governor of New Hampshire
Concord, New Hampshire

Jason Millman
Prof. of Educational Research Methodology
Cornell University
Ithaca, New York

Honorable Richard P. Mills
Commissioner of Education
State Department of Education
Montpelier, Vermont

William J. Moloney
Superintendent of Schools
Calvert County Public Schools
Prince Frederick, Maryland

Mark D. Musick
President
Southern Regional Education Board
Atlanta, Georgia

Mitsugi Nakashima
Hawaii State Board of Education
Honolulu, Hawaii

Michael T. Nettles
Professor of Education & Public Policy
University of Michigan
Ann Arbor, Michigan

Honorable Edgar D. Ross
Senator
Christiansted, St. Croix
U.S. Virgin Islands

Fannie Simmons
Mathematics Specialist
Midlands Improving Math and
Science Hub
Columbia, South Carolina

Marilyn A. Whirry
12th Grade English Teacher
Mira Costa High School
Manhattan Beach, California

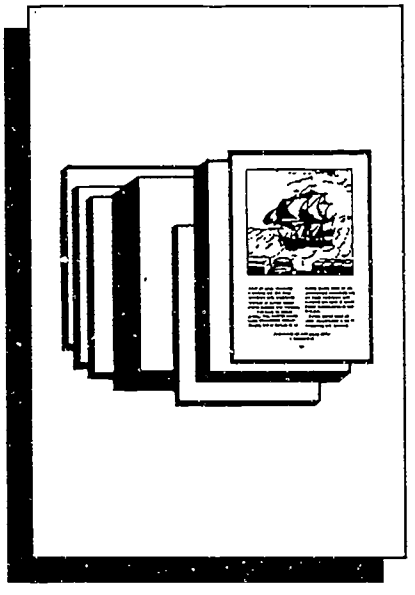
Sharon P. Robinson (ex-officio)
Assistant Secretary
Office of Educational Research
and Improvement
U.S. Department of Education
Washington, DC

Roy Truby
Executive Director, NAGB
Washington, DC

NATIONAL CENTER FOR EDUCATION STATISTICS

Reading Assessment Redesigned

Authentic Texts and Innovative Instruments in NAEP's 1992 Survey



Judith A. Langer
Jay R. Campbell
Susan B. Neuman
Ina V.S. Mullis
Hilary R. Persky
Patricia L. Donahue

Report No. 23-FR-07 January 1995

THE NATION'S REPORT CARD

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

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

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

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

National Center for Education Statistics
Emerson J. Elliott
Commissioner

Education Assessment Division
Gary W. Phillips
Associate Commissioner

FOR MORE INFORMATION:

For ordering information on this report, write:

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

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

Library of Congress, Catalog Card Number: 93-86644

ISBN: 0-88685-152-1

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

. 1
This Report 2
Major Findings 3
Summary 4
Introduction 5
Reports from NAEP's 1992 Reading Assessment 6
The Content of NAEP's 1992 Reading Assessment 7
The Conduct of NAEP's 1992 Reading Assessment 9
This Report In Brief 11
Chapter One — A Framework for Reading Literacy 13
Some History 14
An Interactive Theory of Reading 15
Current Views of Reading Literacy Instruction and Assessment 17
NAEP's Reading Framework 19
Purposes for Reading 20
Figure 1.1 1992 NAEP Framework-Aspects of Reading Literacy 21
Types of Interactions With Text 21
The 1992 NAEP Reading Assessment 22
Summary 24
Chapter Two — The Use of Authentic Texts 27
Why Use Authentic Texts? 27
Selecting the Assessment Texts 29
Examples of Texts Used in the 1992 Assessment 30
Amanda Clement 31
Cady's Life 34
Battle of Shiloh 39
Diversity in Assessment Materials 43
Summary 44
Chapter Three — Fourth Graders' Constructed Responses to Reading 45
Constructed-Response Questions in the NAEP Reading Assessment 46
Average Performance on Constructed-Response Questions 47
Table 3.1 Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 4 49

Table 3.2 Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 4, Trial State Assessment 51
Fourth-Grade Responses To Constructed-Response Questions 52
Grade 4: Amanda Clement — Short Constructed-Responses 52
Table 3.3 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Amanda Clement: Compare to Girls in Sports Today," Grade 4 55
Table 3.4 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Amanda Clement: Compare to Girls in Sports Today," Grade 4, Trial State Assessment 56
Table 3.5 Percentage of Acceptable Responses for the Short Constructed-Response Question,"Amanda Clement: Examples of Mandy Not a Quitter," Grade 4 60
Table 3.6 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Amanda Clement: Examples Mandy Not a Quitter," Grade 4, Trial State Assessment 61
Table 3.7 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Amanda Clement: Hank's Role in Her Career," Grade 4 64
Table 3.8 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Amanda Clement: Hank's Role in Her Career," Grade 4, Trial State Assessment 65
Grade 4: Amanda Clement — Extended-Response Question 66
Table 3.9 Percentage of Responses for the Extended Constructed-Response Question, "Amanda Clement: The Umpire in a Skirt"—"If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know." Grade 4 67
Table 3.10 Percentage of Responses for the Extended Constructed Response Question, Amanda Clement: The Umpire in a Skirt — "If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know." Grade 4, Trial State Assessment 76
Summary 77
Chapter Four — Eighth Graders' Constructed Responses to Reading 79
Average Performance on Question Types 80

Table 4.1 Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 8 81
Eighth-Grade Responses to Constructed-Response Questions 82
Grade 8: Cady's Life — Short Constructed-Responses 82
Table 4.2 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Cady's Life: Why Cady's Perspective," Grade 8 .. 85
Table 4.3 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Cady's Life: Something Anne Frank Could Do," Grade 8 89
Table 4.4 Percentage of Acceptable Responses for the Short Constructed-Response Question,"Cady's Life: Slamming Doors Symbolized Closing the Door of Life," Grade 8 93
Grade 8: Cady's Life — Extended-Response Question 94
Table 4.5 Percentage of Responses for the Extended Constructed-Response Question, "Cady's Life — How the Poem 'I am One' helps to Understand Anne Frank's Life," Grade 8 96
Summary 102
Chapter Five — Twelfth Graders' Constructed Responses to Reading 105
Table 5.1 Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 12 107
Twelfth-Grade Responses to Constructed-Response Questions 108
Grade 12: "Battle of Shiloh"— Short Constructed-Responses 108
Table 5.2 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Battle of Shiloh: Two Sources Help a Student," Grade 12 111
Table 5.3 Percentage of Acceptable Responses for the Short Constructed-Response Question, "Battle of Shiloh: Identify Two Conflicting Emotions," Grade 12 115
Grade 12:"Battle of Shiloh"— Extended-Response Question 115
Table 5.4 Percentage of Responses for the Extended-Response Question,"Battle of Shiloh — Information and Ferspective of the Two Differing Accounts," Grade 12 117
Summary 122
Chapter Six — Written Versus Oral Demonstrations of Reading Comprehension 125
Eliciting Written and Oral Responses 127
Scoring Written and Oral Responses 128
Comparing Written and Oral Demonstrations of Comprehension 129

Table 6.1 Comparison Between Percentage of Written and Oral Responses to Comprehension Questions, Grade 4 129
Summary 130
Chapter Seven — Student Achievement in Reading for Different Purposes 133
Average Proficiency in Purposes for Reading for the Nation 134
Table 7.1 Average Proficiency in Purposes for Reading, Grades 4, 8, and 12 135
Percentiles by Purposes for Reading 138
Table 7.2 Proficiency Levels of Students at Various Percentiles by Purposes for Reading, Grades 4, 8, and 12 139
Average Proficiency in Purposes for Reading by Region 140
Table 7.3 Average Proficiency in Purposes for Reading by Region, Grades 4, 8, and 12 140
Average Proficiency in Purposes for Reading by Type of School 141
Table 7.4 Average Proficiency in Purposes for Reading by Type of School, Grades 4, 8, and 12 142
Average Proficiency in Purposes for Reading by Gender 143
Table 7.5 Average Proficiency in Purposes for Reading by Gender, Grades 4, 8, and 12 143
Average Proficiency in Purposes for Reading by Race/Ethnicity 144
Table 7.6 Average Proficiency in Purposes for Reading by Race/Ethnicity, Grades 4, 8, and 12 145
Average Proficiency in Purposes for Reading for States 146
Table 7.7 Average Proficiency in Purposes for Reading, Grade 4, Trial State Assessment 148
Figure 7.1 Comparisons of Average Overall Reading Proficiency, Grade 4, Trial State Assessment 149
Figure 7.2 Comparisons of Average Reading for Literary Experience Proficiency, Grade 4, Trial State Assessment 150
Figure 7.3 Comparisons of Average Reading to Gain Information Proficiency, Grade 4, Trial State Assessment 151
Summary 152
Chapter Eight — The NAEP Reader: Self-Selection While Reading for Literary Experience 155
Administering The NAEP Reader Selection Task 157
Figure 8.1 Story Summaries for The NAEP Reader at Grades 8 and 12 158
Students' Selections of Stories in The NAEP Reader 159

Table 8.1 Percentages of Students Selecting Stories from The NAEP Reader, Grades 8 and 12 159
How Students Make Reading Selections 160
Table 8.2 Summary of the Selection Criteria Indicated by 8th- and 12th-grade Students Choosing Stories from The NAEP Reader 161
Students' Comprehension of What They Selected to Read 162
Table 8.3 Average Percentage of Students with Acceptable Answers on Short Constructed-Response Questions About Stories in The NAEP Reader, Grades 8 and 12 163
Table 8.4 Percentages of Students Demonstrating Essential or Better Comprehension on the Extended-Response Questions About Stories in The NAEP Reader, Grades 8 and 12 164
Table 8.5 Average Percentage of Students Demonstrating Essential or Better Comprehension on the Extended Constructed-Response Questions in Main Assessment Blocks Measuring Reading for Literary Experience, Grades 8 and 12 165
Summary 167
..... 169
Introduction 169
NAEP's Reading Assessment Content 169
Table A.1 Target and Actual Percentage Distribution of Questions by Grade and Reading Purpose 170
Table A.2 Target and Actual Percentage Distribution of Questions by Grade and Reading Stance 170
The Assessment Design 171
National Sampling 172
Table A.3 1992 Student and School Sample Sizes 173
Trial State Assessment Sampling 173
Participation Rates for States and Territories 174
The Sample Participation Guidelines 175
Table A.4 Summary of School and Student Participation Grade 4, Trial State Assessment 179
LEP and IEP Students 180
Data Collection 181

Scoring 182
Table A.5 Percentages of Exact Agreement for Scoring Reliability Samples for Extended-Response Questions 183
Data Analysis and IRT Scaling 184
Linking the Trial State Results to the National Results 186
NAEP Reporting Groups 187
Minimum Subgroup Sampling Size 188
Estimating Variability 188
Drawing Inferences from the Results 189
Acknowledgments 193

The 1992 National Assessment of Educational Progress (NAEP) in reading incorporated many recent advances in theories eading and innovative approaches to assessing reading developme .. *The NAEP Reading Framework*[1] underlying this assessment views reading as a dynamic, interactive, and constructive process. From this perspective, reading is described as a purposeful, meaning-oriented activity that involves a complex interaction between the reader, the text, and the context.

In developing the 1992 NAEP reading assessment, priority was placed on providing students with materials and reading tasks that resembled authentic literacy demands. That is, the texts used in the assessment were selected from publications that would typically be available to students in and out of school. Furthermore, emphasis was placed on having students demonstrate their comprehension through constructed-response questions.

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

Beyond these innovations, the 1992 NAEP reading assessment included additional features and special studies that represented a broad view of reading and literacy development. For example, many eighth and twelfth graders were given opportunities to make literary selections and a sample of fourth graders were involved in one-on-one literacy interviews. Overall, the 1992 NAEP reading assessment represented an important effort in moving large-scale reading assessments closer to the prevailing view of the reading process.

The assessment was administered to nationally representative samples of fourth-, eighth-, and twelfth-grade students attending public and private schools, and to state representative public-school samples of fourth graders in 43 jurisdictions. Nearly 140,000 students were assessed in all. The data were summarized on the NAEP reading proficiency scale ranging from 0 to 500.

This Report

This report serves as a follow-up to the *1992 NAEP Reading Report Card*[2] that presented overall reading achievement results as well as information regarding instructional and home background experiences for the nation. In addition, because the 1992 reading assessment included a state assessment in reading at grade 4, the *Report Card* presented comparative information for those participating states and territories.

In order to highlight the important innovations embodied in the 1992 assessment, this report focuses on those aspects of the reading assessment that were not presented in the *Report Card*. Included in this report is an overview of the theoretical framework underlying the assessment, a description and presentation of reading materials used in the assessment, a discussion of students' performance on constructed-response questions, and a presentation of example questions. Also, the results of students' performance in reading for different purposes is presented in this report. Finally, two special studies conducted in 1992 are highlighted -- a literary selection task, and a comparison of oral and written responses to comprehension questions.

---

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

Major Findings

Along with a discussion of the NAEP reading assessment framework and its innovations, this report includes the following major findings from the 1992 reading assessment:

- At all three grades, students' average performance was highest on multiple-choice questions (63 to 68 percent correct), somewhat lower on short constructed-response questions (51 to 61 percent acceptable), and lowest on extended-response questions (25 to 38 percent essential or better).
- Differences in reading performance by demographic subgroups remained relatively consistent across the different question types in the assessment, with one exception. The advantage of female students over male students in reading achievement was more evident for the short constructed-response questions than for the multiple-choice questions, and the most evident for extended-response questions.
- Fourth graders demonstrated increased performance on constructed-response questions when giving answers orally compared to when they provided written responses.
- Consistent with research about students' exposure to different types of text as they progress through school, students at grade 4 had higher average proficiency in reading for literary experience, whereas, students at grade 8 demonstrated little difference in performance across the three purposes, and students at grade 12 had higher proficiencies in reading to gain information and to perform a task.
- In a literary story-selection task, eighth and twelfth graders demonstrated little clear decision making criteria for selecting stories. For example, 36 percent at grade 8 and 18 percent at grade 12 did not express a specific criterion when asked why they made their story selection.
- When demonstrating comprehension of texts that they had selected from a compendium of seven short stories, eighth and twelfth graders demonstrated relative success in answering the constructed-response questions. For example, across the seven stories, from 35 to 63 percent of the eighth graders, and from 51 to 78 percent of the twelfth graders provided complete answers to an extended-response question about a major conflict in the story.

Summary

This report provides information that may be considered useful by educators, administrators, and researchers who are interested in how large-scale reading assessments are evolving in response to changing perceptions of reading development and assessment procedures. Findings from innovative components of the 1992 NAEP reading assessment are provided in this report, including students' performance on constructed-response questions, students' achievement in different purposes for reading, the results of a response mode comparison at fourth grade, and the results of a literary self-selection task at grades 8 and 12. Along with the *NAEP 1992 Reading Report Card*, this report demonstrates NAEP's ongoing commitment to providing relevant information about the educational progress of the nation's students, and to do so with instruments that reflect current knowledge about instruction and assessment.

Efforts to increase the literacy achievement of students in the United States across the past decade have generated considerable changes in ideas about reading instructional approaches and emphases. In the ten years since the publication of *Becoming a Nation of Readers*, educators and researchers across the country have become mobilized in implementing classroom practices that cultivate a literate environment and foster the development of those attitudes and skills that characterize the "life-long reader."[3] Foremost among these attempts to advance literacy learning has been an awareness that reading activities in the classroom should mirror those of the world outside of school. These activities, more recently referred to as authentic literacy tasks, are those in which ". . . reading and writing serve a function for children, activities such as enjoying a book or communicating an idea in a composition."[4] From this perspective on instruction, reading and

[3]Anderson, R.C., Hiebert, E.H., Scott, J.A., & Wilkinson, I.A.G., *Becoming a Nation of Readers: The Report of the Commission on Reading* (Washington, DC: The National Institute of Education, 1985).

[4]Hiebert, E.H., Becoming Literate Through Authentic Tasks: Evidence and Adaptations. In Ruddell, R.B., Ruddell, M.R., & Singer, H. (Eds.), *Theoretical Models and Processes of Reading*, pp. 391-413, (Newark, DE: International Reading Association, 1994).

responding to reading are viewed as integrated, purposeful activities directed toward the goal of constructing meaning.

One priority that emerged from these reform efforts is a renewed focus on assessment methods and procedures that can support and provide feedback for reading instruction. The push has been toward integrative assessments that reflect quality instruction and involve students in reading tasks that replicate purposeful, engaging reading experiences. Assessment innovations have stressed the need to move beyond reliance on traditional multiple-choice questions as the single format with which students demonstrate their understandings. Written responses to reading, instead, provide students with opportunities to show how they construct meaning, integrate personal knowledge with text, and critically consider textual elements — important goals in students' literacy development.

In the context of these evolving ideas about reading instruction and assessment, the 1992 National Assessment of Educational Progress (NAEP) reading assessment was developed with a view of reading that reflected current reading research and assessment practices. From an interactive, constructive view of the reading process, the *Reading Framework* underlying the assessment set forth specifications that called for the use of whole authentic materials representing different types of reading purposes and drawn from sources typically available to students.[5] In addition, the framework specified that a majority of students' time be spent providing written responses to reading, and thus, demonstrate their abilities to construct, extend, and examine meaning.

Reports from NAEP's 1992 Reading Assessment

The summary results from NAEP's 1992 reading assessment were released in the *NAEP 1992 Reading Report Card for the Nation and the States*.[6] The *Report Card* presented overall reading achievement results for students at grades 4, 8, and 12 for the nation and for various demographic subgroups. Comparative results were included at grade 4 for 43 participating states and territories. In addition, contextual information regarding students' instructional and home background experiences were discussed in light of students' reading proficiency.

---

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

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

As a follow-up to the *1992 NAEP Reading Report Card*, this publication links the approaches used in the assessment to instructional settings, and discusses students' responses to individual constructed-response questions. The discussion and results provided in this report include a focus on the innovative nature of NAEP's reading framework, a discussion of the authentic materials used in the assessment, a highlighting of students' performance on the different types of questions with a specific focus on their answers to constructed-response questions, and a presentation of students' proficiency in reading for different purposes. In addition, this report includes results from two special studies that augmented the 1992 reading assessment — a comparison of response modes in answering comprehension questions, and an examination of students' performance with a self-selection literary task based on "*The NAEP Reader*," a compendium of short stories.

In addition, there is a pair of reports describing the results from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4. In this special study, fourth graders were interviewed in one-on-one situations about their reading habits and instruction, and asked to read aloud. *Interviewing Children About Their Literacy Experiences*[7] provides the results of the conversations conducted with fourth graders in the IRPR study about their reading habits and their classroom activities related to reading. It also describes how these literacy experiences relate to students' overall reading proficiency as determined by their performance in the main portion of the 1992 reading assessment. The companion report, *Listening to Children Read Aloud*[8], focuses specifically on fourth graders' oral reading abilities. This report provides a thorough discussion of the rationale for assessing students' oral reading, the procedures used in conducting such an assessment, as well as the results of their oral reading achievement.

The Content of NAEP's 1992 Reading Assessment

The *Reading Framework* underlying the 1992 assessment was newly developed and adapted by the National Assessment Governing Board (NAGB) specifically for this assessment, including the Trial State

---

[7]Campbell, J.R., Kapinus, B.A., Beatty, A.S., *Interviewing Children About Their Literacy Experiences* (Washington, DC: National Center for Education Statistics, 1995).

[8]Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B., Beatty, A.S., *Listening to Children Read Aloud* (Washington, DC: National Center for Education Statistics, 1995).

Assessment Program. To ensure a forward-looking conceptualization of reading that was responsive to needs of policy makers and educators and that accounted for contemporary research on reading and literacy, a national consensus process was used to develop the framework. The consensus process, which was managed by the Council of Chief State School Officers (CCSSO) under the direction of the National Assessment Governing Board (NAGB), involved a 16-member Steering Committee representing national organizations and a 15-member Planning Committee of reading experts, including educators, researchers, and curriculum specialists. The CCSSO project staff and NAGB continually sought guidance and reaction from a wide range of individuals in the fields of reading and assessment.

In brief, the *Reading Framework* consists of major purposes for reading and, as a cross-cutting dimension, the interactions that readers have with text as they construct, extend, and examine meaning. The purposes include reading for **literary experience**, to **gain information**, and to **perform a task**, although the latter was not assessed at grade 4. The interactions or reading stances include forming an **initial understanding, developing an interpretation, personal reflection and response**, and demonstrating a **critical stance**.

The reading materials included in the assessment consisted of a wide variety of intact texts, reproduced as faithfully as possible from their original sources. Literary texts included short stories, poems, fables, historical fiction, science fiction, and mysteries. Informational materials included biographies, science articles, encyclopedia entries, primary and secondary historical accounts, and newspaper editorials. Reading to perform a task used such documents as instructions, forms, and schedules.

A combination of constructed-response and multiple-choice questions was used as determined by the nature of the reading tasks associated with each text or sets of texts. To better measure the processes readers use, from 60 to 70 percent of the students' response time was devoted to constructed-response questions. There were two types of constructed-response questions, short and extended. The short constructed-response questions required answers from a few words to a few sentences and were evaluated as either acceptable or unacceptable. The extended questions require responses of a paragraph or more, and were evaluated according to a 4-point scale ranging from *unsatisfactory* to *extensive*. Each text or set of texts was accompanied by at least one extended-response question.

The Conduct of NAEP's 1992 Reading Assessment

As with all NAEP assessments, the schools and students participating in the 1992 reading assessment were selected through scientifically designed stratified random sampling procedures. Approximately 26,000 fourth, eighth, and twelfth graders in 1,500 public and private schools across the country participated in the national assessment. In addition, NAEP's voluntary Trial State Assessment Program was conducted in 44 jurisdictions at grade 4.* For each jurisdiction participating in the Trial State Assessment Program, separate state-representative samples of fourth graders were assessed, involving approximately 2,500 students sampled from approximately 100 public schools. Thus, NAEP's Trial State Assessment Program in reading involved approximately 110,100 students.

All NAEP data are collected by trained administrators. Data for the national assessment were collected by a field staff managed by Westat, Inc. However, in accordance with NAEP legislation, data collection for the Trial State Assessment Program was the responsibility of each participating jurisdiction. Uniformity of procedures across states was achieved through training and quality control monitoring by Westat, Inc. Quality control was provided by unannounced, random monitoring of half the sessions in each state. The results of the monitoring indicated a high degree of quality and uniformity across sessions.

Unless the overall participation rate is high for a state or territory, there is a risk that the assessment results for the jurisdiction are subject to appreciable nonresponse bias. It should be noted that even though all jurisdictions met the guidelines for high student participation rates, several states did not satisfy the guidelines for school participation rates (see Procedural Appendix for the guidelines). Further analyses, documented in the *Technical Report of the 1992 Trial State Assessment in Reading*, suggest that nonresponse bias due to varying participation rates was either non-existent or quite small. However, Delaware, Maine, Nebraska, New Hampshire, New Jersey, and New York are designated with asterisks in the tables containing state-by-state results, because they did not satisfy the guidelines.

The assessment booklets, including the approximately two million written responses constructed by students, were scored by National Computer Systems. The constructed-response questions were scored by

*In accordance with the legislation providing for participants to review and give permission for release of their results, the Virgin Islands choose not to release their results. Therefore, data were reported for 43 of 44 jurisdictions.

professional readers who had experience in education. These readers were thoroughly trained to use scoring guides developed by the NAEP Reading Test Development Committee and Educational Testing Service staff. To determine the reliability of the scoring, 25 percent of the students' responses to each question were evaluated by two different scorers. For the nation, the percentage of exact agreement between scorers, averaged across questions, was approximately 89 percent for grade 4, 86 percent for grade 8, and 88 percent for grade 12. For the Trial State Assessment Program at grade 4, the percentage of exact agreement, averaged across all questions for all states and territories, was approximately 91 percent.

The assessment results were analyzed by ETS to determine the percentage of students responding correctly to each multiple-choice or short constructed-response question and the percentage of students responding in each of the four categories for the extended-response questions. Item response theory (IRT) methods were used to summarize results for each of the reading purposes in the framework (two purposes at grade 4 — literary and informational — as well as the third — to perform a task — at grades 8 and 12). As an analysis innovation for the 1992 assessment, a partial-credit scaling procedure employing a specialized IRT method was used to account for students' responses according to the 4-point guides used with the extended-response questions. An overall composite scale was developed by weighting each reading purpose according to its importance in the framework (see the Procedural Appendix). The NAEP reading proficiency scales, for each of the purposes and the overall scale, range from 0 to 500. Unless otherwise noted, all changes or differences discussed in this report are statistically significant at the .05 level of significance. This means that the observed differences are unlikely to be due to change or to sampling variability. These "confidence intervals" are described in greater depth in the Procedural Appendix.

Throughout the development and conduct of the assessment, NCES and its contractors worked closely with the Trial State Assessment NETWORK, which includes representatives from all interested states. Federal funding permitted regular NETWORK meetings, where state education personnel met with staff members from NCES, the contractors, NAGB, and CCSSO to review NAEP materials, plans, procedures, and data.

This Report In Brief

The *Reading Framework* is presented in Chapter One, including a discussion about the importance of the "thinking" aspects of literacy in today's information society. Special emphasis is placed on the use of authentic, purposeful, thoughtful reading tasks both in the classroom and in assessment. Chapter Two describes the types of reading materials included in the 1992 assessment, and provides an illustrative text from each of the three grades assessed. The fourth-grade selection is a biographical article, while the eighth-grade example includes a short story with a biographical sketch of the author, Anne Frank, that was paired with a poem by a different author. At grade 12, the example consists of a journal entry by an officer who fought in the Battle of Shiloh juxtaposed with the encyclopedia description of the battle.

Chapters Three through Five contain examples of the constructed-response questions and students' responses to them for each of three grades assessed, respectively. The results were quite consistent across grades:

- At all three grades, students average performance was highest on multiple-choice questions (63 to 68 percent), somewhat lower on short constructed-response questions (51 to 61 percent), and lowest on extended-response questions (25 to 38 percent). The difference generally was larger between short- and extended-response questions than between multiple-choice and short-response questions, especially at grade 12.

- There was, however, a range of performance in the percentages of students providing complete answers to the extended-response questions. For example, only 11 percent of the eighth graders were able to connect the biographical information about Anne Frank to the theme of the poem entitled "I Am One." In contrast, about half the twelfth graders (52 percent) described the unique perspectives provided by the journal and encyclopedia entries about the Battle of Shiloh.

- At all three grades, for all three types of questions, performance differences for students from different subgroups were quite consistent. For example, students from advantaged areas had higher average performance than those from disadvantaged or rural communities, private school students had higher average achievement than public school students, and White students had higher average performance than Black or Hispanic students.

- Also, at all three grades, for all three types of questions, females had higher average performance than males. At grades 8 and 12, this advantage for females over males in reading proficiency was more evident for the short constructed-response questions than for the multiple-choice questions, and the most evident for extended-response questions. For example, at grade 12, the difference between male and female performance on multiple-choice questions was only 2 percent, while the differences for performance on short and extended constructed-response questions was 7 and 11 percent, respectively.

Chapter Six presents a comparison of written and oral performance on three fourth-grade reading comprehension questions as measured by the main NAEP assessment and by the Integrated Reading Performance Record (IRPR) special study. In addition to participating in a literacy interview and demonstrating their oral reading fluency, fourth graders in the IRPR special study provided oral responses to comprehension questions after a second reading of the passage and after a second exposure to the questions. The results of this response-mode comparison revealed an advantage for providing oral responses to comprehension questions.

Chapter Seven summarizes students' average achievement for the different reading purposes. Consistent with research about students' exposure to different types of text as they progress through school, students at grade 4 had higher average proficiency in reading for literary experience, whereas students at grade 8 demonstrated little difference in performance across the three purposes, and students at grade 12 had higher proficiencies in reading to gain information and to perform a task. This pattern generally prevailed across public and private school students, regions, and states.

Chapter Eight contains data from the special study using the "The NAEP Reader." Surprisingly, eighth and twelfth graders showed little clear decision-making criteria in selecting their stories. For example, 36 percent at grade 8 and 18 percent at grade 12 did not express a specific criterion when asked why they made their story selection. Students, however, demonstrated relative success in answering the constructed-response questions about their self-selected stories. For example, across the seven stories, from 35 to 63 percent of the eighth graders, and from 51 to 78 percent of the twelfth graders provided complete answers to an extended-response question about identifying and describing a major conflict in the story they had chosen. (The best performance on a literary extended-response question in the main portion of the assessment was 38 percent complete responses.)

The last 10 years have been important ones in American education. A great deal of knowledge gained from research and classroom practices has coalesced into a large scale effort at systemic reform. Part of this effort has emphasized developing closer links between the goals and methodologies underlying both instruction and assessment, in the belief that all parts of the educational system need to work together in support of the same educational objectives.

At that same time, many educators and researchers have embraced a broader view of reading and the processes that contribute to reading proficiency. Currently, there is a general consensus that reading is more than a simple, unidimensional skill. As described in the *NAEP Reading Framework*, "reading literacy" entails not only being able to read, but also knowing when to read, how to read, and how to reflect on what has been read. Thus, throughout this report the terms "reading assessment" and "reading literacy assessment" are used interchangeably in reference to the 1992 NAEP assessment in reading.

Some History

Since its inception, the National Assessment of Educational Progress has attempted to reflect the current thinking about teaching and testing. In 1969, the National Assessment of Educational Progress (NAEP) was established by the United States Congress, with the mandate to conduct national surveys of student achievement and to report these results to the nation. Since that time, reading achievement, the most fundamental ability taught in school, has been assessed about every four years and more recently, every two years. In order to provide the most relevant and useful data to policy makers, educators, and the general public, NAEP has maintained ongoing consultant-relationships with teachers, researchers, administrators, government leaders, parents, and the business community. Across the past 20 years, these partnerships have helped insure that NAEP reflects a general consensus about important competencies in student literacy as well as current research on effective methods for teaching and assessment.

Since 1980, NAEP's assessments have indicated that by and large students can understand what they read, but that understanding is at a surface and unreflective level.[10] Even twelfth-grade students have difficulty elaborating, explaining, or defending their understandings. Across time, large percentages of students have been able to understand at a superficial level — but more thoughtful reasoning continues to prove difficult for all students.[11] Although students from historically underserved minority groups have shown gains in achievement since the 1970s, two problems continue: 1) the achievement gap between these students and their White classmates, although diminishing, remains large, and 2) all students primarily demonstrate only surface as opposed to more reasoned comprehension. These results, combined with a growing concern about the need for more thought-provoking educational experiences for all students, have led to a variety of calls for higher standards in education.[12] These have

---

[10] National Assessment of Educational Progress, *Reading, Thinking, and Writing* (Denver, CO: Education Commission of the States, 1981).

Applebee, A.N., Langer, J.A., & Mullis, I.V.S., *Learning to Be Literate in America: Reading, Writing, and Reasoning* (Princeton, NJ: Educational Testing Service, 1987).

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

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

[12] The National Council on Education Standards and Testing, *Raising Standards for American Education* (Washington, DC: U.S. Department of Education, 1992).

been augmented by an overall initiative stressing the need for systemic change, with an overt goal to achieve educational equity through attention to content, opportunity, instructional, and delivery standards.[13]

The current NAEP reading framework and its view of reading reflects the prevailing consensus of educators, researchers, parents, community leaders, and policy makers. It addresses reading comprehension and literacy learning in ways that treat all students as thinkers — individuals who have ideas in response to what they read. Furthermore, the framework conceives of instructional and assessment activities as essentially purposeful, thus, engaging, thought-provoking, and complete. With this in mind, the Planning Committee determined that the NAEP reading assessment must contain reading materials and tasks "... so similar to those which students encounter in classrooms and in their own reading that, should teachers choose to do so, they could use the kinds of passages and tasks found on the assessment to set priorities in their classrooms without distorting instruction."[14]

An Interactive Theory of Reading

Instead of thinking of literacy as the ability to read and write, it may be more productive to think of literacy as:

> the ability to think and reason as a literate person.... Here the focus is not just on the reading, but also on the thinking that accompanies it. In this case, literacy can be thought of as a tool.[15]

---

[13] O'Day, J. A., & Smith, M.S., "Systemic Reform and Educational Opportunity," In S.H. Furhman, editor, *Designing Coherent Policy: Improving the System* (San Francisco, CA: Jossey-Bass, 1993).

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

[15] Langer, J.A., & Applebee, A.N., *How Writing Shapes Thinking* (Urban, IL: National Council of Teachers of English, 1987)

This view of literacy focuses on students as active thinkers, and on the spiraling changes that take place when they use their literacy skills to think, rethink, and interpret their knowledge, the world, and themselves. When students are treated as active thinkers and asked engaging questions about their reading they will learn to reflect upon, develop, and explain deeper understandings.[16] From this perspective, reading to develop better understanding includes knowing when to read, how to read, how to reflect on what has been read, and ways to communicate growing understandings.

Reading for deeper meaning involves a dynamic and complex interaction among the reader (attitudes, experiences, and expectations), the text (topic, format, and content), and the context (the environment, activity, questions, and interaction) — over time.[17] Understandings do not develop the moment the reading activity starts, ideas do not become fixed at some point during the reading, and comprehension is not complete even after the final words are read.[18] Further, there are a variety of kinds of knowledge a reader might call on when constructing meaning, and these are affected by the purpose.

---

[16] Garrison, M.B., & Hynds, S., "Evocation and Reflection in the Reading Transaction: A Comparison of Proficient and Less Proficient Readers," *Journal of Reading Behavior*, 23.

Langer, J.A., "Levels of Questioning: An Alternative View," *Reading Research Quarterly*, 20, 1985.

Langer, J.A., *Teacher Disciplinary Thinking in Academic Coursework*, National Research Center on Literature Teaching and Learning (Washington, DC: Office of Educational Research and Improvement, 1993).

Nystrand, M., & Gamoran, A., "Instructional Discourse, Student Engagement, and Literature Achievement," *Research in the Teaching of English*, 25, 261-290, 1991.

Perkins, D., "Teaching for Understanding," *American Educators: The Professional Journal of the American Federation of Teachers*, 17, 28-35, 1993.

[17] Ruddell, R.B., & Unrau, N.J., "Reading as a Meaning-Construction Process: the Reader, the Text, and the Teacher," In Ruddell, R.B., Ruddell, M.R., & Singer, H., (Eds.) *Theoretical Models and Processes of Reading* (Newark, DE: International Reading Association, 1994).

[18] Tierney, R.J., "Redefining Reading Comprehension," *Educational Leadership*, pp. 37-42, March, 1990.

Langer, J.A., "The Process of Understanding: Reading for Literary and Informational Purposes," in *Research in the Teaching of English*, 24, pp. 229-260, 1990.

Current Views of Reading Literacy Instruction and Assessment

The framework developed for NAEP's 1992 reading assessment is based on a variety of research conclusions about the most effective contexts for instruction and assessment. The current views of literacy *instruction* underlying the NAEP framework include:

- **Thought-provoking activities and interactions.** Activities that set problems, invite discussion, or request explanations can be extremely useful in engaging students.[19] The goal of these activities is to take students beyond simply "knowing" the text to understanding how textual ideas relate, why they are important, and how they can be used.
- **Intact, complete texts.** The use of naturally-occurring, authentic texts in classroom instruction and assessment has received increased attention.[20] Unlike isolated exercises, whole texts represent the kinds of everyday or on-the-job reading tasks that have understandable ends for students to think towards. For example, a note and a letter are whole, just as a book is whole. Length is not the issue; a complete text of any length carries with it understandings of the social meanings for which the entire piece was intended, while a short made-up sentence or paragraph may not.
- **Purposeful assignments.** Reading takes place in many different situations for many different purposes. Readers may orient themselves to a particular text very differently, depending on the nature of the text or their reason for reading.[21] Reading to curl up with a mystery, reading to write a history report for school, reading to bake cookies, and reading to do a lab experiment are all different

[19] Calfee, R.C., Dunlap, K.L., & Wat, Y.W., "Authentic Discussion of Texts in Middle Grade Schooling: An Analytic-Narrative Approach," *Journal of Reading*, 37(7), 546-556, 1994.

Beck, I.L., "Reading an Reasoning," *The Reading Teacher*, 42(9), 676-682, 1989.

[20] Hiebert, E.H., "Becoming Literate Through Authentic Tasks: Evidence and Adaptations," In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.). *Theoretical Models and Processes of Reading*, 391-413 (Newark, DE: International Reading Association, 1994).

[21] Langer, J.A., "The Process of Understanding: Reading for Literary and Informative Purposes," *Research in the Teaching of English*, 24(3), 229-260, 1990.

kinds of purposeful activities, requiring different approaches to gaining meaning. Reading instruction involves helping students learn to fulfill different purposes.[22]

- **Integrating reading and writing.** Becoming an author, reading other people's writing, and writing about what has been read are three different types of activities that are closely related. All three activities involve students in thinking, learning, and communicating.[23] Together they may help students gain understandings about the underlying content, structure, and social uses for literacy, as well as how to successfully participate in literate events.[24]

Current approaches to reading *assessment* that are reflected in the NAEP framework include:

- **Assessing with an array of texts and topics.** Different types of texts have different organizations and features that have an effect on how they are read.[25] Consequently, there is increasing agreement among literacy educators and researchers that assessments should involve students in reading and commenting on an array of genres and subgenres with varied content and structures.

- **Engaging students in thought provoking, constructed-response tasks.** Assessments that are intended to measure complex, integrative abilities and processes may need to involve students in more than just selection tasks.[26] Therefore, many educators

---

[22] Blanton, W.E., Wood, K.D., & Moorman, G.B., "The Role of Purpose in Reading Instruction," *The Reading Teacher*, *42*(7), 486-493, 1990.

[23] Langer, J.A., "Reading, Writing, and Understanding: An Analysis of the Construction of Meaning," *Written Communication*, *3*(2), 219-267, 1988.

Squire, J.R., "Composing and Comprehending: Two Sides of the Same Basic Process," *Language Arts*, *60*, 568-589, 1983.

[24] Musthafa, B., "Literary Response: A Way of Integrating Reading/Writing Activities," *Reading Improvement*, *31*(1), 52-58, 1994.

Lewin, L., "Integrating Reading and Writing Strategies Using an Alternating Teacher-led/Student-selected Instructional Pattern," *The Reading Teacher*, *45*(8), 586-591, 1992.

[25] Pearson, P.D., & Camperell, K., "Comprehension of Text Structures," In R. B. Ruddell, M.R. Ruddell, & H. Singer (Eds.). *Theoretical Models and Processes of Reading*, 448-468 (Newark, DE: International Reading Association, 1994).

[26] Resnick, L.B., & Resnick, D.P., "Assessing the Thinking Curriculum: New Tools for Educational Reform," In Gifford, B.R., & O'Connor, M.C. (Eds.) *Changing Assessments: Alternative Views of Aptitude, Achievement, and Instruction* (Boston, MA: Kluwer Academic Publishers, 1992).

and researchers have argued for less reliance on multiple-choice questions in tests of reading comprehension.[27] Measuring how well students think in response to reading involves the kinds of thought-provoking questions teachers might ask to help students learn to develop, analyze, and explain their own ideas.[28]

- **Assessment that reflects quality instruction.** Some educators have expressed concern that certain reading tests do not assess the scope of literacy development and the deeper levels of understanding that are typically the goal of quality reading instruction.[29] As a result, there has been an effort to make assessments look as much like classroom activities as possible rather than like a specially created testing-genre, requiring special test-taking skills.[30] These efforts have led to a proliferation of new assessment techniques referred to as "authentic assessments."[31] The intent of these methods is to replicate as closely as possible the kinds of experiences students encounter in and out of school when they engage in complete activities with purposes.

NAEP's Reading Framework

NAEP's *Reading Framework* for the 1992 assessment was developed by a planning committee and reviewed extensively by specialists across the country to ensure it reflected a consensus about the best in current practice in instruction and assessment.[32] It is summarized below and in Figure 1.1 taken from the booklet describing the framework. The orientation reflects a focus on performance, involving three major purposes for reading and four different types of interactions with text.

---

[27] Valencia, S., & Pearson, P.D., "Reading Assessment: Time for a Change," *The Reading Teacher, 40,* 726-732, 1987.

[28] Hill, C., & Parry, K., "The Test at the Gate: Models of Literacy in Reading Assessment," *TESOL Quarterly, 26*(3), 433-461, 1993.

[29] McAuliffe, S., "A Study of Differences Between Instructional Practice and Test Preparation," *Journal of Reading, 36,* 524-530, 1993.

[30] Wiggins, G., "Assessment: Authenticity, Context, and Validity," *Phi Delta Kappan,* 200-214, 1993.

[31] Valencia, S.W., Hiebert, E.H., & Afflerbach, P.P. (Eds.), *Authentic Reading Assessment: Practices and Possibilities* (Newark, DE: International Reading Association, 1994).

[32] The *Reading Framework for the 1992 National Assessment of Educational Progress* also was adopted for the follow-up reading assessment in 1994. Please see *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, U.S., Department of Education).

Purposes for Reading

Reading for Literary Experience. Readers "step into the world of the story" when they read for a literary experience. They become "insiders," calling on all they know and can imagine about human nature and experience in order to explore interplays among events, emotions, and the human condition. They explore horizons of possibilities about motives, feelings, and eventualities. They take multiple perspectives, see many sides of situations, and always leave room to explore yet another interpretation. It is this act of exploring possibilities that lies at the heart of reading for the literary experience. Such readings usually involve but need not be limited to novels, short stories, poems, plays, and essays.

Reading to Gain Information. When reading to be informed, readers gather, consider, and shape their growing understandings. Since the goal is to gain information, readers focus on the type of knowledge they are after; for example, to find specific pieces of information when preparing a research project, or to get some general information when glancing through a magazine article. While readers also ask questions and explore possibilities, these center around the particular point or kind of information being sought. Different orientations are required than in reading for literary experience, because maintaining a point of reference (a topic or issue) and building understandings about it lies at the heart of the reading to gain information. Also, informational materials tend to have their own text features. This type of reading usually involves articles, informational non-fiction, encyclopedias, and textbooks.

Reading to Perform a Task. When reading to perform a task, readers usually seek a quick and ready application to a situation or task they have in mind or at hand. Such tasks generally involve the reading of documents such as bus or train schedules; directions for games, repairs, or recipes; tax or voter information; and office memos. Readers must use their expectations of the purposes of the documents to guide how they select, understand, and apply the necessary information. At the heart of this type of reading is an informed search for specific information that will enable the person to carry out a predetermined act — to do something that could not have been done without that information.

	Constructing, Extending, and Examining Meaning			
	Initial Understanding	**Developing an Interpretation**	**Personal Reflection and Response**	**Demonstrating a Critical Stance**
	Requires the reader to provide an initial impression or unreflected understanding of what was read.	Requires the reader to go beyond the initial impression to develop a more complete understanding of what was read.	Requires the reader to connect knowledge from the text with his/her own personal background knowledge. The focus here is on how the text relates to personal knowledge.	Requires the reader to stand apart from the text and consider it.
Reading for Literary Experience	What is the story/plot about?	How did the plot develop?	How did this character change your idea of _____?	Rewrite this story with _____ as a setting or _____ as a character.
	How would you describe the main character?	How did this character change from the beginning to the end of the story?	Is this story similar to or different from your own experiences?	How does this author's use of _____ (irony, personification, humor) contribute to _____?
Reading for Information	What does this article tell you about _____?	What caused this event?	What current event does this remind you of?	How useful would this article be for _____? Explain.
	What does the author think about this topic?	In what ways are these ideas important to the topic or theme?	Does this description fit what you know about _____? Why?	What could be added to improve the author's argument?
Reading to Perform a Task	What is this supposed to help you do?	What will be the result of this step in the directions?	In order to _____, what information would you need to find that you don't know right now?	Why is this information needed?
	What time can you get a non-stop flight to X?	What must you do before this step?	Describe a situation where you could leave out step X.	What would happen if you omitted this?

Types of Interactions With Text

Forming an Initial Understanding. When readers finish a piece, they are left with ideas and understandings they have built and changed over the course of reading, but that are still more or less uninspected. Initial understandings involve considering the text as a whole or in a broad perspective to reflect initial impressions, global understandings, any questions that may have arisen, and any hunches that might be considered.

Developing an Interpretation. Developing an interpretation occurs as readers extend their initial impressions and develop more thought-through and elaborated understandings of what they have read. It often involves reflecting on changes over time, exploring motivations, analyzing characters, and seeking explanations. Readers can link information across parts of a text as well as focus on specific information.

Personal Reflection and Response. Personal connections occur when readers relate their understandings and knowledge from the text to their own personal experiences and knowledge. It is from this perspective that background knowledge is used to enhance understanding (as readers agree, disagree, or at least are moved to reconsider what they already know) as well as lead to new understandings. Here, two kinds of connections can occur: prior knowledge can support the development of new understandings, and new understandings can also change background knowledge.

Demonstrating a Critical Stance. Demonstrating a critical stance requires readers to stand apart from the text and consider it objectively. It can involve critical evaluation, comparing and contrasting, and examining aspects of the author's craft. In this case, the student is not so much developing textual meaning, but rather is inspecting it.

The 1992 NAEP Reading Assessment

In continuing NAEP's responsiveness to increased knowledge about reading development and its implications for assessment, the 1992 reading assessment included a variety of unique features. Many of these new elements already have been incorporated in recent state, district, and classroom assessment reform efforts and have proven to be effective tools for measuring students' growth in reading literacy.[33] Because many reading

---

[33] O'Neil, J., "Putting Performance Assessment to the Test," *Educational Leadership*, 14-19, May, 1992.

Buechler, M., *Performance Assessment*, Policy Bulletin, No. PB-813 (Bloomington, IN: Indiana Educational Policy Center, 1992).

Moody, D., *Strategies for Statewide Student Assessment*, Policy Brief, No. 17 (Washington, DC: Office of Educational Research and Improvement, 1991)

curriculum programs now embrace a more integrative approach to instruction (e.g., linking reading and writing, making connections across texts), it was critical that the NAEP reading assessment reflect these current understandings of what students should be able to do in reading. These are represented in the following new features of the 1992 assessment.

- *The use of extended and short constructed questions as well as multiple-choice format.* Providing students with opportunities to construct their own responses allowed for varied interpretations that can result when students with different background experiences and knowledge make sense of the text from slightly different perspectives. Also, constructed-response questions made it possible to evaluate the depth of students' understanding.

- *A framework based on purpose and kinds of reading.* In recognition that readers approach texts differently based on aspects of the text and the perceived purpose for reading, the NAEP reading assessment measured students' abilities to read different types of materials for different purposes. As a result, the reading achievement of the nation's fourth-, eighth-, and twelfth- graders is reported by purpose for reading in addition to overall reading proficiency.

- *Complete and authentic texts that are used in real-life.* In the past, many reading assessments have measured students' comprehension of passages that were condensed or written to meet certain specifications for the assessment. The use of authentic texts, like in NAEP, brings the assessment situation closer to replicating real-life reading tasks.

- *Primary trait scoring of readers' understanding.* The use of primary trait scoring ensures that specific evidence of comprehension in students' responses is the focus of scoring. This allows for the establishing of scoring criteria based on reading and thinking demands rather than scoring based on a comparison of students' responses to each other.

- *Multiple texts related to the same task.* Because so many of the daily literacy demands and needs in today's society require making connections between different texts, the NAEP reading assessment provided opportunities for students to read different texts for the purpose of linking and integrating ideas across the texts.

- *A NAEP Reader for self-selected texts.* Some eighth- and twelfth-graders in the NAEP reading assessment were given the opportunity to make choices about the material they read. In doing so, this special study of reading for literary experience came much closer to representing authentic literacy events than is possible when students are given passages to read on a test whether or not they express interest in those materials.

- *A special oral reading and response study.* In response to the current expanding view of literacy development, a special study at fourth grade integrated measures of oral reading fluency and overall reading proficiency. In addition, the special study gave students an opportunity to respond to constructed-response questions with oral answers. These oral responses were then compared to the students' performance with written responses. This may increase our understanding of the influence of response mode on students' answers to constructed-response questions.

- *A special literacy interview assessment.* A one-on-one literacy interview was conducted with some fourth graders in the NAEP assessment to ascertain the extent and nature of their literacy habits and experiences. As a result, it is possible to examine how specific activities and attitudes may be related to overall reading proficiency.

Summary

In concert with education reform efforts emphasizing a closer link between instruction and assessment, NAEP developed a new and innovative reading assessment beginning in 1992. *The Reading Framework* underlying the assessment was based on research supporting purposeful, and integrated reading activities based on whole texts rather than short made-up sentences or paragraphs. It considered students' performance in situations that involved reading different kinds of materials for different purposes.

The 1992 reading assessment measured three global purposes for reading — reading for literary experience, reading to gain information, and reading to perform a task. (The third purpose for reading — reading to perform a task — was not assessed at grade 4.) Reading for literary experience usually involves reading novels, short stories, plays, and essays. In these reading situations, the reader explores or uncovers experiences through the text and considers the interplay among events, emotions, and

possibilities. Reading to gain information usually involves reading articles in magazines and newspapers, chapters in textbooks, entries in encyclopedias and catalogs, and books on particular topics. These reading situations call for different orientations to text from those in reading for literary experience because readers are specifically focused on acquiring information. Reading to perform a task involves reading various types of materials for the purpose of applying the information or directions in completing a specific task. Reading materials used for this purpose may include schedules, directions, or instructions for completing forms.

The *Reading Framework* asked students to build, extend, and examine text meaning from four stances or orientations. Initial understanding involved comprehending the overall or general meaning of the selection. Developing an interpretation required extending the ideas in the text by making inferences and connections. Reflection and personal response included making explicit connections between ideas in the text and student's own background knowledge and experience. Finally, students were asked to adopt a critical stance and consider how the author crafted the text.

One of the most crucial and unique attributes of NAEP's 1992 reading assessment is the use of authentic reading material. This chapter discusses the rationale underlying the use of authentic reading material, giving examples from the 1992 assessment. Unlike the materials that have been traditionally used to measure reading comprehension, these texts were not prepared especially for the assessment. Moreover, whole stories, articles, or selections from textbooks were used, rather than excerpts or abridgements.

Why Use Authentic Texts?

NAEP's decision to use only authentic texts reflected several issues and concerns, including consistency with the reform movement currently taking place in assessment. Significant efforts are being made to move assessment away from isolated, decontextualized testing of individual skills toward

what has been termed authentic or performance assessment.[34] Fueling this effort is a belief that the manner in which students are assessed should reflect the way they are taught. Consequently, if more complex, integrative abilities are the goal of education, then the assessment of what students have learned should mirror that goal and should require demonstration of these higher-order processes.[35] As teachers often do use assessment tasks to set priorities for what they teach, authentic assessments can contribute to good classroom practice. Reading assessments should therefore feature texts like those used for classroom practices — ones that may be interesting to students and promote thoughtful or engaged reading.

As emphasized in NAEP's *Reading Framework* (see Chapter One), reading is a thinking process that involves a complex interaction among the reader, the text, and the context in which something is read. In contrast to passages in more traditional assessments which are often highly abridged portions of whole texts, authentic texts provide students with realistic reading experiences that are more appropriate for assessing this interaction.

Studies have found that less traditional assessment formats may provide a better indication of students' interactions with texts and the processes that result in comprehension.[36] If passages are edited to make them conform to specifications about length, the amount of "argument" or "character motivation" that conforms to a particular structure, and a preference for concrete topics that can be objectively captured by assessment questions, the measurement of reading achievement may be seriously distorted.

---

[34] Mitchell, R., *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools* (New York, NY: The Free Press, 1992).

Berlak, H., et al., (Eds.) *Toward a New Science of Educational Testing and Assessment* (Albany, NY: State University of New York Press, 1992).

[35] Resnick, L.B., & Resnick, D.P., "Assessing the Thinking Curriculum: New Tools for Educational Reform," In Gifford, B.R., and O'Connor, M.C., *Changing Assessments: Alternative Views of Aptitude Achievement and Instruction* (Boston MA: Kluwer Academic Publishers, 1992).

Wolf, D.P., Lemahieu, P.G., & Eresh, J., "Good Measure: Assessment as a Tool for Educational Reform," *Educational Leadership*, 49(3), 14-19, 1992.

[36] Seda, I., "Assessment Format and Comprehension Performance," Paper presented at the 34th annual meeting of the International Reading Association (New Orleans, LA: 1989).

Everyday reading tasks, whether they occur in the classroom, at home, or in the workplace, demand a variety of skills suitable for texts that vary in complexity, abstractness, genre, and even appearance. In general, individuals need to be able to read and comprehend a wide array of texts for different purposes. Traditional assessments may fail to prepare students "... for real, 'messy' uses of knowledge in context — the 'doing' of a subject."[37] As the demands of the workplace in our changing economy become increasingly complex, reading tasks may become even more challenging than at present. Assessments that feature texts like those that people must read everyday may supply more information about the kinds of skills people use to read effectively.

Selecting the Assessment Texts

Consistent with NAEP's *Reading Framework*, the texts selected for the 1992 assessment were drawn from materials occurring naturally in the environments of students at grades 4, 8, and 12. Texts were drawn from a wide variety of sources, including books of short stories, magazines, textbooks, "how to" materials, and documents. In order to address concerns about the possibility of some students being familiar with these materials, texts were not drawn from basal readers, but from books and magazines. Articles taken from magazines were taken from those published in 1990 or earlier, because students assessed in 1992 were very unlikely to have been familiar with these magazines. (Even if some students were familiar with any of the materials in the assessment, the design of the questions accompanying the texts ensured that students had to carefully reread and reconsider the text in order to respond to the questions.)

Each text was chosen to reflect one of the three broadly-based reading purposes included in the assessment — for literary experience, to gain information, and to perform a task. Rather than using conventional readability estimates, teachers judged the difficulty of the texts according to length, complexity of arguments, abstractness of concepts, unusual points of view, and shifting time frames. NAEP made sure that assessment reading materials would meet high instructional standards by sending the pool of texts initially selected by teachers and teacher educators to another set of teachers for review. Teacher reviewers were recruited through state

[37] Wiggins, G., Assessing Students Performance; Exploring the Purpose and Limits of Testing, *Authenticity, Context, and Validity*, pp. 207-208 (San Francisco: CA, Jossey-Bass Publishers, 1993).

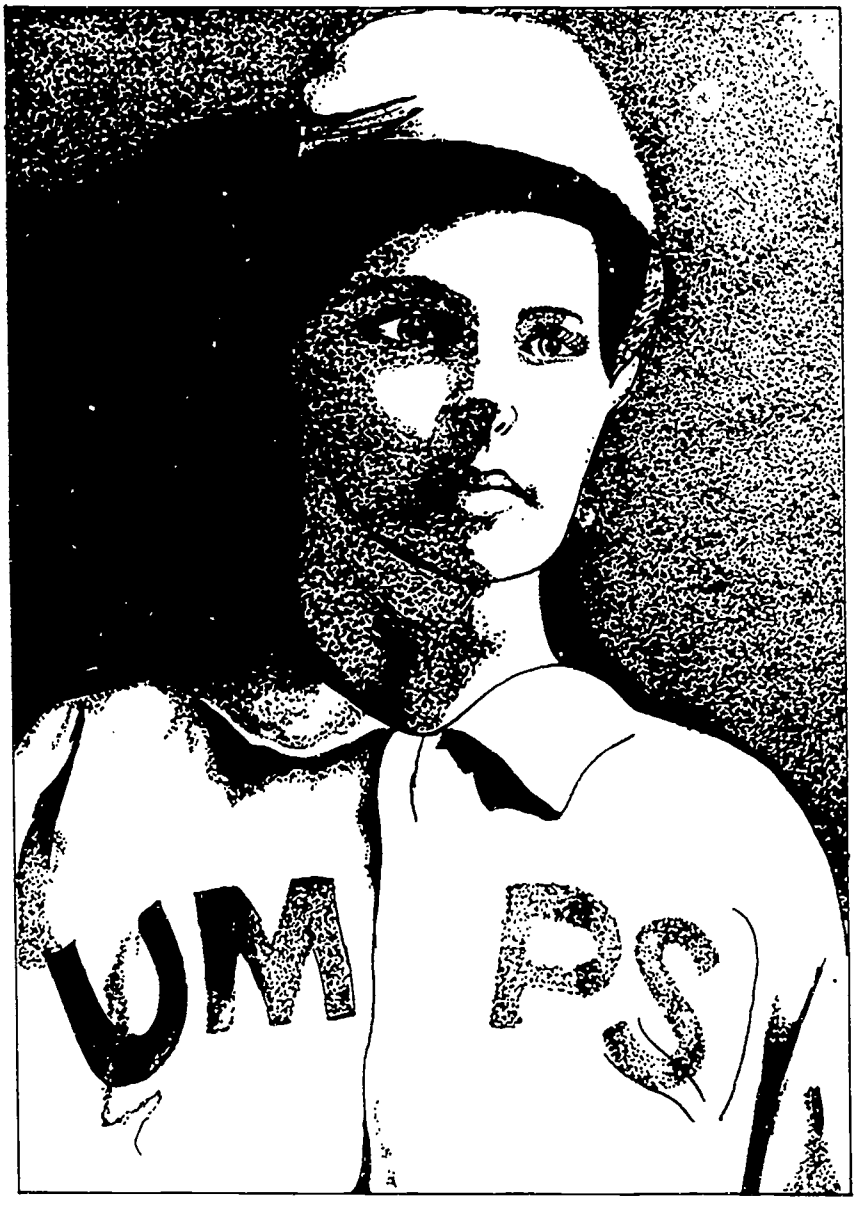
assessment programs and reading supervisors. They reviewed these passages with several aims: to confirm a grade level appropriateness, to indicate whether or not a text was suitable developmentally or culturally, to evaluate a text's structure and cohesiveness, and to confirm that the text occurred in students' environments either inside or outside of school.

Examples of Texts Used in the 1992 Assessment

The following article is an example of the material used in the fourth-grade assessment. This text, *Amanda Clement: The Umpire in a Skirt*, is representative of the informational material children read in terms of level of difficulty, genre, and content. Note that the picture accompanying the article appeared in the assessment, just as it did in the original source.

Amanda Clement: The Umpire in a Skirt

Marilyn Kratz



IT WAS A HOT SUNDAY AFTERNOON in Hawarden, a small town in western Iowa. Amanda Clement was sixteen years old. She sat quietly in the grandstand with her mother, but she imagined herself right out there on the baseball diamond with the players. Back home in Hudson, South Dakota, her brother Hank and his friends often asked her to umpire games. Sometimes she was even allowed to play first base.

Today, Mandy, as she was called, could only sit and watch Hank pitch for Renville against Hawarden. The year was 1904, and girls were not supposed to participate in sports. But when the umpire for the preliminary game between two local teams didn't arrive, Hank asked Mandy to make the calls.

Mrs. Clement didn't want her daughter to umpire a public event, but at last Hank and Mandy persuaded her to give her consent. Mandy eagerly took her position behind the pitcher's mound. Because only one umpire was used in those days, she had to call plays on the four bases as well as strikes and balls.

Mandy was five feet ten inches tall and looked very impressive as she accurately called the plays. She did so well that the players for the big game asked her to umpire for them—with pay!

Mrs. Clement was shocked at that idea. But Mandy finally persuaded her mother to allow her to do it. Amanda Clement became the first paid woman baseball umpire on record.

Mandy's fame spread quickly. Before long, she was umpiring games in North and South Dakota, Iowa, Minnesota, and Nebraska. Flyers, sent out to announce upcoming games, called Mandy the "World Champion Woman Umpire." Her uniform was a long blue skirt, a black necktie, and a white blouse with UMPS stenciled across the front. Mandy kept her long dark hair tucked inside a peaked cap. She commanded respect and attention—players never said, "Kill the umpire!" They argued more politely, asking, "Beg your pardon, Miss Umpire, but wasn't that one a bit high?"

Mandy is recognized in the Baseball Hall of Fame in Cooperstown, New York; the Women's Sports Hall of Fame; and the Women's Sports Foundation in San Francisco, California. In 1912 she held the world record for a woman throwing a baseball: 279 feet.

Mandy's earnings for her work as an umpire came in especially handy. She put herself through college and became a teacher and coach, organizing teams and encouraging athletes wherever she lived. Mandy died in 1971. People who knew her remember her for her work as an umpire, teacher, and coach, and because she loved helping people as much as she loved sports.

The Amanda Clement article required fourth-grade students to work with the biographical genre. Students had to sort through factual information about Amanda Clement in order to answer questions about the passage. (Please see examples of these questions in Chapter Four).

Other materials used in the assessment required fourth graders to read and examine different genres. For example, students also read and answered questions about an African folk tale called *Hungry Spider and the Turtle*. The tale is from a collection of West African stories like those often used in classrooms. The story is about how the character of Turtle, tricked out of a meal while relying on the hospitality of Spider, finds a way to cleverly teach the greedy Spider a lesson. Students, in order to understand the story, had to grasp the different points of view of Spider and Turtle, follow shifts in the time frame, and extract the lesson the story teaches from the course of action and conversation between Spider and Turtle.

Another fourth-grade article called *Blue Crabs* describes the habits and appearance of the blue crab, and explains how the crabs are captured. The article was taken from the journal *Highlights*; its combination of narrative and expository writing and its focus on animals is typical of the kind of reading young children do in and outside of school. The material included some basic scientific information, which is appropriate for students who must read for information across different content areas. The article also features illustrations of the crab and of a mechanism used to trap them. Students could use the illustrations to help them understand the text of the article, as they would have done if they had been reading the piece in its original source in school or at home.

Sometimes students were asked to integrate information across pieces, as in the following texts used at the eighth grade. Students were asked to read all three pieces, and were required to examine the story by Anne Frank, *Cady's Life*, in the light of the biographical information given in the box, and the poem. The Cady's Life set of texts and questions thus asked students to perform tasks much like those they might perform in a classroom; students looked at information about an author before reading that author's work, and sought to better understand one author by reading the writing of another. Also, as in the classroom, students were required to use more than one genre to think about a single topic.

THE FICTION OF

ANNE FRANK

ANNE FRANK

is best known as the writer of *Anne Frank: The Diary of a Young Girl*. She kept this diary while she, her parents, her sister, and four other Jews hid in the "Secret Annex" (the attic of a building in Holland) to escape persecution by Hitler and the Nazis during World War II. Anne was thirteen years old when she began keeping her diary on June 12, 1942. Two years later, in August 1944, the Nazis raided the Annex. Anne died seven or eight months later in a concentration camp. She was fifteen years old.

Anne's diary was first published in 1947. Since then it has been translated and published throughout the world. Through the publication of her diary, Anne has come to symbolize to the world the six million Jews killed by the Nazis.

Although Anne's diary is read throughout the world, her fiction is not as well known. In 1943-1944, Anne wrote a number of stories and began a novel, now published in *Tales from the Secret Annex*. Anne states in her diary that she wanted to be a famous writer. Her fiction, like her diary, shows that she was indeed talented. The following excerpt is from her unfinished novel, *Cady's Life*.

CADY'S LIFE

by Anne Frank

was a hard time for the Jews. The fate of many would be decided in 1942. In July they began to round up boys and girls and deport them. Luckily Cady's girl friend Mary seemed to have been forgotten. Later it wasn't just the young people, no one was spared. In the fall and winter Cady went through terrible experiences. Night after night she heard cars driving down the street, she heard children screaming and doors being slammed. Mr. and Mrs. Van Altenhoven looked at each other and Cady in the lamplight, and in their eyes the question could be read: "Whom will they take tomorrow?"

One evening in December, Cady decided to run over to Mary's house and cheer her up a little. That night the noise in the street was worse than ever. Cady rang three times at the Hopkens's and when Mary came to the front of the house and looked cautiously out of the window, she called out her name to reassure her. Cady was let in. The whole family sat waiting in gym suits, with packs on their backs. They all looked pale and didn't say a word when Cady stepped into the room. Would they sit there like this every night for months? The sight of all these pale, frightened faces was terrible. Every time a door slammed outside, a shock went through the people sitting there. Those slamming doors seemed to symbolize the slamming of the door of life.

At ten o'clock Cady took her leave. She saw there was no point in her sitting there, there was nothing she could do to help or comfort these people, who already seemed to be in another world. The only one who kept her courage up a little was Mary. She nodded to Cady from time to time and tried desperately to get her parents and sisters to eat something.

Mary took her to the door and bolted it after her. Cady started home with her little flashlight. She hadn't taken five steps when she stopped still and listened: she heard steps around the corner, a whole regiment of soldiers. She couldn't see much in the darkness, but she knew very well who was coming and what it meant. She flattened herself against a wall, switched off her light, and hoped the men wouldn't see her. Then suddenly one of them stopped in front of her, brandishing a pistol and looking at her with threatening eyes and a cruel face. "Come!" That was all he said, and immediately she was roughly seized and led away.

"I'm a Christian girl of respectable parents," she managed to say. She trembled from top to toe and wondered what this brute would do to her. At all costs she must try to show him her identity card.

"What do you mean respectable? Let's see your card."

Cady took it out of her pocket.

"Why didn't you say so right away?" the man said as he looked at it. "*So ein Lumpenpack*!"* Before she knew it she was lying on the street. Furious over his own mistake, the German had given the "respectable Christian girl" a violent shove. Without a thought for her pain or anything else, Cady stood up and ran home.

After that night a week passed before Cady had a chance to visit Mary. But one afternoon she took time off, regardless of her work or other appointments. Before she got to the Hopkens's house she was as good as sure she wouldn't find Mary there, and, indeed, when she came to the door, it was sealed up.

Cady was seized with despair. "Who knows," she thought, "where Mary is now?" She turned around and went straight back home. She went to her room and slammed the door. With her coat still on, she threw herself down on the sofa, and thought and thought about Mary.

Why did Mary have to go away when she, Cady, could stay here? Why did Mary have to suffer her terrible fate when *she* was left to enjoy herself? What difference was there between them? Was she better than Mary in any way? Weren't they exactly the same? What crime had Mary committed? Oh, this could only be a terrible injustice. And suddenly she saw Mary's little figure before her, shut up in a cell, dressed in rags, with a sunken, emaciated face. Her eyes were very big, and she looked at Cady so sadly and reproachfully. Cady couldn't stand it anymore, she fell on her knees and cried and cried, cried till her whole body shook. Over and over again she saw Mary's eyes begging for help, help that Cady knew she couldn't give her.

"Mary, forgive me, come back . . ."

Cady no longer knew what to say or to think. For this misery that she saw so clearly before her eyes there were no words. Doors slammed in her ears, she heard children crying and in front of her she saw a troop of armed brutes, just like the one who had pushed her into the mud, and in among them, helpless and alone, Mary, Mary who was the same as she was.

*"*Such a bunch of scoundrels.*"

I AM ONE

I am only one,
But still I am one.
I cannot do everything,
But still I can do something;
And because I cannot do everything
I will not refuse to do the something that I can do.

—EDWARD EVERETT HALE

Another eighth-grade selection called *Creating a Time Capsule* was a step-by-step explanation of how to put together a time capsule that would inform future generations about life in the twentieth century. Reading to follow directions is a task suitable for this age group and one that is encountered in a variety of contexts. The article was taken from the journal *Cobblestone*, a history magazine designed for young people, and so is illustrative of the kind of reading eighth-grade students might do in school or at home.

A Ray Bradbury short story used at both the eighth and twelfth grades was brief but quite challenging. The story conveys character motives through dialogue and symbol; its difficulty lies in its economical use of these devices. The story is appropriate for an innovative assessment because students' abilities to understand these literary tools are a crucial aspect of reading, both in school and for personal pleasure. Moreover, the story's genre is similar to much science fiction, familiar to and popular with this age group, and its examination of issues surrounding the merits of technology is part of our contemporary context.

As shown below, the journal entry by an officer who fought in the Battle of Shiloh during the United States Civil War, juxtaposed with the encyclopedia entry about the Battle of Shiloh, provided twelfth graders the opportunity for a variety of comparisons.

In order to respond to the questions accompanying the texts, students needed to grasp how each text presented useful information about the same topic, but in very different ways. Students frequently must use a variety of sources to complete research projects in school, and the ability to work effectively with information from different sources is a critical part of reading for information. While working with the Shiloh materials, students had to grasp how one text, the journal entry, shed light on what it was like to actually fight in the war, while the encyclopedia entry provided factual information necessary to place the journal in context.

THE CIVIL WAR IN THE UNITED STATES: THE BATTLE OF SHILOH

Here are two perspectives on the battle of Shiloh which was part of the American Civil War. Each of the two passages was taken from a different source; the first is from a soldier's journal and the second is from an encyclopedia. Read them and see how each passage makes a contribution to your understanding of the battle of Shiloh and the Civil War. Think about what each source tells you that is missing from the other source, as well as what each one leaves out.

Journal Entry

The following journal entry relates the noise, confusion, and horror of the battle of Shiloh as told by a Union officer.

On the evening of the 5th, the 18th Wisconsin infantry arrived and were assigned to General Prentiss' division, on the front. They cooked their first suppers in the field that night at nine o'clock, and wrapped themselves in their blankets, to be awakened by the roar of battle, and receive, thus early, their bloody baptism. Before they had been on the field one day, their magnificent corps was decimated, most of the officers killed.

On going to the field the second day, our regiment strode on in line over wounded, dying, and dead. My office detaching me from the lines, I had an opportunity to notice incidents about the field. The regiment halted amidst a gory, ghastly scene. I heard a voice calling, "Ho, friend! ho! Come here." I went to a pile of dead human forms in every kind of stiff contortion; I saw one arm raised, beckoning me. I found there a rebel, covered with blood, pillowing his head on the dead body of a comrade. Both were red from head to foot. The live one had lain across the dead one all that horrible, long night in the storm. The first thing he said to me was "Give me some water. Send me a surgeon—won't you! What made you come down here to fight us? We never would have come up there." And then he affectionately put one arm over the form, and laid his bloody face against the cold, clammy, bloody face of his friend.

I filled his canteen nearly—reserving some for myself—knowing that I might be in the same sad condition. I told him we had no surgeon in our regiment, and that we would have to suffer, if wounded, the same as he; that other regiments were coming, and to call on them for a surgeon; that they were humane.

"Forward!" shouted the Colonel; and 'Forward' was repeated by the officers. I left him.

The above recalls to mind one of the hardest principles in warfare—where your sympathy and humanity are appealed to, and from sense of expediency, you are forbidden to exercise it. After our regiment had been nearly annihilated, and were compelled to retreat under a galling fire, a boy was supporting his dying brother on one arm, and trying to drag him from the field and the advancing foe. He looked at me imploringly, and said, "Captain, help him—won't you? Do, Captain; he'll live." I said, "He's shot through the head; don't you see? and can't live—he's dying now." "Oh, no, he ain't, Captain. Don't leave me." I was forced to reply, "The rebels won't hurt him. Lay him down and come, or both you and I will be lost." The rush of bullets and the yells of the approaching enemy hurried me away—leaving the young soldier over his dying brother.

At home I used to wince at the sight of a wound or of a corpse; but here, in one day, I learned to be among the scenes I am describing without emotion. My friend and myself, on the second night, looking in the dark for a place to lie down, he said, 'Let's lie down here. Here's some fellows sleeping.' We slept in quiet until dawn revealed that we had passed the night among sprawling, stiffened, ghastly corpses. I saw one of our dead soldiers with his mouth crammed full of cartridges until the cheeks were bulged out. Several protruded from his mouth. This was done by the rebels. On the third day most of our time was employed in burying the dead. Shallow pits were dug, which would soon fill with water. Into these we threw our comrades with a heavy splash, or a dump against solid bottom. Many a hopeful, promising youth thus indecently ended his career.

I stood in one place in the woods near the spot of the engagement of the 57th Illinois, and counted eighty-one dead rebels. There I saw one tree, seven inches in diameter, with thirty-one bullet holes. Such had been death's storm. Near the scenes of the last of the fighting, where the rebels precipitately retreated, I saw one grave containing one hundred and thirty-seven dead rebels, and one side of it another grave containing forty-one dead Federals.

One dead and uniformed officer lay covered with a little housing of rails. On it was a fly-leaf of a memorandum-book with the pencil writing: 'Federals, respect my father's corpse.' Many of our boys wanted to cut off his buttons and gold cord; but our Colonel had the body religiously guarded.

My poor friend, Carson, after having fought and worked, and slaved from the beginning of the war, unrequited, comparatively, and after having passed hundreds of hair-breadth escapes, and through this wild battle was killed with almost the last shot. A round shot took off his whole face and tore part of his head. Poor Carson! We all remember your patriotism, your courage, your devotion. We will cheer, all we can, the bereaved and dear ones you have left.

Encyclopedia Entry

The last account you will read of the battle comes from an encyclopedia.

SHILOH, Battle of, shīlō, one of the most bitterly contested battles of the American Civil War, fought on April 6 and 7, 1862, in southern Tennessee, about 100 miles (160 km) southwest of Nashville. The first great battle of the war had been fought at Bull Run (Manassas) in Virginia in July 1861, nearly a year before. It had ended in a temporary stalemate in the eastern theater. In the West, Kentucky tried to remain neutral, but by the end of 1861 both sides had sent troops into the state.

In February 1862, Union General Ulysses S. Grant captured forts Henry and Donelson on the Tennessee and Cumberland rivers in northern Tennessee near the Kentucky boundary, taking about 11,500 men and 40 guns. The whole Confederate line of defense across Kentucky gave way. The Confederates were forced to retreat to Murfreesboro, Tenn., southeast of Nashville, as other Union forces moved toward Nashville.

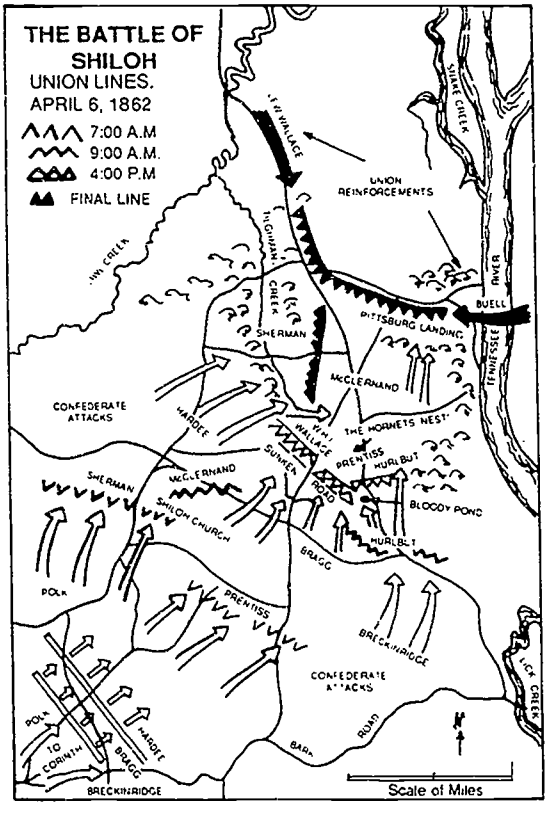
With the Southern press clamoring for his removal, General Albert Sidney Johnston, commanding the Confederate forces in the region, began to assemble the scattered troops. He decided to designate Corinth, in the northeast corner of Mississippi, as the concentration point for the army.

Assembling of the Armies. By the end of March, Johnston and his second-in-command, General Pierre G.T. Beauregard, managed to gather in Corinth more than 40,000 men, including a few units from as far away as the Gulf of Mexico. These were organized into three corps, commanded by Generals Leonidas Polk, Braxton Bragg, and William J. Hardee. There was also a small reserve corps under General John C. Breckinridge.

Meanwhile, General Henry W. Halleck, who was Grant's department commander, had ordered Grant's troops to make a reconnaissance southward along the Tennessee River. They encamped near Pittsburg Landing, on the west side of the river, about 5 miles (8 km) north of the Mississippi boundary. There they awaited the arrival of another large Union force under General Don Carlos Buell, which had been ordered southward from Nashville to join them.

Grant's army of 42,000 men was divided into six divisions. Five of these, a total of 37,000, were near Pittsburg Landing. One division, under General Lew Wallace's command, was stationed 6 miles (9 km) to the north. Buell's army marching from Nashville was almost as large as Grant's; together they would far outnumber the concentration of forces that the Confederates could put in the field.

General Johnston saw that he must strike Grant's army before Buell arrived. The Confederates started northward from Corinth on the afternoon of April 3, intending to attack at dawn on the 5th, but a violent rainstorm turned the dirt roads into a sea of mud. The attack was postponed from the 5th to Sunday, April 6, but on the 5th the



leading division of Buell's army arrived on the other side of the Tennessee River, only 7 miles (11 km) away.

That night the armies encamped only 2 miles (3 km) apart, with the Union forces, whose advanced units were about 4 miles (6 km) west of the river, wholly unaware of their danger. Neither they nor their leaders expected an attack. They were not disposed for defense, nor had any trenches been dug for their protection. Early in the morning of April 6, a suspicious brigade commander in General Benjamin M. Prentiss' Union division sent a small force forward to investigate the nearby woods. At dawn they exchanged shots with the Confederate outpost, but it was too late to give warning of the attack, which burst on the Union camps.

Confederate Attack. For the assault, General Johnston had chosen an unusual formation. He formed his troops in three lines, with Hardee's corps in front, Bragg's corps in a second parallel line, and then Polk's and Breckenridge's reserve corps.

The Confederates charged straight to their front into the divisions of Prentiss and General William Tecumseh Sherman, who held the right flank near the Old Shiloh Church. They and General John A. McClernand's division made a brief stand. Many men fought valiantly, but others broke and fled. When Grant, who had been absent from the field, arrived he found all five of the divisions fighting desperately in what seemed like a hopeless struggle. He had already sent for Buell's troops, and now he sent for Lew Wallace to join him.

The Union forces had retreated about halfway to the river to a new position, naturally strong, with open fields on each side and a sunken road in front. Here, in the center, in a position known to history as "The Hornets' Nest," the Confederates were halted for hours. They could not take it by assault, but gradually the Union troops on each flank were forced back. Johnston fell mortally wounded. Beauregard took command, and the attack continued.

Finally "The Hornets' Nest" was surrounded. General William H.L. Wallace was killed trying to lead his division out. Prentiss was forced to surrender, but time was running out for the Confederates. They made a last attack on the Union left toward Pittsburg Landing to cut off the escape of the Union forces, but Buell's troops were now arriving.

Union Counterstroke. On the next day, Grant attacked. Of the soldiers who had fought on the first day, he had only about 7,000 effectives, (soldiers ready for battle), but Lew Wallace had arrived with his 5,000, and Buell had supplied 20,000 more. To oppose these, the Confederates could muster only about 20,000 men. For hours they held the line in front of Shiloh Church, but at last they withdrew in good order from the field.

The Battle of Shiloh, the second great battle of the war, was a tremendous shock to the people of the North and the South. When the reports were published, they found that each side had lost about 25% of the troops engaged—the Confederates about 10,700, the Union more than 13,000. The people suddenly realized that this was to be a long and bloody war.

One of the most innovative twelfth-grade selections required students to perform a reading task that most adults in the United States must undertake — that of understanding a federal income tax form. Students were given an unrevised 1040EZ form, and asked a series of questions about how to go about filling it in. Finally, they were asked to actually complete the form. The kinds of reading skills necessary for filling in the form are those used everyday by people in both school and employment settings. Such skills include the ability to understand directions and to apply those directions to the performance of a task; the ability to perform a complex task in the appropriate sequence; and the ability to integrate tabular and graphic information with textual information for the purpose of performing a task.

Another twelfth-grade informational text was a long article about sperm whales taken from the journal *Natural History*. Students were given 50 minutes to read the article and to answer questions about it. The article presents complicated scientific information, and its length and somewhat technical style are typical of the kind of material older high school students might be required to read across various content areas. The text also included illustrations, and was arranged, as in the original source, in the double column format characteristic of some journals.

Diversity in Assessment Materials

All of the authentic texts described above are like what students encounter in their classrooms and when they read on their own. One of the texts described above, the tale called *Hungry Spider and the Turtle*, was a good choice for use on the assessment not just because it represents a frequently read genre, but also because it is an example of literature from a different cultural tradition. Other texts not already mentioned that were used in the assessment were also distinguished by their focus on the experiences of people from various backgrounds. For example, an article for fourth and eighth graders discussed the experiences of European immigrants when they arrived at Ellis Island. A brief piece for eighth- and twelfth-grade students presented the story of a man who fled with his family from communist North Vietnam.

Given the current emphasis in today's classrooms on the literatures and experiences of racial and ethnic minorities, students are more likely to have read multicultural literary works. Because part of the use of authentic materials for NAEP meant choosing texts that would reflect classroom and outside reading, an effort was made to include the range of material

students might read in and out of school. The NAEP Reader, discussed in Chapter Six, was especially designed to include authentic materials that represented a range of content, genre, and cultural distinctiveness. Students received a diverse variety of short stories, and were given the opportunity to select the one they wanted to answer questions about (see Chapter Eight for further information).

Summary

NAEP's 1992 reading assessment represented an innovative effort to measure the reading achievement of our nation's students in grades 4, 8, and 12. The naturally occurring reading materials used in the assessment are regarded as a crucial aspect of this innovative effort, and are viewed as the most appropriate assessment instruments for measuring the reading ability of students. The use of authentic texts, that provided more realistic reading experiences than previous reading assessments, reflects several current concerns. Among these are a new focus on performance assessment and the utility of assessments accurately reflecting classroom practices and goals, an understanding of reading as a complex process with many components, the virtues of non-traditional assessment formats, and the importance of including diverse materials for assessment use.

Students learn to become thoughtful readers by being expected to form their own initial understandings of what they have read, by reflecting on and reforming these initial ideas into more fully developed understandings, by enriching their growing understandings through personal reflections, and by considering both the text and their own understandings in an objective and critical manner. It is through this process of constructing and extending meaning that higher levels of reading literacy are achieved.

Previous NAEP assessments have reported that in general, students can build superficial and straightforward understandings but have difficulty developing more thoughtful and elaborated responses. These findings have resulted in a call for more authentic assessments as well as for instructional tasks that require students to read complete (as opposed to abridged or rewritten) texts for more complex goals. Question formats where students are required to construct their own written responses provide students with the opportunities to present and explain their understandings.

Constructed-Response Questions in the NAEP Reading Assessment

In response to the cumulative research on comprehension as well as to the long-lived national reform effort to improve students' abilities to reason effectively, NAEP always has included constructed-response questions in its assessments of reading achievement, in addition to multiple-choice questions. The assessment of reading comprehension using multiple-choice questions has been commonplace for some time, both in classroom tests as well as in large-scale measures. Some advantages to multiple-choice testing that are typically cited include the objectivity of scoring and the ability to have broader content coverage since questions can be answered more quickly, thus, more questions can be included on the test.[38]

More recently, many educators in the field of reading have expressed concerns that multiple-choice questions may not fully capture the diversity of students' interpretations and perspectives in their reading experiences.[39] Furthermore, recent conceptualizations of the reading process portray comprehension as meaning construction. That is, readers use ideas from the text to build meaning based on ideas and experiences that they bring to the reading situation.[40] These newer understandings have resulted in a call for more constructed-responses questions in tests of reading comprehension, so that students may demonstrate their ability to construct meaning and to support their own interpretations.[41] The newly developed 1992 assessment substantially increased the number of questions requiring fourth, eighth, and twelfth graders to reflect on and write about their understandings. Two constructed-response formats were included, together comprising the majority of the assessment questions and from 60 to 70 percent of the students' response time.

---

[38] Bennett, R.E., & Ward, W.C., (Eds.), *Construction Versus Choice in Cognitive Measurement* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1993).

[39] Farr, R., "Putting it all Together: Solving the Reading Assessment Puzzle," *The Reading Teacher, 46,* 26-37, 1992.

[40] Ruddell, R.B., & Unrau, N.J., "Reading as a Meaning-Construction Process: The Reader, The Text, and The Teacher," In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.). *Theoretical Models and Processes of Reading*, 996-1056 (Newark, DE: International Reading Association, 1994).

[41] Valencia, S., & Pearson, P.D., "Reading Assessment: Time for a Change," *The Reading Teacher, 40,* 726-732, 1987.

The first type — short constructed-response questions — prompted students to think and write briefly about their understandings. The second type — extended constructed-response questions — were designed to prompt greater thought and reflection and hence required somewhat longer explanations. In comparison to the multiple-choice questions that required students to select among an array of already developed responses, both constructed-response question types required students to generate their own ideas in response to the questions and to communicate them in writing. Each answer to the short constructed-response questions was scored as either acceptable or unacceptable. Responses to the extended questions were evaluated according to a four-point scale as: extensive, essential, partial, or unsatisfactory. The assessment results were analyzed by ETS to determine the percentage of students responding correctly to each multiple-choice or short constructed-response question and the percentage of students responding at each of the four score levels for the extended constructed-response questions.

Average Performance on Constructed-Response Questions

Short-constructed response questions required students to write a phrase or a sentence or two of global observations, general conclusions, or basic interpretations. While such questions do not demand of students the same depth of understanding and length of response required for extended-response questions, they nevertheless did require students to probe the text and generate their own thoughtful responses about what they have read. Short constructed-response questions are therefore useful for measuring how students are engaging in various reading processes, for example, analyzing and critically considering a text, or bringing their own knowledge and experiences to their understanding of a text.

In general, the extended questions required the students to think beyond their initial impressions, to more fully consider various aspects of the piece and their reactions to it, and to discuss their ideas. In short, such items required students to engage in extended thought and language.

More specifically, these questions prompted the students to consider and explain the larger significance of what they had read, to make and explain connections between what they had read and real-life situations; to project and explain situations from others' points of view (both within and outside the text); to relate important information or situations to outcomes,

ideas, and emotions; or to examine the relevance of their own responses in relation to the pieces they had read. Thus, each extended-response question invited the students to revisit their thinking about the text in order to consider possibilities and develop further understandings — and to explain them in ways that provided evidence of the kinds of thoughtful understanding appropriate to the piece, the reading purpose, and the question.

This chapter will present information regarding fourth-grade students' overall performance on three question types — multiple-choice, short constructed-response, and extended response. Sample constructed-response questions (both short and extended) will be presented along with actual responses provided by students in the NAEP assessment, illustrating the range of performance on each type of question. In addition, the specific reading processes and competencies that are displayed in students' constructed responses will be discussed.

The 1992 NAEP reading assessment included a state-by-state assessment at the fourth grade in addition to the national assessments at grades 4, 8, and 12. Therefore, the performance of fourth graders in participating states is presented along with the national results in this chapter. Since individual state involvement in the state-by-state assessment was voluntary, only the participating 43 jurisdictions are included in this presentation. Comparable state data are not available for grades 8 and 12, and are not presented in Chapters 4 and 5.

Table 3.1 presents the average percentage of successful responses from fourth graders for each of the three types of questions. In general, as would be anticipated, students had the greatest difficulty with the extended-response questions.

Table 3.1

Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 4, 1992 Reading Assessment

	EXTENDED RESPONSE	SHORT CONSTRUCTED-RESPONSE	MULTIPLE-CHOICE
	Average Percentage Essential or Better	Average Percentage Acceptable	Average Percentage Correct
Nation	25 (0.7)	52 (0.6)	63 (0.5)
Region			
Northeast	29 (2.2)	55 (2.4)	65 (1.7)
Southeast	23 (1.5)	50 (1.5)	61 (1.3)
Central	25 (0.8)	52 (0.6)	63 (0.6)
West	24 (1.1)	50 (0.8)	62 (0.6)
Race/Ethnicity			
White	28 (0.8)	55 (0.7)	66 (0.6)
Black	14 (0.9)	38 (1.3)	50 (0.8)
Hispanic	19 (1.4)	43 (1.1)	57 (1.0)
Gender			
Male	22 (0.8)	49 (0.8)	61 (0.6)
Female	28 (0.9)	54 (0.7)	64 (0.5)
Type of Community			
Advantaged Urban	35 (2.3)	63 (2.2)	73 (1.1)
Disadvantaged Urban	12 (1.2)	34 (1.8)	48 (1.3)
Extreme Rural	24 (1.5)	52 (1.7)	63 (1.2)
Other	25 (0.8)	52 (0.7)	63 (0.6)
Type of School			
Public	24 (0.7)	50 (0.7)	62 (0.5)
Private*	32 (1.6)	60 (1.0)	69 (1.1)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private school . he sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

On average, slightly fewer than two-thirds of the fourth-grade students (63 percent) provided correct answers to the multiple-choice items. The short-constructed response questions posed somewhat more difficulty, with students averaging 52 percent acceptable responses. The questions requiring extended responses proved to be the most difficult; students averaged only 25 percent essential or better responses across these questions.

Although average performance differed across the three types of questions, the **relative** performance of students in various demographic subgroups was quite similar (Table 3.1). On each set of questions, females had higher average performance than males. Among types of communities, students from advantaged urban communities had the highest average performance and students from extreme rural and "other" communities had higher average performance than students from disadvantaged communities. White students had higher average performance than Black or Hispanic students and Hispanic students had higher average performance than Black students. Also, private school students out-performed their public school counterparts on all question types. The only differences which appeared to vary systematically with question type were for gender; the performance gap between males and females was smallest for multiple-choice questions (3 percent) and largest for extended-response questions (6 percent).

Table 3.2 presents fourth graders' average performance on the three question types for states participating in the state-by-state assessment. The state assessments only included students attending public schools, in contrast to the national assessment which also included private school students. Thus, the national and regional results provided for comparison with the state data are based only on students attending public schools.

According to the state results, a pattern similar to the national average was observed. Students demonstrated the highest performance on multiple-choice questions, lower performance on short constructed-response, and the lowest on extended response questions. Across the states, the average percentage of correct responses to multiple choice questions ranged from 49 percent to 68 percent, the average percentage of acceptable responses to short constructed-response questions ranged from 34 percent to 59 percent, and the average percentage of essential or better responses to extended questions ranged from 12 percent to 32 percent. Students in all participating states demonstrated the most difficulty in providing extended responses.

Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 4, 1992 Trial State Reading Assessment

Public Schools	EXTENDED RESPONSE Average Percentage Essential or Better	SHORT CONSTRUCTED-RESPONSE Mean Percentage Acceptable	MULTIPLE-CHOICE Mean Percentage Correct
Nation	24 (0.7)	50 (0.7)	62 (0.5)
Northeast	28 (2.3)	55 (2.6)	65 (1.8)
Southeast	22 (1.7)	48 (1.5)	59 (1.4)
Central	24 (1.0)	51 (0.6)	63 (0.6)
West	23 (1.4)	49 (0.9)	61 (0.7)
States			
Alabama	21 (0.9)	48 (1.0)	57 (0.9)
Arizona	22 (0.6)	48 (0.7)	60 (0.7)
Arkansas	22 (0.7)	49 (0.8)	60 (0.6)
California	23 (0.8)	46 (1.1)	58 (0.9)
Colorado	26 (0.8)	52 (0.8)	63 (0.5)
Connecticut	29 (1.0)	56 (0.8)	66 (0.6)
Delaware	25 (0.9)	50 (0.6)	61 (0.5)
District of Columbia	15 (0.6)	37 (0.5)	51 (0.4)
Florida	23 (0.7)	48 (0.7)	60 (0.6)
Georgia	24 (0.9)	50 (0.9)	61 (0.8)
Hawaii	20 (0.9)	46 (1.0)	56 (0.7)
Idaho	26 (0.7)	54 (0.6)	64 (0.5)
Indiana	27 (0.7)	56 (0.8)	64 (0.7)
Iowa	29 (0.7)	58 (0.7)	67 (0.6)
Kentucky	24 (0.8)	49 (0.7)	61 (0.7)
Louisiana	19 (0.8)	45 (0.7)	56 (0.7)
Maine	30 (0.9)	59 (0.6)	68 (0.6)
Maryland	25 (0.7)	51 (0.8)	60 (0.8)
Massachusetts	30 (0.8)	59 (0.6)	68 (0.6)
Michigan	26 (1.2)	52 (0.8)	63 (0.8)
Minnesota	27 (0.8)	55 (0.8)	66 (0.6)
Mississippi	18 (0.6)	42 (0.8)	55 (0.6)
Missouri	26 (0.8)	54 (0.7)	64 (0.6)
Nebraska	28 (0.7)	55 (0.7)	65 (0.5)
New Hampshire	32 (0.8)	59 (0.8)	68 (0.7)
New Jersey	30 (1.1)	57 (0.9)	67 (0.9)
New Mexico	24 (0.8)	49 (1.1)	61 (0.7)
New York	26 (0.7)	53 (0.8)	63 (0.6)
North Carolina	24 (0.7)	50 (0.8)	60 (0.5)
North Dakota	28 (0.7)	58 (0.7)	67 (0.5)
Ohio	27 (0.7)	53 (0.7)	63 (0.7)
Oklahoma	25 (0.8)	55 (0.7)	65 (0.6)
Pennsylvania	29 (0.9)	55 (0.8)	65 (0.7)
Rhode Island	25 (0.8)	53 (1.0)	64 (0.9)
South Carolina	23 (0.8)	48 (0.8)	60 (0.6)
Tennessee	23 (0.9)	50 (0.8)	61 (0.7)
Texas	24 (0.7)	51 (0.9)	61 (0.8)
Utah	27 (0.8)	55 (0.8)	65 (0.6)
Virginia	28 (0.9)	55 (0.9)	65 (0.7)
West Virginia	24 (0.8)	52 (0.8)	62 (0.6)
Wisconsin	29 (0.8)	57 (0.7)	67 (0.5)
Wyoming	27 (0.8)	56 (0.6)	67 (0.6)
Territory			
Guam	12 (0.6)	34 (0.6)	49 (0.5)

Fourth-Grade Responses To Constructed-Response Questions

This section presents examples of the constructed responses provided by fourth-grade students. These sample questions were selected from among the set of questions and reading materials that were released from the 1992 assessment. In particular, examples are provided for three short constructed-response questions and one extended-response question, with all four questions addressing the autobiographical selection about Amanda Clements. The entire text presented in Chapter Two is summarized below.

> **Amanda Clement: The Umpire in a Skirt** is an autobiographical essay about the first paid woman baseball umpire on record. Hired in 1904, she is now recognized in the Baseball Hall of Fame. The article describes how she learned the sport at an early age by being asked to umpire for her brother and his friends, how well accepted she became in her profession, and what she did in her later life. Amanda's story is told against the context that in 1904 girls were not supposed to participate in sports.

Grade 4: Amanda Clement — Short Constructed-Responses

Students' responses to short constructed-response questions were scored according to a two-level rubric, such that a response was either acceptable or unacceptable. Responses scored as unacceptable indicated little or no understanding of the passage and question. Responses scored as acceptable indicated that the student had grasped both the passage and the question and was able to answer the question successfully.

An important reading skill is the ability to bring outside experiences and knowledge to an understanding of a text. The following short constructed-response question asked students to apply this ability to the Amanda Clement passage.

QUESTION 1: *Tell two ways in which Mandy's experience would be similar or different if she were a young girl wanting to take part in sports today.*

Unacceptable responses reflected a lack of understanding of Mandy's experience, often invoking knowledge related to the text's topic, but in ways irrelevant to the text's concerns and the question's intent. Two examples of unacceptable responses follow.

The following example of an acceptable response indicates both an understanding of the obstacles Mandy confronted and an ability to tell whether those obstacles would be the same or different in the light of current circumstances. Acceptable responses focused on various ideas, such as how girls today are allowed to play sports, how baseball games today have more than one umpire, and how some sports are still inaccessible to women.

As shown in Table 3.3, about one-third of the fourth graders (32 percent) provided acceptable responses to the question about comparing Mandy's experience to sports today. These fourth graders were able to relate Amanda's situation to the context of contemporary sports opportunities for girls. In order to provide an acceptable response to this question, students had to make a connection between text ideas and their own ideas about the world around them. By doing so, students could realize a different perspective on information provided in the passage. However, making this connection was relatively difficult for fourth graders. The majority of the students (58 percent) provided unacceptable responses, and another 10 percent omitted the question (or provided irrelevant or indecipherable responses).

There were no differences between regions in the proportions of students receiving an acceptable score on this question. However, significantly more White students than Black or Hispanic students gave acceptable responses. Also, female students outperformed their male counterparts. Significantly more students from advantaged urban communities received an acceptable score compared to students from disadvantaged and "other" communities. Students in "other" communities also showed higher performance on this question than students from disadvantaged urban communities. In addition, more private school fourth graders than public school fourth graders provided acceptable responses to this question. The state-by-state results presented in Table 3.4 generally mirrored the low levels of national performance on this question at grade 4.

At least two factors may influence a reader's ability to relate text ideas to world knowledge. First, the reader may lack adequate understanding of the passage, and in turn, be unable to make an appropriate connection. Second, the reader may have limited prior experiences from which to draw in making such connections. However, being able to integrate personal knowledge with knowledge gained from reading is generally considered paramount among the reading abilities necessary for critical understanding.

It is important to remember that the scoring of questions like this one in which students had to relate passage ideas to their own ideas did not take into account their understanding of world issues. Rather, the intent of this question, and others like it in the NAEP reading assessment, was to measure students' comprehension of the reading material. Therefore, being able to make a connection that is consistent with ideas in the text and demonstrates understanding of the passage was considered to be adequate — without judging the accuracy of the reader's world knowledge.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Compare to Girls in Sports Today,* Grade 4, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	32 (1.2)	58 (1.6)	10 (1.1)
Region			
Northeast	33 (3.1)	58 (3.0)	9 (2.0)
Southeast	30 (1.9)	60 (1.9)	10 (2.2)
Central	36 (2.9)	56 (4.2)	8 (2.0)
West	29 (1.5)	60 (2.6)	12 (2.4)
Race/Ethnicity			
White	35 (1.5)	56 (1.9)	8 (1.2)
Black	23 (3.3)	64 (3.3)	14 (3.0)
Hispanic	23 (3.1)	68 (3.2)	9 (2.1)
Gender			
Male	27 (2.0)	63 (2.5)	10 (1.4)
Female	37 (2.0)	54 (1.9)	9 (1.6)
Type of Community			
Advantaged Urban	44 (4.1)	49 (4.7)	7 (3.1)
Disadvantaged Urban	18 (4.6)	62 (4.5)	19 (4.0)
Extreme Rural	32 (5.1)	59 (1.8)	8 (4.2)
Other	32 (1.5)	59 (1.8)	9 (1.1)
Type of School			
Public	30 (1.3)	60 (1.7)	10 (1.2)
Private*	48 (3.3)	44 (3.3)	8(1.5)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Table 3.4

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Compare to Girls in Sports Today,* Grade 4, 1992 Trial State Reading Assessment

Public Schools	Acceptable	Unacceptable	No Response
Nation	30 (1.3)	60 (1.7)	10 (1.2)
Northeast	30 (3.4)	61 (3.3)	9 (2.4)
Southeast	28 (1.9)	62 (2.3)	11 (2.5)
Central	34 (3.5)	58 (4.6)	7 (2.0)
West	28 (1.1)	61 (2.9)	12 (2.6)
States			
Alabama	29 (2.1)	65 (2.1)	6 (0.7)
Arizona	32 (2.0)	59 (2.3)	9 (1.2)
Arkansas	32 (2.1)	63 (2.3)	5 (0.9)
California	28 (2.3)	60 (2.5)	12 (1.6)
Colorado	37 (2.3)	55 (2.5)	8 (1.2)
Connecticut	40 (2.2)	52 (2.4)	9 (1.4)
Delaware	36 (2.5)	56 (2.3)	8 (1.1)
District of Columbia	19 (1.6)	71 (2.3)	10 (1.5)
Florida*	28 (1.6)	62 (2.0)	10 (1.3)
Georgia	35 (1.5)	59 (1.6)	7 (1.3)
Hawaii	27 (2.1)	62 (2.2)	11 (1.6)
Idaho	36 (1.7)	57 (2.0)	7 (1.1)
Indiana	40 (2.0)	55 (2.0)	5 (1.0)
Iowa	44 (2.3)	51 (2.3)	5 (0.7)
Kentucky	36 (2.0)	58 (2.0)	6 (1.0)
Louisiana	23 (1.6)	69 (1.7)	8 (1.0)
Maine*	43 (2.7)	50 (2.6)	7 (1.3)
Maryland	37 (2.5)	57 (2.8)	7 (0.9)
Massachusetts	45 (2.1)	49 (2.2)	6 (1.1)
Michigan	34 (2.6)	60 (2.5)	6 (0.9)
Minnesota	39 (2.5)	53 (2.4)	8 (1.1)
Mississippi	24 (1.8)	68 (2.0)	8 (1.3)
Missouri	42 (2.4)	52 (2.6)	6 (0.9)
Nebraska*	37 (2.2)	55 (2.2)	8 (1.0)
New Hampshire*	46 (2.4)	46 (2.4)	7 (1.1)
New Jersey*	41 (2.9)	51 (2.9)	8 (1.3)
New Mexico	32 (2.6)	61 (2.5)	7 (1.2)
New York*	35 (2.2)	54 (2.5)	11 (1.3)
North Carolina	36 (2.2)	59 (2.1)	5 (0.7)
North Dakota	41 (2.2)	52 (2.6)	7 (1.5)
Ohio	40 (2.3)	54 (2.4)	6 (1.1)
Oklahoma	37 (2.5)	57 (2.7)	5 (1.0)
Pennsylvania	41 (2.1)	52 (2.0)	7 (1.0)
Rhode Island	38 (2.3)	54 (2.3)	9 (1.3)
South Carolina	27 (2.1)	67 (2.2)	5 (0.8)
Tennessee	36 (2.3)	58 (2.3)	6 (1.1)
Texas	31 (1.9)	62 (2.0)	7 (1.3)
Utah	36 (2.3)	56 (2.3)	8 (1.3)
Virginia	40 (2.0)	53 (2.1)	7 (1.0)
West Virginia	35 (2.2)	58 (2.3)	7 (0.9)
Wisconsin	44 (2.0)	51 (2.2)	6 (0.9)
Wyoming	39 (2.4)	54 (2.6)	8 (1.2)
Territory			
Guam	14 (1.6)	73 (2.2)	13 (1.6)

* Did not satisfy one or more of the guidelines for school sample participation rates (see Appendix B for details).

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Another short constructed-response question required fourth graders to collect evidence from a text to support an interpretation about a character or theme in the text, as in the following example.

QUESTION 2: *Give three examples showing that Mandy was not a quitter.*

Unacceptable responses, similar to the two examples shown below, typically demonstrated a weak grasp of how Mandy is portrayed in the passage, and an inability to cite specific information.

Acceptable responses indicated an understanding of how the passage presents Mandy's character, and an ability to choose specific information about Mandy from the passage that could be called upon to prove that she was not a quitter. Such responses usually referred to Mandy's determination to play, or to her career as a teacher and umpire, or both. Three examples follow.

As presented in Table 3.5, 43 percent of students' responses to this question were rated as acceptable, 51 percent of responses were scored as unacceptable, and 6 percent omitted the question. As with the previous question, there were no differences between regions in proportion of acceptable responses. In addition, no significant difference was observed between male and female fourth graders on this question. However, significantly more White students than Black and Hispanic students gave acceptable responses. Also, advantaged urban students out-performed disadvantaged urban students and more students from "other" communities than from disadvantaged urban communities received acceptable scores. A higher proportion of private school students than public school students provided acceptable responses.

The state-level data are shown in Table 3.6. Overall, from about one-third to one-half of fourth-graders provided acceptable responses to this question — a range of performance similar to that across the national reporting groups.

Fourth graders were somewhat more successful with this question than with the previous example question. One difference may have been that this question required students to make connections between events and situations in the story, rather than connections between the passage and personal knowledge. The fact that Mandy was "not a quitter" was quite evident in the article. Further, several circumstances were described in the passage that clearly supported this characterization of her.

Table 5.5

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Examples of Mandy Not a Quitter,* Grade 4, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	43 (1.9)	51 (2.0)	6 (0.9)
Region			
Northeast	47 (5.2)	47 (4.8)	6 (1.6)
Southeast	43 (2.8)	50 (2.6)	7 (1.4)
Central	48 (3.5)	48 (2.9)	4 (1.5)
West	35 (4.1)	56 (5.1)	9 (2.7)
Race/Ethnicity			
White	48 (2.3)	47 (2.4)	5 (1.0)
Black	32 (4.2)	57 (3.7)	11 (2.6)
Hispanic	28 (4.2)	62 (4.4)	10 (2.5)
Gender			
Male	41 (2.8)	52 (3.0)	7 (1.1)
Female	45 (2.2)	49 (2.1)	6 (1.2)
Type of Community			
Advantaged Urban	57 (5.1)	37 (3.7)	5 (2.5)
Disadvantaged Urban	27 (4.2)	58 (4.2)	14 (3.1)
Extreme Rural	43 (6.3)	47 (5.8)	9 (5.0)
Other	42 (2.3)	53 (2.4)	5 (0.9)
Type of School			
Public	42 (2.1)	52 (2.2)	6 (1.1)
Private*	50 (3.3)	44 (3.3)	6 (1.4)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Examples of Mandy Not a Quitter,* Grade 4, 1992 Trial State Reading Assessment

Public Schools	Acceptable	Unacceptable	No Response
Nation	42 (2.1)	51 (2.2)	6 (1.1)
Northeast	47 (5.8)	47 (5.4)	6 (1.7)
Southeast	42 (2.6)	52 (2.7)	7 (1.6)
Central	48 (4.0)	49 (3.5)	4 (1.8)
West	33 (4.1)	57 (5.4)	9 (2.8)
States			
Alabama	39 (2.3)	57 (2.1)	4 (0.9)
Arizona	41 (2.1)	52 (1.9)	6 (1.0)
Arkansas	42 (2.8)	53 (2.6)	4 (0.8)
California	34 (2.2)	56 (2.4)	9 (1.5)
Colorado	40 (2.0)	54 (2.1)	6 (0.9)
Connecticut	44 (2.0)	49 (2.3)	7 (1.3)
Delaware*	43 (2.7)	51 (2.8)	6 (1.2)
District of Columbia	33 (2.4)	58 (2.2)	9 (1.2)
Florida	36 (2.3)	56 (2.2)	7 (1.0)
Georgia	46 (2.3)	48 (2.3)	5 (1.6)
Hawaii	35 (2.5)	58 (2.5)	7 (1.3)
Idaho	41 (2.1)	54 (2.0)	4 (1.0)
Indiana	51 (2.0)	45 (2.0)	4 (0.8)
Iowa	44 (1.9)	52 (2.1)	5 (0.9)
Kentucky	41 (1.9)	55 (2.2)	5 (0.9)
Louisiana	38 (1.8)	54 (1.9)	8 (1.1)
Maine*	49 (2.8)	47 (2.9)	4 (1.0)
Maryland	44 (2.1)	52 (2.1)	5 (0.8)
Massachusetts	46 (2.0)	49 (2.0)	5 (1.0)
Michigan	44 (2.7)	52 (2.7)	4 (0.9)
Minnesota	44 (2.1)	50 (2.2)	7 (1.1)
Mississippi	35 (2.3)	57 (2.3)	8 (1.5)
Missouri	47 (2.2)	49 (2.2)	4 (0.8)
Nebraska*	43 (2.2)	50 (2.4)	6 (1.1)
New Hampshire*	47 (2.6)	47 (2.7)	6 (1.0)
New Jersey*	49 (2.6)	46 (2.5)	5 (1.1)
New Mexico	39 (3.0)	56 (3.1)	6 (1.1)
New York*	46 (2.1)	46 (2.1)	8 (1.3)
North Carolina	40 (2.1)	55 (2.0)	5 (0.8)
North Dakota	51 (2.6)	47 (2.5)	3 (0.7)
Ohio	47 (2.0)	48 (1.7)	5 (0.9)
Oklahoma	47 (2.3)	49 (2.4)	4 (0.8)
Pennsylvania	47 (2.1)	47 (2.0)	6 (0.9)
Rhode Island	44 (2.2)	49 (1.8)	6 (1.1)
South Carolina	39 (2.3)	55 (2.3)	6 (0.9)
Tennessee	40 (2.1)	56 (2.1)	5 (0.8)
Texas	40 (1.9)	54 (2.0)	5 (1.0)
Utah	40 (2.2)	53 (2.1)	7 (0.9)
Virginia	47 (2.2)	47 (2.2)	5 (0.8)
West Virginia	45 (2.3)	49 (2.0)	6 (1.1)
Wisconsin	45 (2.1)	50 (2.2)	5 (0.8)
Wyoming	44 (2.2)	50 (2.2)	6 (1.1)
Territory			
Guam	26 (1.9)	62 (2.3)	12 (1.6)

* Did not satisfy one or more of the guidelines for school sample participation rates (see Appendix B for details).

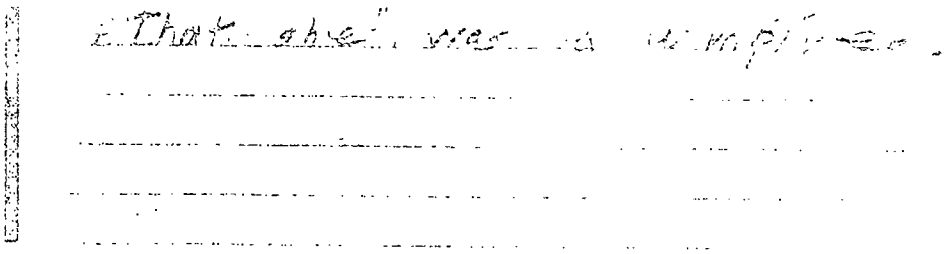
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

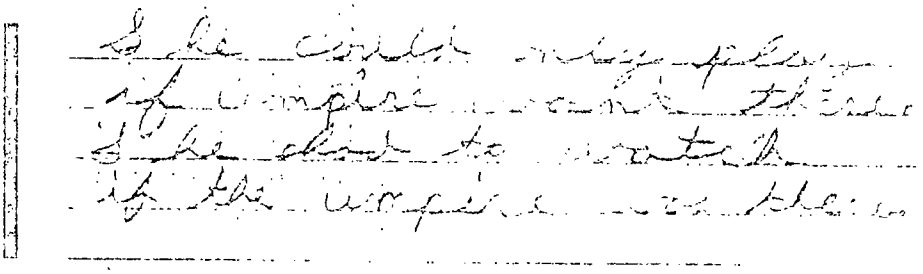
SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

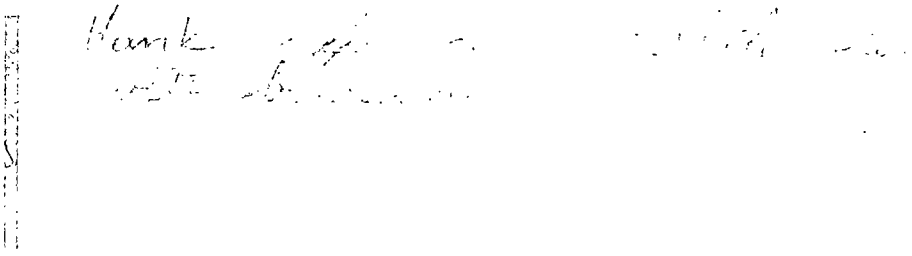
The following relatively straightforward question about the relationship between characters and events was designed to measure students' global understanding of the text.

QUESTION 3: *What was Hank's role in Mandy's early career?*

Unacceptable responses may have showed some minimal grasp of events, but did not indicate an ability to relate events to one another or to characters. Some made reference to umpiring or to Hank, but without relating either to the relationship between Hank's actions and Mandy's life.







Acceptable responses like the following demonstrated an understanding of how Hank assisted Mandy by letting her umpire.

The national and state-by-state data are presented in Tables 3.7 and 3.8, respectively. For the nation as a whole, 42 percent of student responses to this question were scored as acceptable, and 55 percent as unacceptable. Three percent of the students omitted the question. Closely reflecting the range of performance across community types for the nation, the range of acceptable responses to this question for the states was 19 percent to 54 percent.

For this question, significant differences in performance were observed by race/ethnicity, type of community, and type of school. White fourth-graders out-performed their Black and Hispanic counterparts. Fewer students from disadvantaged urban communities provided acceptable responses compared to students from the other three community types. In addition, higher performance was demonstrated by students from advantaged urban communities compared to students from extreme rural communities. As with other questions, students attending private schools provided more acceptable responses than students from public schools.

Students' performance on this question was quite similar to their performance on the previous question. In both cases, fourth graders were being asked to consider situations in the text and make connections

between them or provide a generalization. Here again, slightly over one-half of the fourth graders provided unacceptable responses. With this particular question, students needed to draw on their knowledge of the events surrounding two characters in describing a causal relationship between them. This required not only understanding what each character did, but also understanding the impact of one's actions on the other.

Table 4.7

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Hank's Role in Her Career,* Grade 4, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	42 (2.0)	55 (2.2)	3 (0.7)
Region			
Northeast	50 (4.9)	48 (4.2)	2 (1.1)
Southeast	40 (3.9)	59 (3.8)	1 (0.6)
Central	41 (4.2)	57 (4.4)	2 (0.7)
West	39 (3.3)	56 (4.6)	6 (2.2)
Race/Ethnicity			
White	47 (2.4)	50 (2.5)	2 (0.6)
Black	21 (4.1)	72 (4.3)	6 (2.4)
Hispanic	26 (3.6)	72 (3.9)	2 (1.2)
Gender			
Male	42 (2.6)	56 (2.5)	2 (0.7)
Female	42 (2.3)	54 (2.6)	4 (1.3)
Type of Community			
Advantaged Urban	58 (6.2)	41 (6.3)	2 (1.3)
Disadvantaged Urban	15 (3.7)	79 (4.5)	7 (3.0)
Extreme Rural	35 (5.8)	58 (6.8)	7 (3.3)
Other	43 (2.3)	55 (2.4)	2 (0.6)
Type of School			
Public	41 (2.2)	56 (1.4)	3 (0.8)
Private*	50 (3.9)	50 (3.9)	0 (0.3)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Table 3.8

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Amanda Clement: Hank's Role in Her Career,* Grade 4, 1992 Trial State Reading Assessment

Public Schools	Acceptable	Unacceptable	No Response
Nation	41 (2.2)	56 (2.4)	3 (0.8)
Northeast	50 (6.1)	48 (5.4)	2 (1.3)
Southeast	39 (3.9)	60 (3.9)	1 (0.7)
Central	39 (4.6)	58 (4.9)	2 (0.8)
West	39 (3.3)	55 (4.9)	6 (2.4)
States			
Alabama	34 (2.4)	65 (2.4)	1 (0.5)
Arizona	37 (2.6)	60 (2.7)	3 (0.7)
Arkansas	39 (2.8)	60 (2.6)	1 (0.4)
California	34 (2.9)	62 (2.9)	3 (1.0)
Colorado	44 (2.4)	53 (2.5)	3 (0.8)
Connecticut	49 (2.5)	49 (2.5)	2 (0.6)
Delaware*	41 (3.0)	59 (3.1)	0 (0.3)
District of Columbia	34 (2.8)	64 (2.8)	2 (0.7)
Florida	34 (2.4)	64 (2.3)	2 (0.6)
Georgia	42 (2.0)	56 (2.0)	2 (0.6)
Hawaii	33 (2.7)	64 (2.6)	2 (0.7)
Idaho	46 (2.2)	53 (2.2)	1 (0.4)
Indiana	46 (2.4)	53 (2.4)	1 (0.4)
Iowa	52 (2.1)	47 (2.1)	1 (0.5)
Kentucky	40 (2.1)	58 (2.1)	1 (0.5)
Louisiana	32 (2.7)	67 (2.6)	1 (0.5)
Maine*	52 (2.6)	46 (2.4)	2 (0.8)
Maryland	47 (2.6)	50 (2.5)	3 (0.8)
Massachusetts	54 (2.0)	46 (2.0)	0 (0.3)
Michigan	40 (2.8)	59 (2.8)	2 (0.6)
Minnesota	51 (2.5)	48 (2.3)	1 (0.6)
Mississippi	28 (2.2)	71 (2.2)	1 (0.5)
Missouri	43 (2.1)	56 (2.1)	1 (0.5)
Nebraska*	44 (2.5)	54 (2.5)	1 (0.4)
New Hampshire*	50 (2.6)	49 (2.7)	2 (0.6)
New Jersey*	47 (2.7)	52 (2.7)	1 (0.6)
New Mexico	38 (3.7)	61 (3.7)	1 (0.6)
New York*	48 (2.8)	51 (2.8)	1 (0.5)
North Carolina	40 (2.1)	59 (2.1)	1 (0.5)
North Dakota	45 (2.9)	55 (2.8)	1 (0.3)
Ohio	44 (2.5)	55 (2.5)	1 (0.4)
Oklahoma	49 (2.8)	50 (2.7)	1 (0.4)
Pennsylvania	42 (2.5)	57 (2.5)	1 (0.4)
Rhode Island	44 (2.2)	56 (2.2)	1 (0.5)
South Carolina	32 (2.4)	67 (2.3)	1 (0.4)
Tennessee	36 (2.2)	62 (2.3)	1 (0.4)
Texas	36 (3.3)	62 (3.3)	1 (0.7)
Utah	41 (2.8)	57 (2.9)	2 (0.7)
Virginia	47 (2.3)	52 (2.4)	1 (0.5)
West Virginia	44 (2.0)	55 (2.0)	1 (0.4)
Wisconsin	50 (2.4)	49 (2.3)	1 (0.5)
Wyoming	45 (2.4)	53 (2.2)	1 (0.6)
Territory			
Guam	19 (2.1)	77 (2.2)	3 (1.1)

* Did not satisfy one or more of the guidelines for school sample participation rates (see Appendix B for details).

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages 0.5 percent and less were rounded to 0 percent.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Grade 4: Amanda Clement — Extended-Response Question

Students' responses to the extended-response questions were scored using a rubric reflecting four possible levels of understanding: unsatisfactory, partial, essential, and extensive. Responses scored as **unsatisfactory** reflected little or no understanding, or repeated disjointed or isolated bits from the passage. Responses rated as **partial** demonstrated some understanding, but it was incomplete, fragmented, or not supported with appropriate evidence or argument. Responses scored as **essential** included enough detail and complexity to indicate that students had developed at least generally appropriate understandings of the passage and the question. Responses rated as **extensive** indicated that students had more fully considered the issues and, in doing so, had developed elaborated understandings and explanations.

As readers make sense of what they read, one important ability involves learning to ask questions — questions about the characters, ideas, and events, questions about the relationship between the events described and the readers' own personal experience, and questions about how what is being read relates to the context in which it was written. The extended question that fourth graders were asked about Amanda Clement drew upon their ability to ask questions, as well as their ability to explain the significance of the questions they generated.

QUESTION 4: *If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know.*

Successful responses to this question went beyond surface comprehension to a fuller understanding of Amanda's career in light of her gender, times, personal experiences, or social experiences. Students were asked to provide evidence of such understanding by posing a relevant question not already answered in the passage, and by explaining the relevance of the question in terms of Mandy's life and times, or their own. The results are presented in Table 3.9.

Table 3.9

Percentage of Responses for the Extended Constructed-Response Question, *Amanda Clement: The Umpire in a Skirt* — "If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know." Grade 4, 1992 Reading Assessment

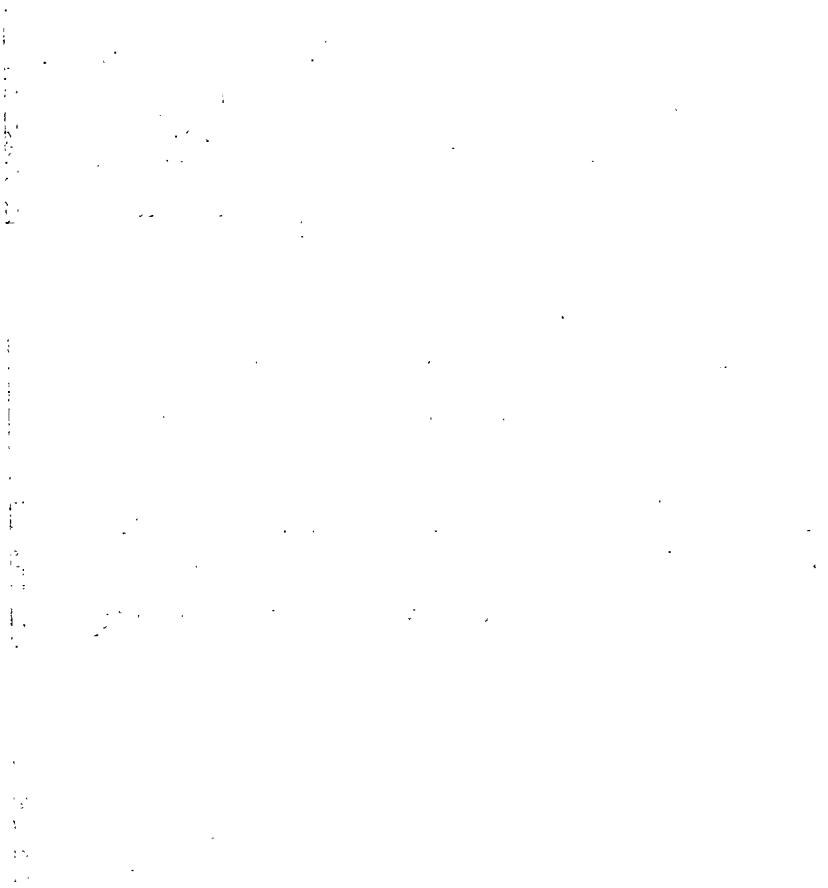
	Not Rated	Unsatis-factory	Partial	Essential	Extensive	Essential or Better
Nation	3 (0.5)	14 (1.2)	50 (1.9)	31 (1.3)	2 (0.4)	33 (1.4)
Region						
Northeast	3 (1.2)	12 (1.8)	49 (3.1)	34 (3.1)	2 (1.1)	36 (3.8)
Soutneast	3 (0.8)	15 (2.8)	47 (4.4)	32 (2.4)	2 (0.8)	34 (2.7)
Central	2 (0.6)	14 (2.7)	52 (4.2)	30 (3.2)	3 (1.1)	33 (3.3)
West	4 (1.2)	15 (2.8)	52 (3.2)	28 (2.1)	1 (0.5)	29 (1.9)
Race/Ethnicity						
White	2 (0.5)	12 (1.4)	50 (2.2)	34 (1.8)	2 (0.6)	37 (1.8)
Black	6 (1.9)	23 (3.2)	51 (4.4)	18 (2.7)	2 (1.1)	20 (3.1)
Hispanic	5 (1.5)	17 (3.3)	48 (4.1)	29 (3.5)	2 (1.1)	31 (3.6)
Gender						
Male	4 (0.8)	15 (1.7)	52 (2.5)	27 (1.9)	1 (0.5)	28 (2.0)
Female	1 (0.5)	13 (1.4)	48 (2.1)	35 (1.9)	3 (0.8)	38 (2.2)
Type of Community						
Advantaged Urban	1 (0.4)	8 (2.4)	50 (4.4)	38 (4.1)	4 (1.5)	42 (4.8)
Disadvantaged Urban	8 (2.4)	24 (2.9)	54 (4.6)	13 (3.1)	2 (1.5)	15 (3.7)
Extreme Rural	5 (2.3)	19 (4.3)	46 (6.4)	28 (4.4)	2 (1.7)	30 (5.1)
Other	3 (0.6)	13 (1.4)	51 (2.1)	32 (1.7)	2 (0.5)	34 (1.7)
Type of School						
Public	3 (0.6)	14 (1.3)	51 (2.1)	30 (1.4)	2 (0.5)	32 (1.6)
Private*	2 (0.7)	11 (2.7)	45 (3.2)	39 (2.3)	3 (1.0)	42 (2.2)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

* The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

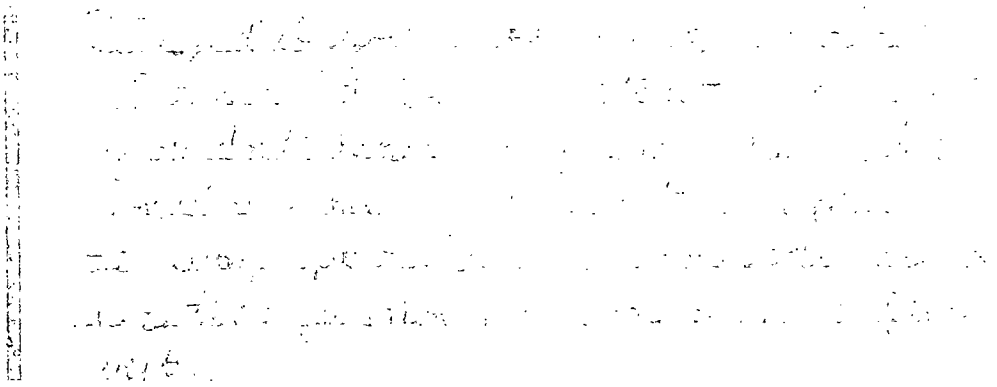
Unsatisfactory understanding was reflected in responses that demonstrated little or no understanding of Mandy's life or career in that they cited isolated or unrelated bits of information from the passage, or posed a question unrelated to Mandy's career or situation. For example:



For the nation as a whole, 14 percent of the students' responses were rated as unsatisfactory, and another 3 percent did not respond at all or responded irrelevantly (not ratable included I don't know, off-task, and illegible responses).

Responses reflecting **partial** understanding demonstrated some understanding of Mandy's life or career by posing at least a relevant question. Approximately one-half (50 percent) of fourth graders demonstrated partial understanding. For example:

Responses reflecting **essential** understanding demonstrated an overall understanding of Mandy's life and career. Some 31 percent of students' responses were scored at this level. They contained at least one question specifically related to Mandy's career with a relevant explanation about the importance of that question. For example:



Responses reflecting **extensive** understanding demonstrated a richer understanding of the passage, indicating that the student has considered the more complex social or personal issues suggested by the passage. These responses, for example, might have contained questions about issues or feelings that emerge from consideration of the potential problems Mandy faced, placing her in a historical and social context. Very few students — two percent nationally — provided responses such as these. For example:

Across all demographic subgroups, a very small proportion (1 to 4 percent) of fourth graders attained an extensive score with this question. In demonstrating at least essential level comprehension, however, White students out-performed Black students and females surpassed males. More students from advantaged urban communities scored essential or better than did students from disadvantaged urban communities. Also, the higher performance of private school students compared to public school students continued to be apparent with this question, as more private school fourth graders displayed at least essential level comprehension.

As shown in Table 3.10, the percentages of success for public school fourth graders in jurisdictions participating in the trial state assessment were similar to those for the nation. For four states, Maine, New Hampshire, Pennsylvania, and Wisconsin, at least 33 percent of the students provided essential or better responses.

Generating thoughtful questions about ideas in text may be considered one of the hallmarks of critical reading abilities. Questioning and exploring additional information are ways in which readers can extend their understanding of a passage. This question about Amanda Clement gave fourth graders an opportunity to demonstrate their ability to extend their understanding and examine the relevance of their own questions about the text — a somewhat complex task. Only about one-third of fourth graders were able to provide responses that demonstrated at least essential comprehension.

The fact that 50 percent of students demonstrated partial level understanding suggests that a great many fourth graders were successful with at least one part of this task. That is, they were able to generate a question, but then could not explain the importance of their question. Many responses displayed circular reasoning in their explanation for why the answer to their question would be important. For example, a statement like, "It would be important because I would want to know ' does not adequately explain the relevance of a student's question. These findings may suggest that, while a majority of fourth graders could generate a pertinent question, most of them were unable to extend their textual understandings by providing a critical examination of their self-generated questions — clearly a higher-level reading ability.

Table 3.10

Percentage of Responses for the Extended Constructed-Response Question, *Amanda Clement: The Umpire in a Skirt* — "If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know." Grade 4, 1992 Trial State Reading Assessment

	Not Rated	Unsatis-factory	Partial	Essential	Extensive	Essential or Better
Nation	3 (0.6)	14 (1.3)	51 (2.1)	29 (1.4)	2 (0.5)	32 (1.6)
Northeast	3 (1.4)	12 (1.9)	49 (3.6)	34 (3.9)	2 (1.1)	36 (4.6)
Southeast	3 (0.9)	15 (3.0)	48 (4.6)	32 (2.3)	2 (0.9)	34 (2.8)
Central	2 (0.6)	15 (2.7)	52 (4.6)	28 (3.4)	3 (1.3)	31 (3.6)
West	4 (1.3)	15 (2.4)	53 (3.7)	26 (2.3)	1 (0.6)	27 (2.2)
States						
Alabama	3 (0.7)	18 (1.5)	55 (2.3)	24 (2.2)	1 (0.5)	25 (2.2)
Arizona	3 (0.7)	17 (1.6)	55 (2.0)	24 (1.8)	1 (0.5)	25 (1.8)
Arkansas	3 (0.7)	17 (1.4)	56 (2.2)	24 (2.1)	1 (0.4)	25 (2.1)
California	6 (1.2)	17 (1.5)	51 (2.2)	24 (1.7)	2 (0.5)	26 (1.8)
Colorado	2 (0.6)	15 (1.3)	55 (1.9)	26 (1.6)	1 (0.4)	28 (1.6)
Connecticut	2 (0.6)	11 (1.3)	55 (2.1)	30 (2.2)	2 (0.4)	32 (2.2)
Delaware*	3 (0.9)	15 (1.4)	54 (2.5)	27 (1.9)	1 (0.6)	28 (2.1)
District of Columbia	3 (0.6)	23 (1.5)	54 (1.7)	18 (1.4)	1 (0.4)	19 (1.5)
Florida	4 (0.7)	16 (1.4)	54 (1.5)	26 (1.4)	1 (0.3)	26 (1.4)
Georgia	4 (0.9)	15 (1.4)	58 (2.1)	22 (1.8)	1 (0.4)	23 (1.9)
Hawaii	3 (0.6)	20 (1.8)	54 (2.1)	22 (2.0)	1 (0.4)	23 (2.0)
Idaho	4 (0.9)	12 (1.5)	57 (2.0)	27 (1.8)	1 (0.5)	28 (1.8)
Indiana	2 (0.6)	9 (1.1)	59 (1.8)	29 (1.8)	2 (0.4)	30 (1.9)
Iowa	3 (0.6)	11 (1.5)	55 (2.2)	29 (2.1)	2 (0.6)	32 (2.2)
Kentucky	4 (0.7)	12 (1.3)	52 (1.8)	31 (1.8)	1 (0.4)	32 (1.8)
Louisiana	3 (0.7)	19 (1.6)	55 (2.2)	22 (1.8)	0 (0.2)	23 (1.8)
Maine*	3 (0.7)	11 (1.1)	53 (2.7)	30 (2.6)	2 (0.7)	33 (2.7)
Maryland	3 (0.7)	15 (1.1)	54 (2.1)	27 (2.1)	1 (0.3)	28 (2.1)
Massachusetts	2 (0.6)	9 (1.2)	57 (2.1)	30 (2.0)	3 (0.7)	32 (2.1)
Michigan	2 (0.6)	14 (1.5)	57 (2.1)	25 (1.9)	2 (0.5)	27 (1.8)
Minnesota	4 (0.8)	13 (1.4)	52 (1.6)	30 (1.9)	1 (0.4)	31 (2.0)
Mississippi	3 (0.8)	24 (1.9)	51 (2.1)	22 (2.0)	1 (0.4)	23 (2.1)
Missouri	3 (0.8)	11 (1.4)	58 (2.4)	27 (2.1)	1 (0.4)	28 (2.2)
Nebraska*	2 (0.6)	14 (1.5)	56 (1.8)	27 (1.9)	1 (0.4)	28 (1.9)
New Hampshire*	3 (0.8)	9 (1.3)	54 (1.9)	32 (1.9)	2 (0.5)	34 (2.0)
New Jersey*	2 (0.8)	12 (1.4)	54 (2.5)	30 (2.2)	1 (0.6)	31 (2.2)
New Mexico	2 (0.6)	18 (1.9)	51 (1.7)	28 (1.8)	1 (0.6)	29 (1.9)
New York*	4 (0.7)	16 (1.4)	51 (1.8)	28 (1.6)	1 (0.4)	29 (1.5)
North Carolina	3 (0.5)	15 (1.3)	53 (1.8)	28 (1.6)	1 (0.3)	29 (1.6)
North Dakota	2 (0.7)	8 (1.4)	59 (2.7)	30 (2.4)	1 (0.5)	32 (2.4)
Ohio	3 (0.6)	12 (1.1)	53 (1.7)	30 (1.6)	1 (0.4)	31 (1.7)
Oklahoma	1 (0.5)	14 (1.4)	53 (2.7)	30 (2.6)	1 (0.4)	31 (2.7)
Pennsylvania	2 (0.6)	13 (1.3)	51 (2.0)	32 (2.0)	2 (0.7)	34 (1.9)
Rhode Island	3 (0.7)	15 (1.7)	54 (2.3)	27 (2.0)	1 (0.5)	28 (2.0)
South Carolina	3 (0.6)	17 (1.8)	54 (2.2)	25 (2.0)	1 (0.4)	26 (2.0)
Tennessee	3 (0.6)	17 (1.2)	53 (2.0)	26 (1.9)	1 (0.5)	28 (1.9)
Texas	2 (0.7)	17 (1.8)	54 (2.4)	26 (2.1)	1 (0.3)	27 (2.1)
Utah	3 (0.8)	12 (1.3)	56 (2.1)	28 (1.8)	2 (0.5)	30 (1.9)
Virginia	3 (0.8)	11 (1.0)	58 (1.8)	26 (1.7)	2 (0.5)	28 (1.8)
West Virginia	3 (0.7)	11 (1.2)	54 (2.0)	31 (2.0)	1 (0.3)	32 (2.1)
Wisconsin	2 (0.5)	10 (1.3)	55 (2.4)	32 (2.0)	1 (0.6)	33 (2.0)
Wyoming	3 (0.6)	14 (1.1)	54 (2.0)	28 (2.1)	1 (0.4)	29 (2.0)
Territory						
Guam	9 (1.2)	31 (2.1)	43 (2.2)	17 (1.6)	0 (0.2)	17 (1.5)

* Did not satisfy one or more of the guidelines for school sample participation rates (see Appendix B for details).

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value of the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages 0.5 percent and less were rounded to 0 percent.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Summary

This chapter described fourth graders' performance on specific short constructed-response questions and an extended-response question in the context of their performance on the entire assessment. In general, one would expect that students would experience more difficulty answering questions that ask them to explore and manipulate the information they read, and the more manipulation, the greater the difficulty. This, in fact, was the case in this assessment, with students' performance overall being higher on multiple-choice items, somewhat lower on short constructed-response questions, and lowest on extended-response questions. On average, 67 percent of fourth-grade students provided acceptable responses to multiple choice questions, 51 percent gave acceptable answers to short constructed-response questions, but only 26 percent were able to answer extended-response questions demonstrating at least essential comprehension. Examples of constructed-response questions and students' answers show that students have considerable difficulty providing extended responses.

With one example question in this chapter, fourth-graders demonstrated considerable difficulty connecting personal knowledge with information in the article. Only 32 percent were able to explain how Mandy's experience with sports would be similar or different if it had occurred in the present. Their performance was slightly better on two other example questions which required students to connect ideas within the text or to support a generalization. With these two questions, from 42 to 43 percent were able to provide acceptable responses. Fourth graders' responses to the extended-response question about Amanda Clement showed that most of them could at least generate a question about the article (83 percent with at least *Partial* responses), but had more difficulty in explaining the relevance of their questions (33 percent *Essential or Better*).

The example questions presented in this chapter provide a glimpse of the information that can be gained by examining students' constructed responses to reading. When readers are asked to construct a response, they must take their understanding of the text and do something with it. At the very least, this may require readers simply to communicate their understanding. By doing so, readers demonstrate how much of the text's meaning they have grasped to the point of being able to describe it. However, as displayed through the examples in this chapter, the constructed-response questions used in the NAEP assessment typically required readers to go beyond simply communicating their understanding. Instead, students were required to display a range of stances with the text, to make connections between ideas in the passage, and to integrate personal knowledge with text information.

This chapter continues the discussion of constructed-response questions in the NAEP reading assessment by focusing on eighth graders' performance on the different question types as well as their performance on sample constructed-response questions. The same question formats were used at eighth grade as were used with fourth graders: multiple-choice, short constructed-response, and extended response. However, the reading materials differed from those used at the fourth grade in difficulty, length, complexity, and topic. In addition, at eighth grade some students were given sets of reading materials representing different text genre. With these tasks, students were required not only to demonstrate comprehension of each passage, but also to integrate ideas across the texts. The sample questions presented in this chapter were part of an eighth-grade reading task involving multiple passages.

Average Performance on Question Types

Table 4.1 presents the average percentage of successful responses for each of the three types of questions. National data only are available for eighth and twelfth graders' performance on constructed-response questions since the state assessment was not conducted at these grades. Similar to the performance of fourth graders, students in the eighth grade also had the greatest difficulty with the extended-response questions. About two-thirds (67 percent), on average, provided correct responses to the multiple-choice questions. In comparison, only about one-half (51 percent) gave acceptable responses to short constructed-response questions and about one-fourth (26 percent) demonstrated at least essential comprehension in their answers to extended-response questions.

For all question types, students in the Southeast had lower average performance than students in any of the other three regions of the country. On short constructed-response questions the difference with the West was not significant. Students in advantaged urban communities had higher average performance for all question types than students in extreme rural, disadvantaged urban, or other types of communities. Students in disadvantaged urban communities had lower average performance than students in any of the other three types of communities. Private school students out-performed public school students.

For all question types, White students performed significantly better than both Black and Hispanic students. Hispanic students, however, showed a significant advantage over Black students on the more difficult extended-response questions. Females' advantage over males in eighth-grade reading performance increased as the complexity and difficulty of the question type increased. That is, while the difference between males and females for average correct response on multiple-choice questions was only 4 percent, the difference for short constructed-response questions was 7 percent and the difference for extended-response questions was 11 percent. While females consistently out-performed their male counterparts, the gap was smallest for multiple-choice questions. These findings are consistent with other research observing advantages for female students over male students with written response formats in assessment.[42]

---

[42] Mazzeo, J., Schmitt, A., & Bleistein, C., *Exploratory Analyses of Some Possible Causes for the Discrepancies in Gender Differences on Multiple-Choice and Free-Response Sections of the Advanced Placement Examinations* (Princeton, NJ: Educational Testing Service, Draft Report, 1990).

Bieland, H.M., & Griswold, P.A., Use of a Performance Test as a Criterion in a Differential Validity Study, *Journal of Educational Psychology*, 74 713-721, 1982.

Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 8, 1992 Reading Assessment

	EXTENDED RESPONSE	SHORT CONSTRUCTED-RESPONSE	MULTIPLE-CHOICE
	Average Percentage Essential or Better	Average Percentage Acceptable	Average Percentage Correct
Nation	26 (0.5)	51 (0.5)	67 (0.4)
Region			
Northeast	28 (1.2)	53 (1.2)	69 (0.7)
Southeast	22 (0.6)	48 (0.9)	64 (0.7)
Central	27 (1.1)	53 (1.4)	69 (1.2)
West	25 (0.9)	50 (0.7)	67 (0.7)
Race/Ethnicity			
White	29 (0.6)	55 (0.6)	71 (0.5)
Black	15 (0.7)	38 (0.8)	56 (0.7)
Hispanic	18 (0.8)	39 (0.8)	58 (0.7)
Gender			
Male	20 (0.5)	47 (0.6)	65 (0.6)
Female	31 (0.7)	54 (0.6)	69 (0.5)
Type of Community			
Advantaged Urban	37 (1.2)	61 (1.2)	76 (1.0)
Disadvantaged Urban	15 (1.1)	38 (1.0)	56 (0.9)
Extreme Rural	27 (2.0)	52 (2.0)	69 (1.8)
Other	25 (0.7)	51 (0.6)	67 (0.5)
Type of School			
Public	24 (0.5)	49 (0.5)	66 (0.4)
Private*	34 (1.0)	61 (1.1)	75 (0.9)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Eighth-Grade Responses to Constructed-Response Questions

A sense of a literary tradition evolves out of readers' abilities to see the relationships among diverse authors and works. These relationships take many forms, including the use of conventional characters (the martyr, the heroine), conventional settings ("It was a dark and stormy night"), established genres (fairy tale, short story, novel), shared predicaments, and common themes. The example responses in this section were provided by eighth graders in response to a group of materials representing three different genres (see full text in Chapter Two). In brief:

> **Cady's Life** is a story written by Anne Frank when she was hiding in an attic to escape persecution by Hitler and the Nazis during World War II. It is told in the first person by a Christian girl named Cady and is about her experiences with and sorrow for her friend Mary, who was Jewish and who was eventually arrested along with the rest of her family. The story was preceded with biographical information about Anne Frank. Also, the story was paired with a poem, "I Am One," by Edward Everett Hale, in which Hale acknowledges that while one person cannot do everything, "I will not refuse to do the something that I can do."

Grade 8: Cady's Life — Short Constructed-Responses

Some short constructed-response questions were designed to determine whether students were able to stand back from a passage and consider why its author has used a particular style or approach. For example, the question below asked students in the eighth grade to employ their critical powers to think about the point of view taken in the short story written by Anne Frank, "Cady's Life."

QUESTION 1: *Why did the author write this story from the perspective of Cady, a Christian?*

Unacceptable responses to this question either did not focus on perspective at all, or showed confusion about why the author might have been interested in using Cady's feelings to frame the story. Such responses indicated a difficulty in grasping how a particular perspective might function in a text.

Acceptable responses like the following indicated an understanding of what a reader could learn from Cady's perspective, and the utility of Cady's perspective for Frank. These responses often focused on how Frank wanted to explore what Christians felt about the fate of the Jews.

According to the data presented in Table 4.2, approximately half of the students (51 percent) provided responses to this question that were scored as unacceptable, and 11 percent of the students omitted the question. Just slightly more than one-third (38 percent) of eighth-graders were able to take a critical stance with this question and provide an acceptable response — suggesting that this may be a somewhat difficult reading ability for these students. Asking readers to consider why the author presents information from a particular perspective requires the ability to take a critical stance with the passage. That is, readers must step back from their text-based understanding, think objectively about how the author has crafted the piece, and make evaluative decisions about why the author may have done so.

As with other questions, students from disadvantaged urban communities did not perform as well as students from the remaining community types. Whereas 56 percent of advantaged urban students, 43 percent of extreme rural students, and 38 percent of students from "other" communities provided acceptable responses, only 20 percent of students from disadvantaged urban communities were able to do so. Students from advantaged urban communities also demonstrated an advantage over their counterparts in communities designated as "other." Students from the Northeast provided significantly more acceptable

responses than students from the Southeast. Forty-three percent of White students achieved acceptable responses, significantly more than the 23 percent for both Black and Hispanic students. Female students (43 percent) provided more acceptable responses than male students (33 percent). A significantly higher percentage of private school students than public school students provided acceptable responses.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Cady's Life: Why Cady's Perspective,* Grade 8, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	38 (1.2)	51 (1.3)	11 (0.8)
Region			
Northeast	43 (1.8)	44 (1.7)	13 (1.4)
Southeast	34 (2.3)	57 (2.3)	9 (2.4)
Central	40 (2.8)	54 (2.9)	6 (1.0)
West	37 (2.4)	49 (3.1)	14 (1.6)
Race/Ethnicity			
White	43 (1.4)	50 (1.8)	8 (1.1)
Black	23 (2.3)	56 (2.9)	20 (2.7)
Hispanic	23 (3.0)	57 (3.5)	20 (2.3)
Gender			
Male	33 (1.5)	53 (2.0)	14 (1.3)
Female	43 (2.1)	49 (2.2)	8 (0.9)
Type of Community			
Advantaged Urban	56 (4.2)	40 (5.0)	5 (2.1)
Disadvantaged Urban	20 (3.5)	55 (4.0)	26 (3.1)
Extreme Rural	43 (4.8)	48 (2.6)	8 (4.2)
Other	38 (1.4)	53 (1.5)	10 (1.1)
Type of School			
Public	36 (1.4)	53 (1.5)	12 (1.0)
Private*	58 (2.9)	38 (2.8)	4 (1.1)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Another important reading skill is the ability to use one text, possibly of a different genre, to better understand another. The question below required students to use their understanding of the poem "I Am One" to think about the biographical information provided at the beginning of "Cady's Life."

QUESTION 2: *For Anne Frank, what was "the something that I can do?"*

Unacceptable responses indicated a lack of understanding of the texts themselves and of how to use different texts or ger res together. They were characterized by vague statements about what Anne Frank may have been feeling that indicated a weak grasp of her circumstances. Other responses offered an interpretation of the poem without any real attempt to use the poem to consider Frank's life.

Acceptable responses showed an understanding of the different kinds of information in the story and in the biographical piece; students who wrote these responses were able to distinguish between Frank as a writer and the characters Frank created, and thus to think about Frank's life in the abstract, and to consider the meaning of the creation of "Cady's Life" in the context of the poem. Many of the responses discussed Anne Frank's decision to write about her experiences.

This question proved to be at least as difficult for eighth graders as was the last example question, although the task itself was quite different. With this question, students were called upon to integrate their understandings from two texts representing different genre. Being able to relate ideas and build new interpretations based on more than one passage has been referred to as "intertextuality."[43] The importance of this ability has received increased attention as educators acknowledge the diversity in types and forms of materials that readers in today's society must negotiate in order to build more complete understandings of various topics and issues.

Disappointingly perhaps, results for this question indicated that only 33 percent of eighth graders could make a connection between a poet's ideas and the facts surrounding a historical figure's life presented in biographical sketch (see Table 4.3). Half of the student responses to this question were scored as unacceptable, and another 17 percent of eighth graders omitted the question. Forty-three percent of student responses from private schools were scored as acceptable, while 32 percent of student responses from public schools were scored as acceptable. Thirty-seven percent of White students' responses were scored as acceptable, compared to 22 percent of Black students' responses, and 16 percent of Hispanic students' responses. Students from advantaged urban, from extreme rural, and from "other" communities all provided a significantly higher percentage of acceptable responses than students from disadvantaged urban communities. No significant differences were observed between regions or between male and female students.

---

[43] Hartman, D.K., "The Intertextual Links of Readers Using Multiple Passages: A Postmodern/Semiotic/Cognitive View of Meaning Making." In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical Models and Processes of Reading* (Newark, DE: International Reading Association, pp. 616-636, 1994).

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Cady's Life: Something Anne Frank Could Do*, Grade 8, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	33 (1.4)	50 (1.3)	17 (1.1)
Region			
Northeast	37 (1.9)	48 (2.9)	15 (2.2)
Southeast	34 (4.2)	50 (4.1)	16 (2.7)
Central	32 (2.2)	54 (1.9)	13 (1.3)
West	29 (2.6)	49 (1.9)	22 (2.0)
Race/Ethnicity			
White	37 (1.7)	47 (1.6)	15 (1.3)
Black	22 (2.3)	56 (2.6)	23 (2.8)
Hispanic	16 (2.4)	62 (3.5)	22 (2.5)
Gender			
Male	31 (2.0)	49 (2.2)	21 (1.6)
Female	35 (1.5)	52 (1.5)	13 (1.1)
Type of Community			
Advantaged Urban	46 (5.4)	42 (5.8)	12 (2.2)
Disadvantaged Urban	20 (2.7)	57 (3.2)	24 (2.9)
Extreme Rural	40 (3.3)	47 (3.9)	14 (3.4)
Other	32 (1.5)	51 (1.3)	17 (1.3)
Type of School			
Public	32 (1.5)	51 (1.3)	18 (1.2)
Private*	43 (4.4)	48 (4.4)	9 (1.4)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

The question that follows required students to think critically about and to interpret the text of "Cady's Life." This question tapped students' ability to understand figurative language, an important reading skill.

QUESTION 3: *Explain what the author means when she says that slamming doors symbolized the closing of the door of life.*

Unacceptable responses demonstrated an inability to interpret the text and to explain the author's meaning. They were often vague, either restating the question or presenting thoughts about the text without explanation.

Acceptable responses showed an understanding of the text and of the symbol of the slamming doors sufficient for interpreting how the symbol reveals one of the text's important meanings. Acceptable responses like those shown below focused on how the slamming doors meant that people were being taken away and probably killed, or prevented from returning to their ways of life.

As shown in Table 4.4, this was an easier question for eighth graders than either of the previous two example questions. More than one-half of the students provided acceptable responses. With this question, they were asked to explain the author's use of figurative language. Given the events of the story surrounding the abduction of Jews, "the closing of the door of life" should have represented rather straightforward symbolism for students. However, as demonstrated in the example unacceptable responses, many students interpreted the phrase literally or failed to connect it to the story's description of how Jews were being treated. Thirty-nine percent of student responses to this question were scored as unacceptable, and 7 percent of students omitted the question. The percentage of acceptable student responses from disadvantaged urban communities was significantly lower (38 percent) than all other types of communities. Fifty-nine percent of White students' responses received an acceptable score, compared to 42 percent of Hispanic students' responses and 40 percent of Black students' responses. Female students earned a higher percentage of acceptable scores than male students. Once again, private school students performed significantly better than public school students. No significant differences were observed between regions.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Cady's Life: Slamming Doors Symbolized Closing the Door of Life,* Grade 8, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	54 (1.7)	39 (1.9)	7 (0.8)
Region			
Northeast	52 (1.9)	41 (2.5)	7 (1.3)
Southeast	51 (2.8)	42 (3.2)	7 (1.5)
Central	62 (4.8)	33 (5.1)	5 (1.3)
West	53 (2.9)	39 (3.1)	8 (1.8)
Race/Ethnicity			
White	59 (1.9)	36 (2.1)	6 (0.9)
Black	40 (4.0)	50 (4.6)	10 (2.2)
Hispanic	42 (3.0)	48 (3.4)	10 (2.2)
Gender			
Male	50 (2.0)	41 (2.1)	9 (1.2)
Female	59 (2.3)	37 (2.3)	4 (0.8)
Type of Community			
Advantaged Urban	63 (5.6)	30 (4.5)	7 (2.7)
Disadvantaged Urban	38 (3.3)	48 (4.0)	14 (2.9)
Extreme Rural	64 (7.3)	34 (6.9)	2 (1.3)
Other	54 (1.9)	40 (2.2)	6 (0.9)
Type of School			
Public	53 (1.9)	40 (2.1)	7 (0.9)
Private*	68 (3.9)	29 (3.9)	3 (1.1)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Grade 8: Cady's Life — Extended-Response Question

The extended question posed to eighth graders who read Anne Frank's short story, "Cady's Life," was based on another intertextual task. It asked students to explore the relationships between Frank's own life as elaborated in the introduction, and the poem by a different author.

QUESTION: *How does the poem "I Am One" help you to understand Anne Frank's life? Use information from the introduction to the story to explain your ideas.*

To respond to this question beyond a cursory level, students needed to understand both the poem and the information about Anne Frank's life sufficiently to perceive connections between them, to describe at least one issue they both deal with, and to explain the relationship based on background information about Anne Frank's life provided in the introduction. As with other extended-response questions, responses were scored on a four-point scale, from *unsatisfactory* to *extensive*.

Unsatisfactory understanding was reflected in responses that exhibited little or no understanding of the poem or of Anne Frank's life, or did not posit a relationship between the two. Often they focused on trivial or tangential issues. For example:

As displayed in Table 4.5, 62 percent of the eighth graders provided responses indicating unsatisfactory understanding, and another 10 percent did not respond at all. Thus, nearly three-quarters of Grade 8 students were unable to demonstrate even a partial understanding of the relationship between the poem and Anne Frank's life.

Percentage of Responses for the Extended Constructed-Response Question, *Cady's Life — How the Poem 'I Am One' Helps to Understand Anne Frank's Life*, Grade 8, 1992 Reading Assessment

	Not Rated	Unsatis-factory	Partial	Essential	Extensive	Essential or Better
Nation	10 (1.0)	62 (1.3)	17 (1.0)	9 (0.8)	2 (0.4)	11 (1.0)
Region						
Northeast	10 (2.0)	58 (2.8)	18 (1.7)	11 (1.7)	3 (1.3)	13 (2.3)
Southeast	11 (2.4)	66 (2.2)	15 (2.3)	7 (1.3)	1 (0.5)	9 (1.5)
Central	8 (1.4)	64 (2.8)	17 (1.8)	9 (1.2)	3 (0.8)	12 (1.7)
West	12 (1.8)	59 (2.6)	18 (1.5)	9 (1.7)	2 (0.8)	11 (2.1)
Race/Ethnicity						
White	9 (1.2)	59 (1.6)	19 (1.1)	11 (0.9)	3 (0.6)	13 (1.2)
Black	12 (2.3)	74 (2.5)	10 (2.4)	2 (1.0)	1 (0.5)	3 (1.1)
Hispanic	16 (6.4)	65 (3.5)	15 (3.1)	3 (1.1)	0 (0.5)	4 (1.2)
Gender						
Male	14 (1.4)	65 (1.9)	14 (1.2)	6 (0.9)	1 (0.6)	7 (1.0)
Female	7 (1.1)	58 (1.8)	20 (1.6)	11 (1.3)	3 (0.6)	15 (1.4)
Type of Community						
Advantaged Urban	6 (1.8)	46 (3.7)	23 (3.1)	18 (3.8)	6 (2.1)	24 (4.3)
Disadvantaged Urban	13 (2.5)	71 (2.3)	14 (2.3)	2 (0.9)	0 (0.0)	2 (0.9)
Extreme Rural	8 (3.0)	59 (3.9)	22 (3.9)	8 (3.0)	4 (1.6)	11 (3.8)
Other	11 (1.3)	63 (1.6)	16 (1.1)	8 (1.0)	2 (0.4)	10 (1.1)
Type of School						
Public	11 (1.1)	63 (1.5)	17 (1.0)	8 (0.9)	2 (0.5)	9 (1.1)
Private*	3 (1.0)	52 (2.8)	21 (3.5)	18 (3.6)	5 (1.2)	23 (3.6)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

* The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Partial understanding was indicated by responses that provided some evidence that the student understood the relationship between the poem and Anne Frank's life, but these responses usually suggested a relationship without concrete explanation or relevant examples. Some 17 percent of eighth grade students provided responses at this level. For example:

Essential understanding was demonstrated by responses that suggested a relationship between the poem and Anne Frank's life, and explained the relationship in terms of some straightforward aspects of the war, Anne's reactions to it, and her inability to stop it. For example:

Only 9 percent of the students were able to demonstrate this level of understanding of the poem and its relationship to Anne Frank.

Extensive understanding was reflected in responses that showed evidence of richer understandings, using the relationship between the poem and Anne Frank's life to discuss the larger significance of her life, such as how she preserved history through her writing, perhaps saving others from her fate. For example:

This kind of fuller understanding of the texts and their implications was evident in only 2 percent of the students' responses. Consequently, for this extended-response question, subgroup differences in attaining the highest score level (extensive) were slight.

Essential or better understanding was demonstrated by only 11 percent of the nation's students. Again, as overall performance was so low, subgroup differences were relatively small. Of the subgroups examined, students from advantaged urban communities performed best overall, but even in this group only 24 percent demonstrated at least essential understanding, and over half (52 percent) did not respond or responded unsatisfactorily. At the other extreme, among students from disadvantaged urban communities, only 2 percent demonstrated essential or extensive understanding, and fully 84 percent were not able to respond with at least partial understanding. Students from communities classified as "other" provided significantly more essential or extensive responses than students from disadvantaged urban communities. Overall, 13 percent of White students demonstrated essential or extensive understanding as compared to 4 percent of Hispanic and 3 percent of Black students. Overall, fifteen percent of female students' responses evidenced essential or

extensive understanding, as compared to only 7 percent of male student responses. And, overall, performance by private school students (23 percent essential or better) was significantly better than public school students (9 percent essential or better).

Once again, with this question students were being asked to demonstrate intertextual understanding. In this task, however, students were required to go beyond specific ideas and to consider more global interpretations from the poem in making connections to Anne Frank's life. It is clear that this extended-response question was among the hardest for eighth graders in the NAEP assessment.

The poem's theme of striving to make a difference, even when it is unclear how influential one's individual efforts will be, is strikingly consistent with the way Anne Frank lived her life, as described in the biographical sketch. However, most eighth-grade students were unable to make this connection. The difficulty that students had with this question as well as with the second short constructed-response example question may point to the challenging nature of making intertextual connections. Perhaps this is a skill with which eighth-grade students have had little practice, and thus, few opportunities to develop.

Summary

The information provided in this chapter allows for a more in-depth view of eighth graders' reading performance than is possible from simply examining their overall proficiency. It was apparent that these students, much like their younger counterparts in fourth grade, found the multiple-choice questions easier than either the short constructed-response or extended-response questions. About two-thirds (67 percent), on average provided correct responses to the multiple-choice questions. In comparison, only about one-half (51 percent) gave acceptable responses to short constructed-response questions and about one-fourth (26 percent) demonstrated at least essential comprehension in their answers to extended-response questions. The more in-depth examinations of text meaning that were required by the extended-response questions proved to be substantially more difficult for eighth graders than the answers required for short constructed-response questions.

The sample questions presented in this chapter showed that students had moderate success in explaining relatively straightforward use of symbolism. Those students who were unsuccessful in interpreting the

author's use of figurative language, typically responded to the question literally instead of considering the symbolic use of language. On average, eighth graders had more difficulty in another question where they were asked to explain the author's use of perspective. One reason why this question may have been more difficult for these students is the need to take a critical stance in considering why the author chooses a certain technique and style. Perhaps, thinking objectively about the way a piece is written may be less familiar to students than thinking about the ideas being expressed within the passage. In order to take a critical stance, readers must "step outside" of the text. That is, they must not only think about the ideas within the text, but also think critically about the text itself.

Two of the sample questions in this chapter explicitly required students to link their understanding of two passages representing different genre. Constructing intertextual meaning appeared to be difficult for most eighth graders, as demonstrated in their responses to both a short constructed-response and an extended-response question. On the short constructed-response question, only one-third of the eighth-graders gave acceptable responses — linking a specific idea from the biographical sketch about Anne Frank to the general message of a poem. Moreover, their performance on the extended-response question provided additional evidence of eighth graders' difficulty with integrating and communicating content from multi-genre texts. Only 11 percent of the students were able to connect their understanding of a poem's theme with the essential information provided in a biographical sketch and cite support from the text for their ideas.

By the time students reach grade 12, they typically have had many opportunities to interact with different types of texts in various situations. Furthermore, it may be important for them to have had opportunities to respond to reading in various formats, particularly in writing. This chapter completes the examination of students' performance on constructed-response questions in the NAEP reading assessment by focusing on the oldest students in the NAEP assessment — twelfth graders. As with the younger students, twelfth graders were asked to respond to the same three types of questions — multiple-choice, short constructed-response, and extended response. The reading materials included in the assessment were appropriate for this more advanced level, reflecting the types of reading demands that students in their final year of secondary education would be expected to meet. Similar to the selection of materials presented to eighth-grade students, some of the reading tasks at this level required the integration of more than one text.

Table 5.1 presents the average percentage of successful responses for each of the three types of questions. As with the other two grades, twelfth-graders displayed a progression in performance across the question types — multiple-choice showing the highest percentage of successful performance, with short constructed-response questions and extended-response questions following, respectively. One slight variation in the pattern was observed at twelfth grade compared to the other grades. The difference between average percentage of acceptable response for short constructed-response questions and average percentage correct with multiple-choice questions was only 7 percent; whereas, the difference was 16 percent for eighth graders and 11 percent for fourth graders. In contrast to the results at grades 4 and 8, twelfth graders were likely to perform nearly as well on short constructed-responses questions as on multiple-choice questions. Similar to the younger students, however, they had considerable difficulty, on average, with their extended-response questions.

While nearly two-thirds of twelfth graders on average were providing acceptable answers to short constructed-response and multiple-choice questions (61 percent and 68 percent respectively), just slightly more than one third (38 percent) were able to provide responses at the essential level or better on extended-response questions. The disparity between male and female students' performance on the three question types was evident at grade 12, as it was at the other two grades. For twelfth graders, the difference between males' and females' performance on multiple-choice questions was only 2 percentage points; however, this difference increased to 7 percentage points with short constructed-response questions and 11 percentage points with extended-response questions.

Among the community types, twelfth-grade students from advantaged urban communities performed best on all types of questions with 68 percent and 73 percent respectively providing acceptable answers to short constructed-response and multiple-choice questions, and 46 percent receiving essential or better on extended-response questions. On all question types, students from communities classified as "other" and students from extreme rural communities out-performed students from disadvantaged urban communities. Also, students from communities designated as "other" out-performed their counterparts from extreme rural communities on multiple-choice and extended-response questions. On a regional basis, students from the Southeast provided the lowest percentage of acceptable responses on all question types. Disparity between public and private school students' responses was evident across all question types with private school students demonstrating higher performance.

Average Student Performance on Constructed-Response and Multiple-Choice Questions, Grade 12, 1992 Reading Assessment

	EXTENDED RESPONSE	SHORT CONSTRUCTED-RESPONSE	MULTIPLE-CHOICE
	Average Percentage Essential or Better	Average Percentage Acceptable	Average Percentage Correct
Nation	38 (0.5)	61 (0.4)	68 (0.3)
Region			
Northeast	40 (0.5)	62 (0.9)	68 (0.5)
Southeast	34 (0.8)	58 (0.7)	65 (0.8)
Central	41 (1.3)	64 (0.7)	69 (0.5)
West	38 (1.1)	62 (0.8)	69 (0.6)
Race/Ethnicity			
White	42 (0.6)	64 (0.5)	71 (0.3)
Black	28 (1.3)	51 (1.0)	58 (0.9)
Hispanic	31 (1.2)	54 (1.1)	62 (0.9)
Gender			
Male	33 (0.6)	58 (0.4)	67 (0.4)
Female	44 (0.6)	65 (0.5)	69 (0.4)
Type of Community			
Advantaged Urban	46 (1.6)	68 (1.1)	73 (1.0)
Disadvantaged Urban	30 (1.6)	54 (1.6)	60 (1.3)
Extreme Rural	35 (1.0)	60 (1.1)	66 (0.7)
Other	39 (0.7)	62 (0.5)	69 (0.4)
Type of School			
Public	37 (0.6)	55 (0.6)	63 (0.5)
Private*	48 (1.2)	68 (0.4)	74 (0.3)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

* The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Twelfth-Grade Responses to Constructed-Response Questions

The example set of materials given twelfth-grade students is summarized below. These materials were presented as a task to gain information through reading. The use of different types of source materials is typical when readers seek to understand more fully the range of issues surrounding a historical event. The full set of texts can be found in Chapter Two.

> **The Civil War in the United States: The Battle of Shiloh** contains materials from two sources, each providing a different perspective on the battle of Shiloh. The first is from a soldier's journal and provides a personal account of the war; the second is from an encyclopedia entry. Before reading, students are asked to read both texts and to see how each one makes a contribution to their understanding of the battle and of the Civil War. They are also asked to think about what each source tells that is missing from the other.

Grade 12: Battle of Shiloh — Short Constructed-Responses

The Battle of Shiloh selection required twelfth-grade students to work with two different sources about the same topic. In order to work successfully with the sources, students had to understand how each was an example of a different genre, and how each genre could provide unique insight into the Battle of Shiloh. The ability to utilize several sources while grasping the distinctions between them is a crucial component of reading for information. One of the two short constructed-response questions in the set asked students to employ this ability.

QUESTION 1: *How could reading these two sources help a student learn about the Battle of Shiloh?*

Unacceptable responses often reworded the question, or presented vague descriptions of what could be learned from either source without specific references to the Battle, or to what each source might in particular contribute. These responses showed an inability to draw distinctions between the sources.

Acceptable responses demonstrated an understanding of how the two sources, each in a particular way, add to an understanding of the Battle. Generally, acceptable responses explained that the journal gave personal information, and the encyclopedia strictly factual information about the battle.

The national results are presented in Table 5.2. Generally, twelfth-graders had little difficulty recognizing the unique contributions of the encyclopedia passage and the primary source journal entry — nearly three-fourths (73 percent) provided acceptable responses. As discussed in Chapter Seven of this report, students, on average, move ahead by twelfth grade in reading to gain information compared to reading for literary experience. This question, regarding how two different sources might contribute to learning about a topic, may be a familiar situation for these students. Encouragingly, most twelfth graders were able to cite some advantage for reading about a topic from different perspectives.

Within subgroups, students' performance on this short constructed-response question followed patterns similar to other questions at twelfth grade. Females demonstrated a higher percentage of acceptable scores than males, and private school students received more acceptable scores than public school students. Among the community types, significantly more students from advantaged urban communities received an acceptable score. Students in communities classified as "other" showed higher performance than students from disadvantaged urban communities. On this question, there was no significant statistical difference between acceptable scores by White and Hispanic students. However, Black students gave significantly fewer acceptable responses than White or Hispanic students. Twelfth graders from the West had significantly more acceptable scores than students from the Southeast region of the country.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Battle of Shiloh: Two Sources Help a Student,* Grade 12, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	73 (1.6)	25 (1.5)	2 (0.5)
Region			
Northeast	72 (3.7)	24 (3.4)	3 (1.1)
Southeast	66 (3.4)	30 (3.1)	3 (1.3)
Central	74 (3.0)	23 (3.0)	3 (1.0)
West	78 (2.2)	22 (2.3)	1 (0.4)
Race/Ethnicity			
White	77 (1.9)	21 (1.7)	2 (0.5)
Black	53 (3.7)	42 (3.5)	5 (1.8)
Hispanic	70 (5.4)	28 (5.6)	2 (1.3)
Gender			
Male	69 (2.1)	27 (2.1)	4 (0.8)
Female	77 (1.9)	22 (1.9)	1 (0.4)
Type of Community			
Advantaged Urban	87 (2.7)	11 (2.0)	2 (1.0)
Disadvantaged Urban	59 (4.4)	37 (4.3)	4 (1.0)
Extreme Rural	63 (5.5)	33 (5.5)	3 (2.3)
Other	74 (1.7)	24 (1.7)	2 (0.5)
Type of School			
Public	71 (1.7)	27 (1.6)	3 (0.6)
Private*	89 (2.0)	10 (1.8)	2 (0.8)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Another basic component of reading is the ability to make inferences about complicated representations of thought and feeling in a text. The following question required students to interpret and to make inferences about the Battle of Shiloh journal passage, in order to better understand the perspective of the officer who is the journal's narrator.

QUESTION 2: *Identify two conflicting emotions displayed by the Union officer in his journal entry. Explain why you think the battle of Shiloh caused him to have these conflicting feelings.*

Unacceptable responses tended to present feelings the officer mentioned in his journal, but failed to identify genuinely conflicting feelings, or to explain why the battle might have generated conflicting feelings.

Acceptable responses indicated an ability to understand and interpret the different references the officer makes to his feelings in the journal, and an ability to explain those references by referring to the context of the battle. Many students discussed how the officer felt both compassion and anger towards the enemy. Some described how he had to suppress feelings of sympathy in order to preserve himself.

Recognizing the presence of conflicting ideas within a passage and understanding the basis for the conflict may be considered one element of critical reading. Particularly with reading material that describes a personal account of an emotional event like a battle, readers may need to be aware of the inconsistences in human reactions and take into account their understanding of the context in making their interpretations. This question required readers to take this type of stance in thinking about ideas being expressed in the Union officer's journal entry.

Although they were not as successful with this question as they were with the last example question, more than one-half (58 percent) of the nation's twelfth-grade students were able to provide an acceptable response (see Table 5.3). Sixty-two percent of White students' responses were scored as acceptable, compared with only 43 percent of Black students' responses and 41 percent of Hispanic students' responses. Sixty-three percent of female students and 52 percent of male students achieved an acceptable score, a significant advantage for the female students. Once again, among the community types, the performance of students from advantaged urban communities achieved the highest percentage of acceptable scores. Students from communities classified as "other" performed significantly better than students from disadvantaged urban communities. As with other questions, students from private schools out-performed their public school counterparts.

Percentage of Acceptable Responses for the Short Constructed-Response Question, *Battle of Shiloh: Identify two Conflicting Emotions,* Grade 12, 1992 Reading Assessment

	Acceptable	Unacceptable	No Response
Nation	58 (1.6)	36 (1.5)	6 (0.7)
Region			
Northeast	58 (3.0)	36 (2.1)	7 (1.2)
Southeast	53 (3.5)	37 (3.5)	9 (1.5)
Central	57 (2.2)	39 (2.2)	5 (0.9)
West	61 (3.5)	33 (3.3)	6 (1.5)
Race/Ethnicity			
White	62 (1.8)	32 (1.6)	5 (0.8)
Black	43 (3.5)	44 (3.2)	12 (1.8)
Hispanic	41 (4.3)	49 (5.1)	9 (1.1)
Gender			
Male	52 (1.9)	39 (1.9)	9 (2.9)
Female	63 (2.3)	33 (2.1)	4 (0.7)
Type of Community			
Advantaged Urban	70 (2.6)	26 (2.5)	4 (1.2)
Disadvantaged Urban	42 (3.9)	50 (3.9)	8 (1.8)
Extreme Rural	46 (5.5)	44 (4.7)	10 (2.9)
Other	59 (2.0)	35 (2.0)	6 (0.8)
Type of School			
Public	56 (1.7)	38 (1.6)	7 (0.7)
Private*	71 (2.5)	26 (2.8)	3 (1.1)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Grade 12: Battle of Shiloh — Extended-Response Question

As students explore new topics, they will discover many different and sometimes conflicting accounts of similar incidents or phenomena. To develop their understanding, they must learn to reconcile different versions with one another, creating their own syntheses that recognize the differing

perspectives that authors may take, and the similarities and differences among them. One of the questions about the Battle of Shiloh explored twelfth graders' understanding of the differences in point of view and perspective in the two accounts they had read.

QUESTION 3: *Each account of the battle of Shiloh gives us information that the other does not. Describe what each account includes that the other does not. Does this mean that both accounts provide a distorted perspective of what happened in the battle?*

To respond to this question, students needed to understand not only the information included in each of the sources, but also the nature of the different kinds of information inherent in each particular genre, a personal account and encyclopedia entry, and the differences in content and experience derived from reading each.

Unsatisfactory understanding was reflected in responses that did not accurately describe what is included or alluded to in either selection, provided only an unsupported opinion about the perspectives, or listed details from one or both passages. For example:

Overall, 6 percent of twelfth-grade students demonstrated unsatisfactory understanding in their responses to this question, and another 10 percent did not attempt to respond at all (Table 5.4). Thus, 16 percent of the nation's twelfth graders were unable to demonstrate even partial understanding of the passages.

Percentage of Responses for the Extended-Response Question, *Battle of Shiloh — Information and Perspective of the Two Differing Accounts,* Grade 12, 1992 Reading Assessment

	Not Rated	Unsatis-factory	Partial	Essential	Extensive	Essential or Better
Nation	10 (0.8)	6 (0.8)	32 (1.6)	21 (1.2)	31 (1.4)	52 (1.6)
Region						
Northeast	11 (2.0)	6 (1.8)	29 (2.1)	20 (1.8)	33 (2.6)	53 (3.5)
Southeast	13 (1.3)	7 (1.0)	35 (1.9)	21 (1.3)	24 (1.5)	45 (1.8)
Central	6 (1.3)	4 (1.1)	34 (3.6)	23 (2.2)	34 (2.7)	56 (2.8)
West	10 (1.6)	8 (2.0)	29 (3.9)	21 (3.2)	32 (3.3)	53 (3.5)
Race/Ethnicity						
White	7 (0.8)	5 (0.8)	31 (1.8)	23 (1.4)	34 (1.9)	57 (1.8)
Black	18 (3.2)	11 (2.1)	34 (3.2)	16 (2.7)	21 (2.7)	36 (3.4)
Hispanic	17 (2.9)	6 (2.1)	43 (3.5)	13 (3.9)	21 (3.2)	34 (4.8)
Gender						
Male	12 (1.0)	7 (0.9)	35 (2.3)	23 (1.5)	23 (1.8)	46 (2.1)
Female	8 (1.2)	6 (1.0)	28 (1.6)	20 (1.8)	38 (2.0)	58 (1.9)
Type of Community						
Advantaged Urban	5 (1.8)	2 (1.7)	27 (2.6)	29 (4.2)	37 (3.8)	66 (2.8)
Disadvantaged Urban	18 (2.4)	6 (2.2)	38 (3.8)	16 (2.0)	23 (3.2)	39 (3.6)
Extreme Rural	11 (3.6)	10 (1.9)	31 (2.5)	27 (4.0)	21 (3.6)	48 (2.8)
Other	9 (1.0)	7 (1.0)	32 (2.1)	20 (1.3)	32 (1.8)	52 (2.1)
Type of School						
Public	11 (0.9)	7 (0.9)	33 (1.8)	22 (1.3)	28 (1.6)	49 (1.8)
Private*	6 (1.6)	2 (0.8)	23 (2.0)	17 (2.2)	52 (3.1)	69 (2.5)

The standard error of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (See Appendix for details).

*The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Partial understanding was demonstrated by responses that provided accurate information, or information from one source, or drew information from both passages but did not provide an opinion about perspectives. Nearly a third of the nation's students (32 percent) provided responses demonstrating this level of understanding. For example:

Essential understanding was demonstrated by responses that showed understanding by providing ideas and perspectives drawn from both sources. Some 21 percent of the twelfth-grade students demonstrated understanding at this level. For example:

Extensive understanding was evident in responses that showed a richer understanding of both passages and the differing perspectives they bring to the reader. These responses discussed several ideas included in each passage, and the different perspectives offered in each source. Some 31 percent of the nation's twelfth-grade students demonstrated understanding at this level. For example:

As shown in Table 5.4, more than twenty percent of the students from all groups demonstrated extensive understanding on this extended-response question. Among regions, northeast and central students performed (at the extensive level) statistically better than students from the southeast. Thirty-four percent of White students demonstrated extensive understanding, compared with 21 percent for both Black and Hispanic students. Statistically significant differences in performance at this highest level of comprehension favored females (38 percent) over males (23 percent), and advantaged urban (37 percent) over disadvantaged urban and extreme rural communities (23 percent and 21 percent respectively). Also, more students from communities classified as "other" than students from extreme rural communities demonstrated extensive understanding on this question. As with previous questions, students from private schools performed significantly better than public school students at the highest level of performance.

Across the entire assessment, only 38 percent of twelfth graders on average were able to respond with essential or better understanding on extended-response questions. With this example question, however, more than one-half (52 percent) provided answers that were rated as essential or better. In all subgroups, more than a third of the students achieved essential or better. Statistical differences within subgroups reflected higher performance by Central region students over Southeast region students, females over males, private school students over public school students, and White students over both Black and Hispanic students. Among communities, advantaged urban had the most students with essential or better responses (66 percent); and students from communities classified as "other" (52 percent) performed significantly better than students from disadvantaged urban communities (39 percent).

Such results suggest that by the twelfth-grade level many students understand the usefulness of multiple resources in gaining information about a particular topic. The ability to synthesize information from various sources is a necessary skill, both for students who plan to pursue further studies at post-secondary institutions, as well as for students who seek career opportunities after high school.

Summary

Examining students' constructed responses to reading allows us to observe the way students think about what they read and their success in constructing and extending text-based understanding. Twelfth graders displayed a pattern of performance across the three question types similar to that of their younger counterparts — extended-response questions appeared to be more difficult for them than short constructed-response or multiple-choice questions. However, twelfth graders' success with the short constructed-response questions, in comparison to younger students, was closer to their performance on multiple-choice questions.

The sample twelfth-grade questions in this chapter were associated with a reading task in which students were asked to read for the purpose of gaining information. As described later in Chapter 7 of this report, twelfth graders demonstrated higher proficiency, on average, with the informational purpose for reading than with reading for literary experience. Correspondingly, these example questions displayed some of the twelfth graders' strengths when reading to gain information. For example, when asked why reading two sources may help a student learn about a particular

topic, 73 percent were able to provide acceptable responses. With a more difficult question that tapped students' understanding of conflicting ideas in a passage, more than half (58 percent) were able to identify the conflict as well as its cause. Furthermore, more than half (52 percent) were also able to provide complete responses that demonstrated essential comprehension to an extended question about the unique perspectives offered by two types of passages about the same topic.

If the goal of reading education is to promote deeper and more complete understanding, then the responses to reading that are required of students should reflect this desired goal. The sample questions discussed in this chapter, as well as in Chapters 3 and 4, are indicative of the kinds of reading tasks that students were given in the NAEP reading assessment. As demonstrated, students needed to go beyond simply communicating understanding, they were also required to extend and examine their understanding. Through constructed-response questions, such as the ones used by NAEP, reading assessments are able to examine the more complex aspects of reading for meaning, and thus, support the emerging view of reading as an interactive, constructive process.

The 1992 NAEP reading assessment at grade 4 was supplemented with an individual interview that was conducted with a nationally-representative subsample of 1,136 4th-graders. Combining multiple indicators of developing reading abilities, the *Integrated Reading Performance Record* (IRPR), was developed as a performance-based, instructionally relevant measure of reading ability that incorporated a broad view of literacy. The individualized format allowed for in-depth appraisals of students' reading habits, attitudes, and oral reading fluency, thus, providing a more complete portrait of how the nation's fourth-graders are developing in the area of literacy.

The self-reported information about students' reading experiences, as well as the oral reading performance component, made the IRPR a highly innovative approach to reading assessment that has direct applicability to classroom instruction. The results of this special study are presented in

two separate NAEP reports, *Listening to Children Read Aloud* and *Interviewing Children About Their Literacy Experiences*. Because the IRPR was a special national study, data are not available for individual states. One important component of the IRPR was an investigation of how students' responses to comprehension questions on the NAEP reading assessment may vary by response mode.

As assessment instruments move away from total reliance on multiple-choice testing for the measurement of reading comprehension, many new modes of assessment are being explored and implemented.[44] For example, as illustrated in Chapters Three through Five, the main NAEP reading assessment was composed of approximately 60 percent constructed-response questions, in which students were asked to provide written responses demonstrating their ability to construct their own answers and interpretations of the text. This emphasis on written responses was reflective of NAEP's interactive and constructive view of reading.[45] However, students' performance was lower on the constructed-response questions than on the multiple-choice questions. The performance differences may be due to complexity and/or question format.

By constructing their own answers, students demonstrate that they can think about what they read, integrate their own knowledge and experiences with text information, and extend their understanding beyond the text ideas.[46] The constructed-response questions in the NAEP reading assessment, however, required students to write their answers, drawing on an additional language process (writing) in order to demonstrate understanding. While many researchers have suggested that reading and writing share similar constructive thinking processes and that written responses can reveal the meaning-making process involved in reading,[47] there may be some question as to the effect of young students' writing skills on their ability to demonstrate reading comprehension through written responses.

---

[44] Valencia, S.W., Hiebert, E.H., & Afflerbach, P.P. (Eds.), *Authentic Reading Assessment: Practices and Possibilities* (Newark, DE: International Reading Association, 1994).

Rhodes, L.K., & Shankin, N.L., *Windows Into Literacy: Assessing Learners K - 8* (Portsmouth, NH: Heinemann, 1993).

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

[46] Farr, R., Lewis, M., Faszholz, J., Pinsky, E., Towle, L., & Faulds, B.P., "Writing in Response to Reading," *Educational Leadership*, 66-69, 1990.

[47] Hancock, M.R., "Exploring the Meaning-Making Process Through the Content of Literature Response Journals: A Case Study Investigation," *Research in the Teaching of English*, 27, 335-368, 1993.

As one part of the IRPR, fourth-grade students were asked to read a passage that had been presented to them during the main reading assessment and to answer three of the same comprehension questions, this time orally. While these responses represented the second time students had answered the three questions, and while they were posed after reading the story a second time, it was hoped that some information would be learned about the relationship between these response modes (written and oral) and demonstrations of reading comprehension.

Eliciting Written and Oral Responses

Those fourth-grade students that were sampled for the IRPR special study had also been participants in the main NAEP reading assessment one or two days prior to the IRPR interview. During the main assessment, they had been asked to read two different passages and to respond to a combination of multiple-choice and constructed-responses questions within two separate 25-minute periods. Each of these reading assessment sections included several short constructed-response questions requiring students to answer with one or two sentences, and one extended-response question requiring students to provide at least a paragraph-length response.

One of the passages that these IRPR students had read in the main assessment was presented once again to the students during the individual interview.

More specifically, the passage contained two main characters who were animals — the turtle and the spider. The text was structured in a familiar narrative form with the development of a conflict between rival antagonists, the use of dialogue between characters, and a sequence of events leading to a climactic turning point when one character is able to gain revenge over the other for a trick that had been played early on in the story. Of the five constructed-response questions that had been administered in the main assessment with that story, one extended constructed-response and two short constructed-response questions were asked again during the interview after students had an opportunity to reread the story silently. During the IRPR, however, students did not respond in writing, instead they gave their answers orally.

The three questions were presented to these students on cards and read aloud by the interviewer. Students' responses to these questions were tape recorded. No additional prompting was provided for students so that the experience of responding to these questions during the IRPR interview could be as similar as possible to the situation during the main assessment, except for the mode of responding. Students in the IRPR were informed that the interviewer would not be able to see the students' previously written responses to these questions; therefore, similarities or differences in their answers were not important. Students were encouraged, however, to give as complete answers as possible, just has they had been asked to do during the written assessment.

Scoring Written and Oral Responses

Students' written responses to these three questions in the main assessment were scored in the same manner as all other constructed-response questions in the NAEP reading assessment. For the main assessment, scorers were trained to apply a primary-trait scoring guide in rating the written responses of students. Regular constructed-response questions were scored on a 2-point scale describing either *Acceptable* comprehension or *Unacceptable* comprehension. Extended constructed-response questions were evaluated on a 4-point scale describing increasing levels of comprehension: *Unsatisfactory, Partial. Essential,* or *Extensive.*

The scoring of the three questions that were administered in the IRPR interview study took place in a similar manner, except that scorers listened to students' taped responses rather than reading their written responses. Scorers were trained to apply the same primary-trait scoring guide that had been used to score written responses to the same questions. As a result, a comparison between students' oral and written responses to the same comprehension questions was possible.

Comparing Written and Oral Demonstrations of Comprehension

While interpretations of the results of this component of the IRPR are limited by some aspects of the study (e.g., all students responding orally had read the passage twice), data presented in Table 6.1 reveal some interesting patterns in students oral and written responses to the same comprehension questions.

Comparison Between Percentage of Written and Oral Responses to Comprehension Questions, Grade 4, 1992 Reading Assessment and IRPR

1) *What do Turtle's actions at Spider's house tell you about Turtle?* — Short Constructed-Response Question

	Written Performance	Oral Performance
Acceptable	41 (1.6)	60 (1.4)
Unacceptable	56 (1.5)	40 (1.4)
No Response	4 (0.5)	NA

2) *Who do you think would make a better friend, Spider or Turtle?* — Short Constructed-Response Question

	Written Performance	Oral Performance
Acceptable	71 (2.0)	72 (1.4)
Unacceptable	26 (1.9)	28 (1.4)
No Response	3 (0.6)	NA

3) *Pick someone you know, have read about, or have seen in the movies or on television and explain why that person is like either Spider or Turtle.* — Extended Constructed-Response Question

	Written Performance	Oral Performance
Extensive	10 (1.1)	12 (1.1)
Essential	18 (1.2)	35 (1.5)
Partial	27 (1.5)	29 (1.4)
Unsatisfactory	34 (1.8)	24 (1.4)
No Response	12 (1.4)	NA

The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

The data comparing written and oral performance reveal that when students provided oral responses to the first question, a significantly greater percentage of them were rated as "acceptable" (60 percent compared to 41 percent). On the second short constructed-response question, no significant difference in performance by response mode was observed. It should be noted that the second question was substantially easier for students in general and the appearance of differences between oral and written responses may have been constrained by the ceiling effect of over 70 percent providing acceptable responses through either mode of responding.

Comparing students' oral and written responses for the extended response question again revealed a significant advantage in providing oral responses. When students wrote their answers in the main assessment, only 28 percent demonstrated at least essential understanding. During the IRPR, 47 percent reached this level with their oral responses. This advantage was not as evident at the highest level of understanding, however. There was no significant difference in the percentage of students demonstrating extensive understanding either orally or in writing.

In interpreting the results of this response mode comparison, it is important to remember that these students gave their oral responses after reading the story and answering the same questions earlier in the main written assessment. That is, all written responses preceded their corresponding oral responses. Furthermore, nearly every student responding orally provided an answer. In contrast, a small percentage of students responding in writing did not provide a response by either skipping the question or leaving it blank. This was not the case during the one-on-one interview sessions.

Summary

With an increasing emphasis being placed on alternative procedures for assessing reading comprehension, investigations of how this new generation of assessment methods affect performance has become more important. The 1992 NAEP reading assessment was noteworthy for its reliance on constructed-response questions to measure students' understanding of what they read. These questions required students to demonstrate their understanding and describe their thinking about the passage in writing. While this format for assessment has been recognized as an effective method for observing how students go about constructing meaning from text, there may be some concern for how the process of

writing itself affects the process of meaning-making in reading and how demonstrating comprehension could be related to the mode of responding.

The response mode comparison conducted as a part of the IRPR that has been described in this chapter contributes to our understanding of two important assessment formats — written and oral responses. Although some limits are placed on the interpretation of these results since all students responded in writing before answering the same questions orally, a relatively consistent and significant finding in this study was that fourth-grade students demonstrated higher performance with their oral presentations of comprehension than with their written responses.

As set fourth in the initial chapter of this report, students need to learn to use literacy for various purposes. Because, as maturing readers, students wiil be required to respond effectively to the somewhat different demands that are imposed by different types of texts and contexts, NAEP assessed achievement according to three broad reading purposes — for literary experience, to gain information, and to perform a task.

At grade 4, there were four literary texts and four informational texts, each accompanied by approximately 10 multiple-choice and constructed-response questions. One of the questions was an extended-response question. Reading to perform a task was not assessed at grade 4. Fourth graders were given 25 minutes to read each text and answer the related questions. However, it should be noted that in accordance with a carefully specified sampling design, each fourth grader was asked to complete only two text and question sets. At grades 8 and 12, the assessment consisted of nine 25-minute text and question sets, consisting of three sets for each of the

three purposes. Each set contained a text or multiple texts accompanied by about 10 to 15 questions. Similar to grade 4, each set contained at least one extended-response question. In addition, at grade 8 there were two 50-minute sets of materials — one literary and one informational; and at grade 12 there were three such blocks — one literary and two informational. These sets of materials were based on more extensive texts or provided opportunities for students to compare and contrast materials, and included several extended-response questions. The 50-minute materials assessing literary experience at both grades 8 and 12 was based on a compendium of short stories called "The NAEP Reader," from which students selected a story to read and then answered questions about it. Because students were given the opportunity to exercise self-selection skills, these data were not included as part of the results summarized in this chapter, but the findings are reported in the following chapter.

Item response theory (IRT) methods were used to summarize results for each of the reading purposes. New for the 1992 NAEP assessment, a partial-credit scaling procedure employing a specialized IRT method was used to account for students' responses scored according to the 4-point scoring guides used with the extended-response questions. In addition, an overall composite scale was developed by weighting each subscale according to its importance in the *Reading Framework*. This chapter presents information about students' average proficiency on the NAEP scales, which range from 0 to 500, for the reading purposes.

Average Proficiency in Purposes for Reading for the Nation

Table 7.1 presents overall average proficiency and average proficiencies for the reading purposes for students in grades 4, 8, and 12.[48] As can be seen, overall average performance increased substantially from grade 4 to grade 8, and grade 8 to grade 12, as did average performance within each of the purposes for reading. However, the pattern of performance differed by reading purpose.

[48] Proficiency data are reported in this chapter to illustrate average student performance within the subdomains of reading. The focus in Chapters 3-6 was on students' performance on individual reading tasks; thus, proficiency results were not presented in those chapters but are presented here.

Average Proficiency in Purposes for Reading, Grades 4, 8, and 12, 1992 Reading Assessment

	Overall Average Proficiency	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
Grade 4	218 (1.0)	220 (1.0)	215 (1.2)	**
Grade 8	260 (0.9)	259 (1.0)	261 (1.0)	261 (1.0)
Grade 12	291 (0.6)	289 (0.7)	292 (0.6)	292 (0.7)

The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

** Reading to Perform a Task was not assessed at Grade 4.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

At grade 4, students performed better in reading for literary experience than in reading to gain information. This pattern was in contrast to average performance levels for students at grades 8 and 12. At grade 8, there were no significant differences in students' average performance across the different purposes for reading. However, by the twelfth grade, students performed better when reading to gain information or to perform a task than when reading for literary experience. Although students in higher grades displayed increased proficiency in each of the measured purposes for reading, the differences among proficiencies within each grade indicated a shift in emphasis from narrative to expository text at the upper grades. This is consistent with the view that as students progress through school, reading becomes more integral to the learning of subjects such as geography, science, and social studies, and to the application of these proficiencies in order to complete increasingly complex tasks.

Research has shown that students' abilities to perform effectively across differing reading situations may be influenced by development as well as exposure.[49] Some developmental theorists, for example, argue that students' understanding of narrative precedes their capabilities to interpret nonfictional, informational text. Constructing stories in the mind, or "storying," is considered one of the fundamental ways in which children think about the world.[50] Correspondingly, studies indicate that children's sense of the structure of stories develops rapidly when exposed to many narratives.[51]

The primacy of narrative in many school programs, however, begins to shift as students are required to apply their skills for informational purposes.[52] As students' advance in various curricular subject areas, they also need to learn to cope with texts in which both the structure and content may be less familiar. In a sense, this shift reflects the traditional conception of the differences between early and later schooling. In the earlier grade, students learn to read; in the higher grades they read to learn.

These developmental patterns are reflective of exposure to different types of text in schools. Studies indicate that students are rarely exposed to a regular diet of expository text in the early grades.[53] Similarly, although children may enter school with a firm knowledge of story structure, the bulk of their instructional time as they go through the grades tends to

---

[49] Anderson, R.C., Hiebert, E., Scott, J., & Wilkinson, I.A.G., *Becoming A Nation of Readers: The Report of the Commission on Reading*, Center for the Study of Reading (Washington, DC: National Institute of Education, 1985).

Spiro, R.A., & Taylor, B.M., "On Investigating Children's Transition from Narrative to Expository Discourse: The Multidimensional Nature of Psychological Text Classification." In R.J., Tierney, P.L. Anders, & J.N. Mitchell, *Understanding Readers' Understand: Theory and Practice*, 77-93 (Hillsdale, NJ: Erlbaum, 1987).

[50] Wells, G., *The Meaning Makers: Children Learning Language and Using Language to Learn* (Portsmouth, NH: Heinemann, 1986).

[51] Applebee, A.N., "Children's Narratives: New Directions," *The Reading Teacher*, 34, 137-142, 1980.

Stein, N.L. & Trabasso, T., "What's in a Story: An Approach to Comprehension and Instruction." In R. Glaser, *Advances In Instructional Psychology*, 2, 213-254 (Hillsdale, NJ: Earlbaum, 1982).

[52] Chall, J., *States of Reading Development* (New York, NY: McGraw-Hill, 1983).

[53] Campbell, J.R., Kapinus, B.A., & Beatty, A.S. *Interviewing Children About Their Literacy Experiences* (Washington, DC: National Center for Education Statistics, Government Printing Office, 1995).

Pappas, C., "Is Narrative "Primary"? Some Insights from Kindergartners' Pretend Readings of Stories and Information Books," *Journal of Reading Behavior*, 25, 97-129, 1993.

Hoffman, J., "Teacher and School Effect in Learning to Read." In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson, *Handbook of Reading Research*, 2, 911-950 (New York, NY: Longman, 1990).

Langer, J.A., *Children Reading and Writing* (Norwood, NJ: Ablex, 1986).

increasingly focus on expository text with fewer opportunities to use more complex and varied literary forms.[54] As students move through the transitional and high school years, reading begins to play a supportive rather than a dominant role. Granulate and Wolfe, for example, found that over 60 percent of classroom time in secondary schools was devoted to reading in an instrumental fashion.[55] More often than not, however, the preponderance of reading opportunities was in "short bursts," using reading time to locate bits of information rather than to engage in self-motivated and self-regulated reading for extended periods of time. Similarly, out-of-school reading practices take on a more practical nature. Self-reports of reading habits indicate a significant increase in the percentage of students who read informational materials including at least parts of the newspaper on a regular basis.[56]

Similar patterns have been reported in cross-cultural comparisons. Examining students' purposes for reading in grades 4, 8, and 12 in 22 industrialized and developing countries, Greaney and Neuman reported that students' functions of reading shifted from a primary focus in reading for enjoyment to an emphasis on reading to gain information and for utilitarian purposes.[57] Thus, despite wide variations in teaching practices and reading materials, these findings indicate a reading pattern which seems to take on a universal characteristic.

Given that proficiencies with rhetorical forms may reflect both development and exposure, higher proficiencies in reading for literary experience than for informational text would be expected at the fourth grade level, with a shift to informational and application materials as students reach the higher grades.

---

[54] Alverman, D. & Moore, D., "Secondary School Reading." In R. Barr, M. Kamill, P. Mosenthal, & P.D. Pearson, *Handbook of Reading Research*, 2, 951-983 (New York, NY: Longman, 1990).

[55] Granulate, M.J., & Wolfe, A.E., "Reading Instruction and Material in Content Area Classes: A Summary of Three Observational Studies." Paper presented at the meeting of the Secondary Reading Symposium (Dallas, TX: National Reading Conference, 1981).

[56] Applebee, A.N., Langer, J., & Mullis, I.V.S., *Who Reads Best? Factors Related to Reading Achievement in Grades 3, 7, and 11* (Princeton, NJ: Educational Testing Service, 1988).

Neuman, S.B., "Why Children Read: A Functional Approach," *Journal of Reading Behavior*, 12, 333-336, 1980.

[57] Greaney, V., & Neuman, S.B., "The Functions of Reading: A Cross-Cultural Perspective," *Reading Research Quarterly*, 25, 172-195, 1990.

Percentiles by Purposes for Reading

Table 7.2 shows the national percentiles by purposes for reading proficiency at grades 4, 8, and 12. Across the performance distribution at grade 4, students consistently had higher proficiency in reading for literary experience. Apparently, the highest to the lowest performing fourth graders share a common characteristic of being more competent with literary texts than with expository materials.

For the most part, eighth graders displayed no significant differences in their performance with the three purposes for reading — literary experience, gaining information, or performing a task. There was indication, however, that the very best readers in eighth grade began to shift from the dominance of literary reading evident at fourth grade and, in fact, displayed higher performance in reading to perform a task. At the 90th percentile, eighth graders had higher proficiency in reading to perform a task than either the literary or informative purposes. Also, at the 95th percentile, they had higher proficiency in reading to perform a task than reading to gain information.

As displayed in their average proficiencies, twelfth-grade students excelled in informative and task purposes compared to literary experience. However, this pattern did not remain consistent across the performance distribution. Twelfth graders from the 5th to the 25th percentiles displayed lower proficiency in literary reading compared to the other two purposes. At the 50th percentile, students continued to have higher proficiency in informative reading compared to literary reading; however, there was no significant difference between task oriented reading and the other two purposes. Beginning with the 75th percentile and continuing through the upper range of performance, students returned to higher proficiency in literary reading.

Proficiency Levels of Students at Various Percentiles by Purposes for Reading, Grades 4, 8, and 12, 1992 Reading Assessment

	Average Proficiency	5th Percentile	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile	95th Percentile
Grade 4								
Reading for Literary Experience	220 (1.0)	155 (1.4)	170 (2.0)	196 (1.3)	222 (1.4)	246 (1.3)	267 (1.2)	278 (0.8)
Reading to Gain Information	215 (1.2)	149 (1.5)	165 (1.5)	190 (1.6)	217 (1.3)	241 (1.4)	262 (2.1)	273 (1.6)
Grade 8								
Reading for Literary Experience	259 (1.0)	194 (1.3)	209 (1.4)	234 (1.5)	260 (1.2)	285 (1.1)	306 (1.1)	318 (1.5)
Reading to Gain Information	261 (1.0)	197 (1.1)	213 (1.3)	238 (1.3)	263 (1.1)	286 (1.1)	306 (0.9)	317 (1.2)
Reading to Perform a Task	261 (1.0)	193 (1.6)	210 (1.1)	236 (1.3)	263 (1.1)	289 (1.4)	310 (1.1)	322 (1.4)
Grade 12								
Reading for Literary Experience	289 (0.7)	217 (1.9)	234 (1.3)	262 (1.0)	291 (0.8)	318 (1.1)	341 (0.9)	354 (1.1)
Reading to Gain Information	292 (0.6)	237 (1.2)	251 (1.0)	272 (0.8)	294 (0.6)	314 (0.6)	331 (0.7)	341 (1.3)
Reading to Perform a Task	292 (0.7)	235 (1.2)	248 (0.9)	270 (0.9)	293 (0.9)	316 (0.9)	334 (1.0)	345 (1.3)

The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

An interesting picture emerges from these data of children's development in reading proficiencies for different purposes. Fourth graders generally demonstrate a strong inclination toward literary experiences over informative ones, due either to developmental or curricular influences, or a combination of both, and the NAEP data support these observations. All fourth graders, regardless of their reading performance, were more proficient when reading for literary experience. These data also display a distinct developmental pattern away from the dominance of literary reading abilities to more emphasis on informative reading and task-oriented reading at the higher grades. However, an interesting finding was that among the very best readers at twelfth grade, literary reading purposes reappeared as dominant over informative and task-oriented reading.

Average Proficiency in Purposes for Reading by Region

Table 7.3 presents average proficiency in purposes for reading for students attending school in four regions of the country — Northeast, Southeast, Central and West. The results indicate that the national patterns of proficiency across the purposes for reading at all three grades were generally reflected in the regions.

Average Proficiency in Purposes for Reading by Region, Grades 4, 8, and 12, 1992 Reading Assessment

	Overall Average Proficiency	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
Grade 4				
Northeast	223 (3.7)	225 (3.3)	220 (4.3)	**
Southeast	214 (2.4)	216 (2.3)	212 (2.5)	**
Central	221 (1.4)	222 (1.6)	219 (1.6)	**
West	215 (1.5)	219 (1.5)	210 (1.9)	**
Grade 8				
Northeast	263 (1.8)	262 (1.7)	265 (1.8)	264 (2.1)
Southeast	254 (1.7)	253 (1.7)	255 (1.8)	254 (2.1)
Central	264 (2.2)	261 (2.3)	266 (2.2)	268 (2.2)
West	260 (1.2)	260 (1.4)	259 (1.3)	259 (1.3)
Grade 12				
Northeast	293 (1.2)	290 (1.4)	295 (1.2)	294 (1.5)
Southeast	284 (1.1)	280 (1.4)	286 (1.2)	284 (1.2)
Central	294 (1.1)	293 (1.2)	295 (1.5)	296 (1.3)
West	292 (1.6)	292 (1.9)	293 (1.5)	292 (1.7)

The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

** Reading to Perform a Task was not assessed at Grade 4.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

At grade 4, few significant differences were found in the data. Although, students from the Central region had higher overall average proficiency and higher proficiency in reading to gain information than students from the West.

In contrast, at grade 8, a fairly consistent pattern of differences between the regions was apparent in the average proficiencies as well as in the different purposes for reading. Essentially, students in the Northeast, Central, and West regions of the country had higher proficiencies than students from the Southeast. This pattern was consistent across the three purposes for reading, although the difference between the Southeast and the West was not significant for informative reading. At grade 12, the pattern across the regions was nearly identical to that at grade 8. Average proficiencies in the Southeast were lower than the other three regions.

Across the four regions, proficiencies in general reflected the transitional pattern previously noted for the nation. Relatively higher proficiencies in reading for literary experience were observed at the fourth grade, with comparatively higher proficiencies in reading to gain information and to perform a task seen at grade 12. While instructional practices across regions surely vary, reading for informational purposes and to accomplish tasks may dominate classroom and out-of-school activities at the higher grades, as these types of reading may be perceived to be more closely connected with other forms of classroom communication across the curriculum and with students' real-life needs.

Average Proficiency in Purposes for Reading by Type of School

Table 7.4 presents average proficiency in purposes for reading for students attending public and private schools in grades 4, 8, and 12. Data for students attending private schools include Catholic school students and those attending other (non-Catholic) private schools. At all three grades, students attending private schools exceeded the performance of students attending public schools.

Average Proficiency in Purposes for Reading by Type of School, Grades 4, 8, and 12, 1992 Reading Assessment

	Overall Average Proficiency	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
Grade 4				
Public Schools	216 (1.1)	218 (1.1)	213 (1.2)	**
Private Schools*	232 (2.1)	234 (2.3)	230 (2.1)	**
Grade 8				
Public Schools	258 (1.0)	257 (1.1)	259 (1.0)	259 (1.0)
Private Schools*	278 (2.0)	277 (1.9)	279 (2.2)	281 (2.5)
Grade 12				
Public Schools	289 (0.7)	287 (0.8)	290 (0.8)	290 (0.9)
Private Schools*	307 (1.3)	304 (1.8)	309 (1.3)	307 (1.3)

The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

* The private school sample included students attending Catholic schools as well as other types of private schools. The sample is representative of students attending all types of private schools across the country.

** Reading to Perform a Task was not assessed at Grade 4.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

Patterns of reading proficiency across the various purposes for reading were strikingly similar for students in all types of schools. For example, at grade 4, proficiencies in reading for literary experience were higher than for reading to gain information for each type of school. This pattern changed in grades 8 and 12 when proficiency in reading to gain information or to perform a task was about the same or higher across the two types of schools than in reading literary experience. Therefore, while average performance was lower for students attending public as compared to private schools, the overall pattern across reading purposes remained the same as the national picture.

Average Proficiency in Purposes for Reading by Gender

The data for reading proficiency by different purposes for male and female students are presented in Table 7.5. In general, at all three grades, females had higher average reading proficiency than males in each of the reading purposes.

Average Proficiency in Purposes for Reading by Gender, Grades 4, 8, and 12, 1992 Reading Assessment

	Overall Average Proficiency	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
Grade 4				
Male	214 (1.2)	216 (1.4)	212 (1.4)	**
Female	222 (1.0)	225 (1.0)	218 (1.2)	**
Grade 8				
Male	254 (1.1)	252 (1.3)	255 (1.2)	254 (1.0)
Female	267 (1.0)	267 (1.2)	267 (1.0)	268 (1.2)
Grade 12				
Male	286 (0.7)	283 (0.9)	288 (0.8)	287 (0.9)
Female	296 (0.7)	295 (0.8)	296 (0.9)	297 (0.9)

The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

** Reading to Perform a Task was not assessed at Grade 4.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

At grade 4, girls had higher average proficiency in reading for literary experience than they did in reading to gain information. This difference was less pronounced for boys. At grade 8, performance by gender was very similar among the reading purposes. At grade 12, however, the males performed better in reading to gain information and reading to perform a task than they did in reading for literary experience. In comparison, the females showed essentially no difference in average proficiency from purpose to purpose. Thus, the gender gap was larger for reading for literary experience than it was for either of the more explanatory purposes. These findings are consistent with other research findings that males report reading more nonfiction materials than females.[58]

Average Proficiency in Purposes for Reading by Race/Ethnicity

Table 7.6 shows average reading proficiency in the various reading purposes for students in five racial/ethnic groups.

[58] Langerman, D., Books and Boys: Gender References and Book Selection, *School Library Journal*, 36, pp. 132-136, 1990.

Steve, G., & Wu, L., *Influences of Gender and Adolescent Pleasure Book Reading on Young Adult Media*, Paper presented at the Annual Meeting of the Association for Education in Journalism and Mass Communication (Kansas City, MO: 1993)

Haynes, C., & Richgels, D.J., Fourth Graders' Literature References, *Journal of Educational Research*, 85, pp. 208-219, 1992.

Average Proficiency in Purposes for Reading by Race/Ethnicity, Grades 4, 8, and 12, 1992 Reading Assessment

	Overall Average Proficiency	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
Grade 4				
White	226 (1.2)	228 (1.2)	223 (1.4)	**
Black	193 (1.7)	196 (1.7)	190 (1.9)	**
Hispanic	202 (2.2)	207 (2.5)	196 (2.1)	**
Asian/Pacific Islander	216 (3.3)	218 (3.3)	213 (3.9)	**
American Indian	208 (4.7)	210 (4.8)	204 (4.9)	**
Grade 8				
White	268 (1.2)	266 (1.3)	268 (1.2)	270 (1.2)
Black	238 (1.6)	238 (1.6)	239 (1.6)	236 (1.8)
Hispanic	242 (1.4)	242 (1.4)	242 (1.3)	240 (2.1)
Asian/Pacific Islander	270 (3.1)	271 (3.2)	270 (3.0)	269 (3.6)
American Indian	251 (3.7)	249 (3.2)	253 (4.2)	252 (5.1)
Grade 12				
White	297 (0.6)	296 (0.8)	298 (0.7)	298 (0.8)
Black	272 (1.5)	267 (1.7)	274 (1.5)	275 (1.4)
Hispanic	277 (2.4)	274 (3.3)	280 (2.0)	276 (2.7)
Asian/Pacific Islander	291 (3.2)	286 (3.7)	293 (3.1)	293 (3.7)
American Indian	272 (5.3)	267 (7.2)	274 (4.9)	275 (5.0)

The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

** Reading to Perform a Task was not assessed at Grade 4

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

The grade 4 pattern of better performance with literary than informational materials generally held across racial/ethnic groups. Also, the grade 8 pattern of little or no difference in performance by reading purpose was consistent across racial/ethnic groups. At grade 12, the data displayed some variation for the minority groups. In contrast to the White students who had similar performance across reading purposes, the Black students tended to have higher proficiency for the informational and task oriented materials.

Average Proficiency in Purposes for Reading for States

Table 7.7 presents average proficiencies in reading purposes for NAEP's Trial State Assessment Program which involved fourth-grade students attending public schools. The table is organized by overall average reading proficiency.

The pattern of average proficiency for the two purposes of reading assessed at grade 4 essentially mirrors the national picture. Across participating entities, in general, fourth graders performed better in reading for literary experience than in reading to gain information.

Figures 7.1, 7.2, and 7.3 are provided to help interpret differences in the average proficiencies across jurisdictions in overall reading, as well as in reading for literary experience and reading to gain information. The figure indicates whether or not differences between pairs of participating jurisdictions are statistically significant.[59]

For example, in Figure 7.1, although the average reading proficiencies in the fourth grade appear to be different between New Hampshire (229) and Pennsylvania (222), the difference is not statistically significant and may be due to chance factors such as sampling and/or measurement error. The computations underlying Figures 7.1, 7.2, and 7.3 take the confidence intervals or degree of sampling error associated with the estimates of average proficiency into account, as well as the estimates of average proficiency themselves. Also, the computations underlying these figures were based on data carried out to two decimal places, rather than rounded to whole numbers. As an example, Utah and Pennsylvania have the same average rounded proficiencies (222). However, in Figure 7.1, Utah's average proficiency is shown as statistically different from New Hampshire's, while Pennsylvania's average proficiency is displayed as being not statistically different from that of New Hampshire. This results from the unrounded proficiencies of Utah (221.63) and Pennsylvania (221.95) in combination with their respective standard errors.

As an example of how to read Figures 7.1, 7.2, and 7.3, compare overall average reading proficiency (Figure 7.1) in the state of Ohio to that in each of the other 41 participating states, the District of Columbia, and Guam. Reading vertically down the Figure 7.1 column labeled"Ohio," it can be seen

[59] The significance tests used in these figures are based on a Bonferroni procedure for multiple comparisons. This procedure takes into account all possible comparisons between states in declaring the differences between any two states to be statistically significant. The Bonferroni procedure holds across all possible comparisons to 5 percent the probability of erroneously declaring the averages for any two states to be different when they are not.

that, on average, fourth graders in Ohio scored lower than students in the states listed from New Hampshire through Iowa (the dark gray shaded states), about the same as students in all the states listed from Wisconsin through New Mexico (the white, or unshaded states), and better than students in the jurisdictions listed from South Carolina through Guam (light gray shading).

From Figure 7.1, we see that the cluster of highest-performing states was quite large, consisting of 14 states. The states whose fourth graders had the highest average reading proficiency were New Hampshire, Maine, Massachusetts, North Dakota, Iowa, Wisconsin, Wyoming, New Jersey, Connecticut, Nebraska, Indiana, Minnesota, Virginia, and Pennsylvania.

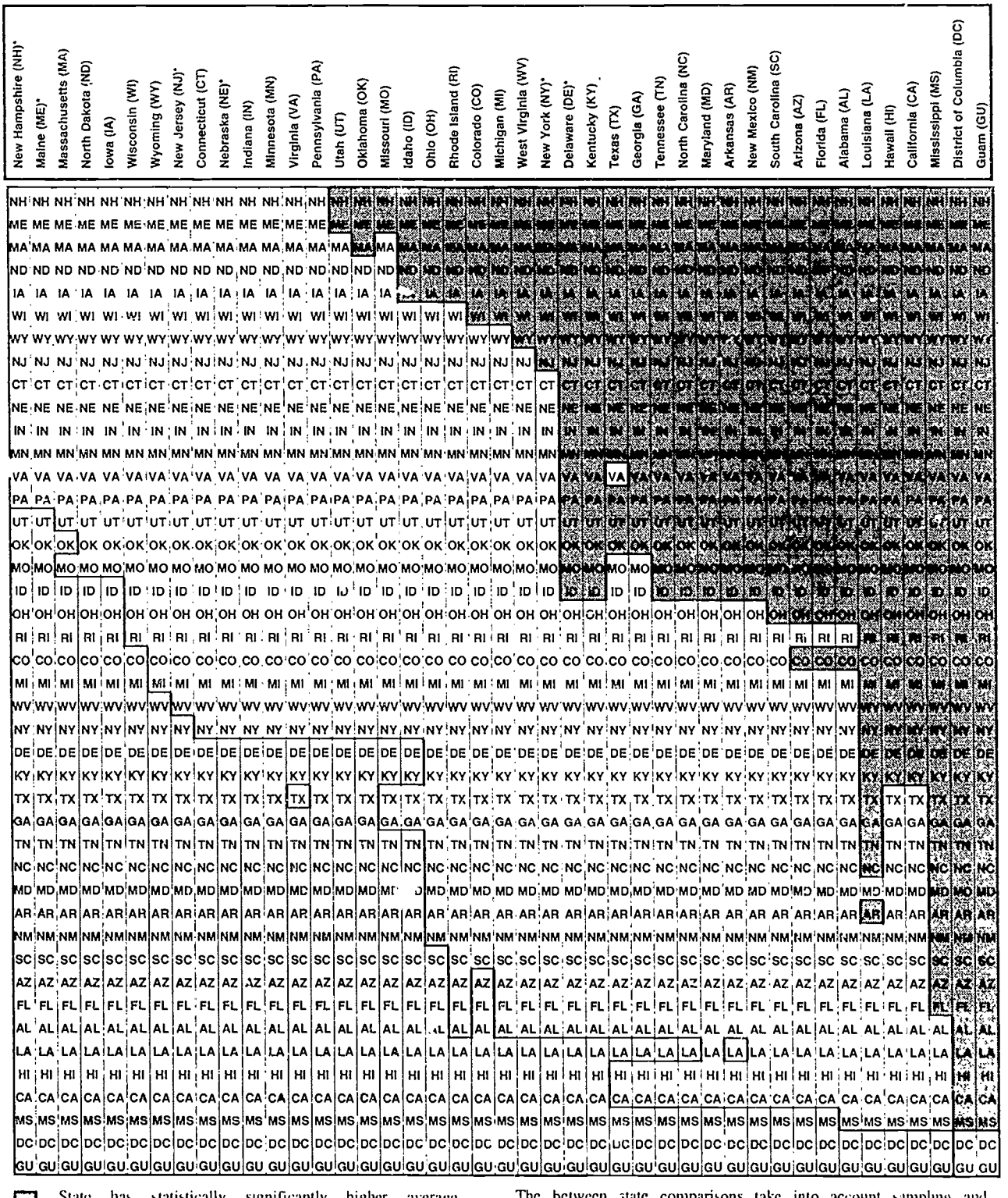
According to Figure 7.2, the top 14 states in reading for literary experience included New Hampshire, Maine, North Dakota, Massachusetts, Wyoming, Iowa, Wisconsin, New Jersey, Connecticut, Indiana, Nebraska, Virginia, Pennsylvania, and Minnesota.

The 16 states with the highest average performance in reading to gain information, as displayed in Figure 7.3, were New Hampshire, Maine, Iowa, Massachusetts, North Dakota, New Jersey, Wisconsin, Minnesota, Oklahoma, Connecticut, Nebraska, Wyoming, Pennsylvania, Indiana, Virginia, and Missouri. Essentially, they were identical to the top-performing states reported for reading for literary experience, with the addition of Oklahoma and Missouri.

Average Proficiency in Purposes for Reading, Grade 4, 1992 Trial State Reading Assessment

	Average Proficiency	Reading for Literary Experience	Reading to Gain Information
Nation	216 (1.1)	218 (1.1)	213 (1.2)
Northeast	221 (4.0)	224 (3.5)	218 (4.7)
Southeast	212 (2.5)	213 (2.5)	209 (2.7)
Central	219 (1.6)	220 (1.8)	217 (1.7)
West	213 (1.7)	217 (1.7)	208 (2.0)
States			
Alabama	208 (1.7)	211 (1.9)	205 (1.8)
Arizona	210 (1.3)	213 (1.2)	207 (1.5)
Arkansas	212 (1.2)	213 (1.4)	210 (1.4)
California	203 (2.1)	206 (2.2)	199 (2.2)
Colorado	218 (1.2)	222 (1.2)	213 (1.4)
Connecticut	223 (1.3)	226 (1.5)	219 (1.6)
Delaware	214 (0.7)	214 (0.7)	213 (0.8)
District of Columbia	189 (0.8)	192 (0.9)	185 (1.0)
Florida	209 (1.3)	212 (1.3)	206 (1.5)
Georgia	213 (1.5)	216 (1.7)	210 (1.5)
Hawaii	204 (1.7)	207 (1.7)	201 (1.9)
Idaho	221 (1.0)	224 (1.2)	217 (1.1)
Indiana	222 (1.3)	225 (1.4)	219 (1.4)
Iowa	227 (1.1)	228 (1.1)	225 (1.5)
Kentucky	214 (1.3)	216 (1.4)	210 (1.4)
Louisiana	205 (1.2)	208 (1.3)	200 (1.3)
Maine	228 (1.1)	230 (1.2)	226 (1.3)
Maryland	212 (1.6)	215 (1.8)	208 (1.6)
Massachusetts	227 (1.0)	230 (1.2)	224 (1.1)
Michigan	217 (1.6)	220 (1.6)	213 (1.7)
Minnesota	222 (1.2)	224 (1.4)	220 (1.3)
Mississippi	200 (1.3)	201 (1.5)	198 (1.3)
Missouri	221 (1.3)	223 (1.3)	219 (1.5)
Nebraska	222 (1.1)	225 (1.2)	219 (1.4)
New Hampshire	229 (1.2)	231 (1.3)	226 (1.4)
New Jersey	224 (1.5)	226 (1.5)	222 (1.7)
New Mexico	212 (1.5)	214 (1.9)	209 (1.6)
New York	216 (1.4)	219 (1.4)	212 (1.9)
North Carolina	213 (1.2)	215 (1.3)	210 (1.3)
North Dakota	227 (1.2)	230 (1.3)	223 (1.4)
Ohio	219 (1.4)	221 (1.4)	216 (1.5)
Oklahoma	221 (1.0)	223 (1.1)	220 (1.1)
Pennsylvania	222 (1.3)	224 (1.3)	219 (1.5)
Rhode Island	218 (1.8)	221 (1.8)	214 (2.0)
South Carolina	211 (1.3)	215 (1.5)	206 (1.6)
Tennessee	213 (1.5)	216 (1.5)	210 (1.8)
Texas	214 (1.6)	216 (1.6)	210 (1.8)
Utah	222 (1.2)	224 (1.3)	219 (1.2)
Virginia	222 (1.4)	225 (1.6)	219 (1.4)
West Virginia	217 (1.3)	219 (1.4)	214 (1.5)
Wisconsin	225 (1.0)	228 (1.2)	222 (1.0)
Wyoming	224 (1.2)	229 (1.1)	219 (1.4)
Territory			
Guam	183 (1.4)	187 (1.7)	177 (1.3)

INSTRUCTIONS: Read *down* the column directly under a state name listed in the heading at the top of the chart. Match the shading intensity surrounding a state postal abbreviation to the key below to determine whether the average reading performance of this state is higher than, the same as, or lower than the state in the column heading.



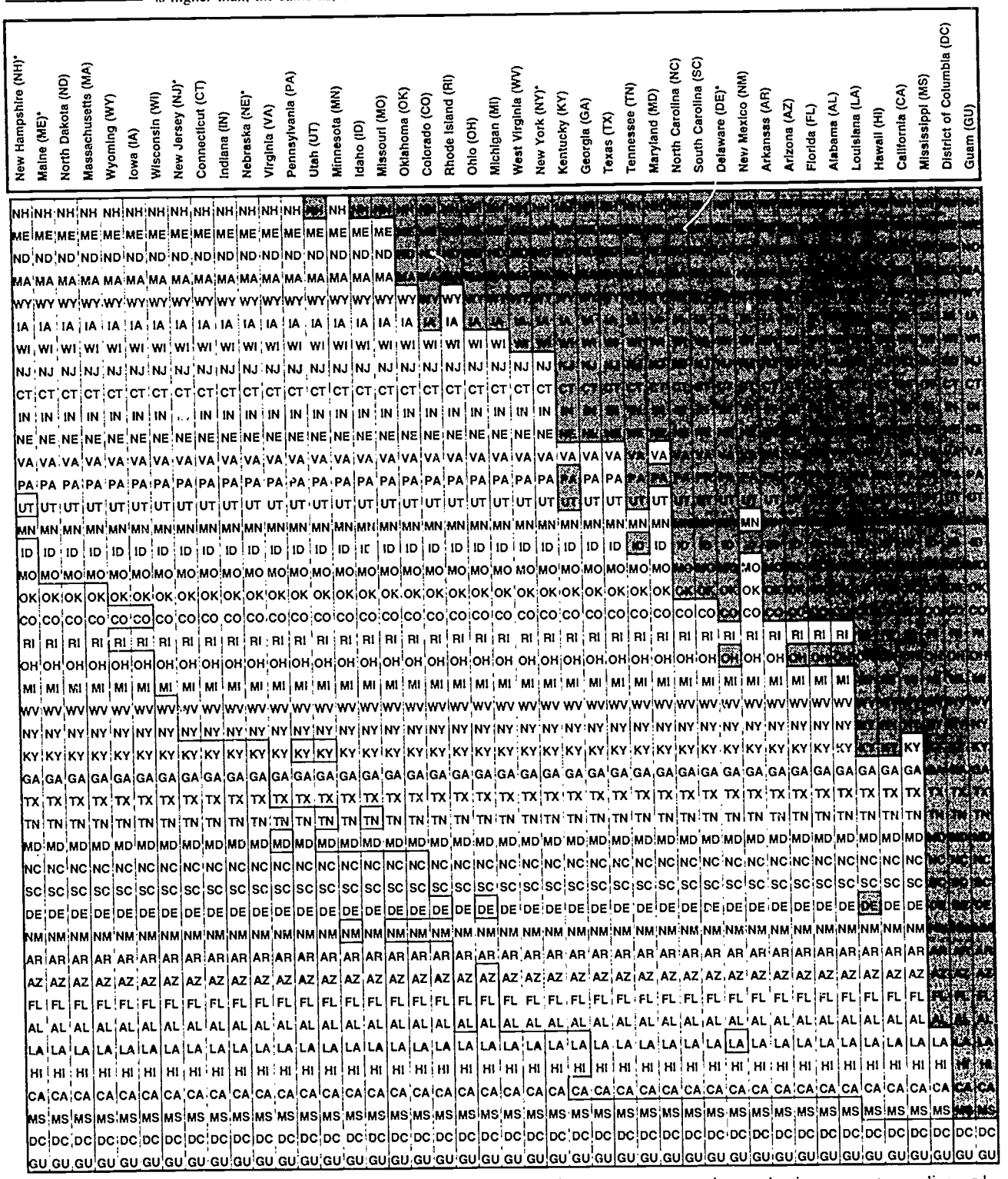
- State has statistically significantly higher average proficiency than the state listed at the top of the chart.
- No statistically significant difference from the state listed at the top of the chart.
- State has statistically significantly lower average proficiency than the state listed at the top of the chart.

The between state comparisons take into account sampling and measurement error and that each state is being compared with every other state. Significance is determined by an application of the Bonferroni procedure based on 946 comparisons by comparing the difference between the two means with four times the square root of the sum of the squared standard errors.

*Did not satisfy one or more of the guidelines for sample participation rates (see Appendix for details).

THE NATION'S REPORT CARD

INSTRUCTIONS: Read *down* the column directly under a state name listed in the heading at the top of the chart. Match the shading intensity surrounding a state postal abbreviation to the key below to determine whether the average reading performance of this state is higher than, the same as, or lower than the state in the column heading.



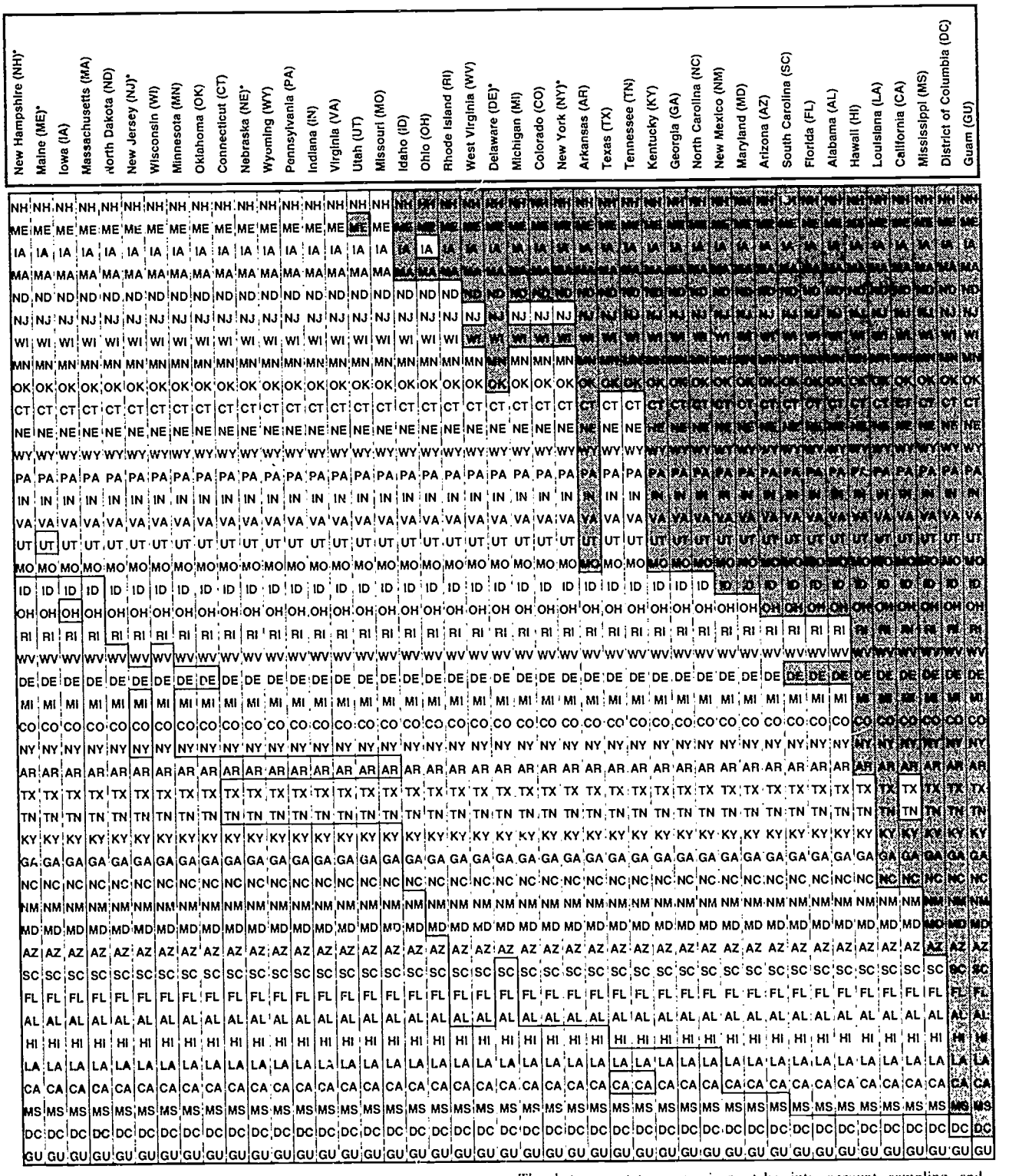
- State has statistically significantly higher average proficiency than the state listed at the top of the chart.
- No statistically significant difference from the state listed at the top of the chart.
- State has statistically significantly lower average proficiency than the state listed at the top of the chart.

The between state comparisons take into account sampling and measurement error and that each state is being compared with every other state. Significance is determined by an application of the Bonferroni procedure based on 946 comparisons by comparing the difference between the two means with four times the square root of the sum of the squared standard errors.

*Did not satisfy one or more of the guidelines for sample participation rates (see Appendix for details).

THE NATION'S REPORT CARD

INSTRUCTIONS: Read *down* the column directly under a state name listed in the heading at the top of the chart. Match the shading intensity surrounding a state postal abbreviation to the key below to determine whether the average reading performance of this state is higher than, the same as, or lower than the state in the column heading.



- State has statistically significantly higher average proficiency than the state listed at the top of the chart.
- No statistically significant difference from the state listed at the top of the chart.
- State has statistically significantly lower average proficiency than the state listed at the top of the chart.

The between state comparisons take into account sampling and measurement error and that each state is being compared with every other state. Significance is determined by an application of the Bonferroni procedure based on 946 comparisons by comparing the difference between the two means with four times the square root of the sum of the squared standard errors.

*Did not satisfy one or more of the guidelines for sample participation rates (see Appendix for details).

Summary

The analysis of reading achievement by purposes for reading showed growth across the nation in average proficiencies at grades 4, 8, and 12. However, consistent with research about students' exposure to different types of text, there were variations in these patterns of growth. Generally, students at grade 4 had higher proficiency in reading for literary experience, whereas students at grade 8 demonstrated little difference in performance across the three purposes, and students at grade 12 had higher proficiencies in reading to gain information and to perform a task. This pattern generally prevailed across public and private school students, regions, and states.

In part, these patterns reflect both development and exposure to different types of text. Many developmentalists hold that narrative or story is a more appropriate genre for young children because childrens' understanding of narrative precedes their ability to grasp informational text; thus, early experiences with stories are considered to facilitate later comprehension of text. Studies of classroom practice indicate that these widely held assumptions about development reflect curriculum practices at different grade levels. Although students have knowledge of exposition, narrative is the mainstay of instructional reading materials found in the early elementary grades.

As learners advance, they develop more efficient processing mechanisms to deal with material outside their immediate experience. Reading becomes more integrally connected with other forms of classroom communication and with the accomplishment of numerous outcomes. Older students spend much more time — both in and out of school — reading expository and informational materials.

Average proficiencies by reading purpose for region revealed that eighth- and twelfth-grade students from the Northeast, Central, and West regions had higher proficiencies than those in the Southeast. Also, students attending private schools had higher proficiencies than those attending public schools.

In general, the patterns of performance shown nationally at grades 4 and 8 also were reflected across gender and race/ethnicity. That is, across groups by gender and race/ethnicity, fourth graders consistently tended to have higher average proficiency in the literary than the informational purpose. Also, eighth graders showed little or no differences in proficiency across reading purposes regardless of gender or race/ethnicity. At grade 12, however, the groups with lower average reading proficiency performed

relatively better with the informational and task-oriented purposes. This finding is supported by the relative performance of twelfth graders at the lower and upper ends of the percentile distribution. As demonstrated in Table 7.2, twelfth graders at or below the 25th percentile in overall reading proficiency demonstrated higher achievement in reading for either informational or task purposes compared to their performance with literary reading. Conversely, at the 90th and 95th percentiles of overall reading proficiency, students had higher achievement with the literary purpose for reading. Furthermore, female and White twelfth graders showed essentially no difference in average reading proficiency across the three reading purposes, whereas, males and Black, Hispanic, Asian/Pacific Islander, and American Indian students tended to perform relatively better with the informational and task oriented purposes than in the literary purpose.

At grade 4, state-by-state analyses of performance by public school students tended to reflect regional differences. State proficiencies were generally consistent across the different purposes for reading with considerable variations in mean performance levels for high performing and low performing states and territories. In general, however, performance in the purposes was consistent with the national picture at grade 4 — higher in reading for literary experience than to gain information.

As an extension of NAEP's 1992 innovations in the assessment of reading, a special national study at grades 8 and 12 was conducted to assess students' ability to engage in an authentic literacy experience involving the self-selection of reading materials. Students were provided with a compendium of seven short stories drawn from grade-appropriate, naturally-occurring sources and asked to select one story to read.

At each grade, *The NAEP Reader*, as the compendium was titled, contained a wide array of literary pieces representing a range of genre — from mysteries to romance — and included culturally-diverse authors and topics. Students were given 50 minutes to read one of the stories and to respond to 12 constructed-response questions, three of which required extended, reflective answers. These questions were written generically, so as to allow for students selecting any one of the seven stories to respond to the same set of questions.

The impetus for developing and administering this unique assessment task grew out of a realization that literacy development involves a multitude of abilities and behaviors that evolve through engagement in personally meaningful activities.[60] Outside of measurement situations, individuals generally select the reading materials that are of particular interest or use to them for some specific reason or purpose. Some researchers have suggested that the ability to make selections of reading materials based on one's own interests and abilities is an important aspect of developing "life-long" literate behaviors.[61] Moreover, reading materials that have been self-selected may promote a sense of ownership in the literacy activity and increase motivation.[62]

The extent to which a standardized testing situation can replicate an authentic reading experience is automatically constrained by the necessity of providing the same reading materials across students, resulting in giving stories and articles to students that may or may not reflect their own interests. While collecting data that can be used for comparing students' reading abilities requires such a measurement approach, *The NAEP Reader* special study was an attempt to move somewhat beyond traditional testing constraints and make assessment more parallel to real-world types of literacy activities and more reflective of quality reading instruction. As such, it served as an appropriate complement to the innovations embodied in the 1992 NAEP reading assessment and clearly portrayed an instructionally-relevant activity.

---

[60] Carlsen, G.R., & Sherrill, A., *Voices of Readers: How We Come to Love Books.* (Urbana, IL: National Council of Teachers of English, 1988).

[61] Hiebert, E.H., Mervar, K.B., & Person, D., "Research Directions: Children's Selection of Trade Books in Libraries and Classrooms," *Language Arts, 67,* 758-763, 1990.

Lesesne, T.S., "Developing Lifetime Readers: Suggestions for Fifty Years of Research," *English Journal, 80,* 61-64, 1991

[62] Turner, J.C., "Situated Motivation in Literacy Instruction," *Dissertation Abstracts International, 53,* University Microfilms No. 93-03, 834, 1992.

Administering *The NAEP Reader* Selection Task

Nationally-representative samples of 2,138 eighth graders and 1,918 twelve graders were selected to participate in this special study. Students involved in the special study were given a copy of *The NAEP Reader* appropriate for their grade, as well as a booklet with twelve constructed-response questions. They were instructed to select a story, read it, and provide answers to the questions within the 50-minute time period.

In order to aid their selection, *The NAEP Reader* included a page of story summaries that gave students some clue as to the characters and plot of each story. In addition, the table of contents included the names of authors so that authorship could have played some role in their selection strategies. The stories were all printed in the same font and format and were equivalent in terms of length. Furthermore, the stories were determined to be similar in level of difficulty by teachers from across the country and by a committee of reading experts involved in text selection.

As previously described, the stories chosen for inclusion in *The NAEP Reader* were representative of a wide variety of literary texts. Several had been written by well-known authors and a mixture of both gender and race/ethnicity was represented among the authors. Figure 8.1 presents the story summaries as they appeared in the front of *The NAEP Reader* at each grade.

Story Summaries for *The NAEP Reader* at Grades 8 and 12, 1992 Reading Assessment

The NAEP Reader: Grade 8		The NAEP Reader: Grade 12	
story #1	Here we have a group of children in a classroom on Venus, where the sun shines for only two hours once every seven years. For one of the children, however, the sun will not shine at all.	story #1	In an attempt to salvage a failing a relationship, Alice asks Georgie to visit with her one winter evening after their break-up. As the evening progresses, their motivations for rekindling the relationship are revealed.
story #2	Being a receptionist for a publishing company got boring awfully fast for sixteen-year-old Becky. It isn't a very exciting way for an aspiring writer to spend the summer. Then obnoxious Mr. REM pops into her life.	story #2	Science rushes us into the future, yet the tools of science that have finally become part of our world are tame and represent access to a simpler past. In this science fiction story, the main character finds a new meaning for the word "nostalgia."
story #3	Confusion surrounds the illness of a young boy who has resigned himself to dying until he learns the truth about his condition.	story #3	Set against the backdrop of a bitter civil war in Dublin, Ireland at the turn of the century, a young man makes a startling discovery about the identity of his enemy.
story #4	Picking fruit all day in the hot sun is hard work. But moving from town to town and starting life over again every few months can be even more difficult.	story #4	For Cecil Rhodes, the catch of the day yields information that will change his life in a swift and calculated way.
story #5	Selling brushes door to door after school is no easy job for Donald. It is difficult to deal with the rejections, to handle the disappointments. But it is even more difficult for Donald to face his mother at home.	story #5	The punishment Nicholas receives from his aunt turns into an afternoon of delight for him in a forbidden room and an ordeal for his aunt who falls into a rain water tank.
story #6	Norman was definitely weird. For one thing all he ever did was read. Willie, on the other hand, was "a real boy" who especially loved baseball. What these two had in common came about only because a mysterious stranger came to town.	story #6	Why would someone write a check for a face cream formula in lipstick on a heart-shaped handkerchief? Who murdered the inventor of the formula? These questions and others are answered in this murder mystery.
story #7	Having the two most brilliant, most athletic, most handsome boys in the class fighting to take you to the dance might sound exciting to some girls. But while Jeff and Steve are fighting over Annie, no one has invited her best friend Brenda to the Valentine's Day dance.	story #7	A picture with a twist emerges when a dishonest portrait salesman crosses the path of Don Mateo — a man who is eager to preserve the memory of his deceased son.

Students' Selections of Stories in *The NAEP Reader*

The overall percentages of students selecting each story, as well as percentages of male and female students choosing particular stories are presented in Table 8.1.

Percentages of Students Selecting Stories from *The NAEP Reader*, Grades 8 and 12, 1992 Reading Assessment

	GRADE 8			GRADE 12		
	Total	Males	Females	Total	Males	Females
Story #1	33 (1.4)	33 (1.7)	33 (1.8)	31 (1.2)	14 (1.1)	48 (1.9)
Story #2	17 (0.8)	14 (1.3)	19 (1.2)	8 (0.6)	11 (0.9)	4 (0.7)
Story #3	15 (1.1)	18 (1.5)	11 (1.1)	18 (1.0)	31 (1.7)	4 (0.8)
Story #4	3 (0.4)	5 (0.8)	1 (0.3)	10 (0.8)	15 (1.3)	5 (0.7)
Story #5	3 (0.4)	4 (0.6)	2 (0.5)	3 (0.4)	3 (0.6)	2 (0.6)
Story #6	15 (0.7)	13 (1.2)	18 (1.2)	20 (1.0)	14 (1.2)	27 (1.5)
Story #7	8 (0.6)	6 (0.7)	10 (1.0)	5 (0.6)	5 (0.6)	6 (0.9)
No Selection	6 (0.6)	6 (0.6)	5 (0.8)	5 (1.0)	6 (1.3)	3 (0.8)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

Approximately one-third of all the students in each grade decided to read the first story in the book. Eighth-grade male and female students demonstrated similar patterns of selection among the seven stories — 33 percent of them selected the first story, a science-fiction piece. The remaining selections displayed a fairly parallel pattern, with the other most frequently selected stories being chosen by 11 to 19 percent of males and females — story 2, story 3, and story 6. A small percentage (6 percent at grade 8 and 5 percent at grade 12) did not indicate a story selection and did not respond to the comprehension questions.

At grade 12, males and females demonstrated more variations in their choices than did male and female students in grade 8. Nearly one-half (48 percent) of female twelfth graders selected the first story about the rekindling of a romantic relationship, while only 14 percent of their male counterparts chose this story. The story most frequently selected by male twelfth graders was the third one, about a young man's experience in an on-going civil war. However, only 4 percent of the females chose to read this story. Interestingly, for both males and females, the predominantly chosen story had a main character of the corresponding gender. This finding would seem to concur with previous research indicating that adolescents tend to select reading materials that include protagonists with whom they can relate or identify.[63]

The story selected by the second largest percentage of female twelfth graders was the sixth story, a murder mystery. More than one-quarter (27 percent) of the females chose this story. The remaining one-fourth of the females not selecting either story 1 or story 6 were spread out fairly evenly among the other five stories, with no more than 6 percent choosing any one of the other stories. As a result, the proportion of female twelfth graders selecting the first or sixth story accounted for 75 percent of the female students. Male twelfth graders demonstrated a wider variation in their selections. Five of the seven stories were selected by at least 10 percent of the male students.

How Students Make Reading Selections

In order to better understand how students go about the process of selecting reading material, students participating in *The NAEP Reader* special study were asked to explain on what basis they chose one story from among the seven they were given. This was a constructed-response question allowing students to describe their own unique strategies. These selection strategies were classified according to eight coding categories pertaining to the primary criteria indicated in the students' answers. The results of students' responses about how they chose a story from *The NAEP Reader* are presented in Table 8.2.

---

[63] Samuels, B.G., "Young Adult Choices: Why Do Students "Really Like" Particular Books?" *Journal of Reading*, 714-719, 1989.

The range of selection criteria used by both eighth- and twelfth-graders seemed to be rather narrow. Sixteen percent of the eighth graders relied on the title and 29 percent used the content of the stories to make their decisions. (Responses that mentioned something about the story's content but did not indicate if this information was acquired from the summaries, titles, or from browsing through the stories were coded as *content seemed interesting*.) An additional 36 percent of eighth graders did not indicate the use of any particular strategy. These students read a story from *The NAEP Reader* but did not indicate that they made their choices based on a specific criteria.

Summary of the Selection Criteria Indicated by 8th- and 12th-grade Students Choosing Stories from *The NAEP Reader*, 1992 Reading Assessment

Selection Criteria	Grade 8	Grade 12
Position in book	4 (0.7)	6 (0.7)
Author	5 (0.6)	6 (0.7)
Length of the story	3 (0.5)	3 (0.5)
Summary in front of the book	3 (0.4)	7 (0.6)
Browsed through the stories	2 (0.4)	21 (1.1)
Title	16 (0.9)	19 (1.1)
Content seemed interesting	29 (1.1)	19 (1.1)
No specific criteria	36 (1.1)	18 (1.4)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

Significantly more twelfth-grade students said that they browsed through the stories as a primary method for making their selection than did eighth graders (21 compared to 2 percent). Another 19 percent of the twelfth graders said that they used the title as selection criteria, while 19 percent also described the story's content as a major factor. Although significantly fewer than the 36 percent of eighth graders, there were still nearly one-fifth (18 percent) of the twelfth graders who did not seem to use any selection strategy in choosing a story to read from among the seven.

In general, both eighth and twelfth graders tended to use the same two or three selection strategies. However, twelfth graders were more likely than eighth graders to take the time to browse through the stories as a part of their decision process. Although several well-known authors were included in the collection (e.g., Ray Bradbury, Ernest Hemingway, and Mark Twain), only 5 percent of eighth graders and 6 percent of the twelfth graders indicated that this entered into their decision-making. There also seemed to be relatively little use of the story summaries which were provided in the front of *The NAEP Reader*. Only 3 percent of the eighth graders and 7 percent of the twelfth graders said that they used the summaries as a primary tool for their selection.

Students' Comprehension of What They Selected to Read

Students in *The NAEP Reader* special study were given 12 constructed-response questions to answer after reading the story that they selected. Nine of these questions were short constructed-response types, requiring a brief response of only one or two sentences. The remaining three questions were extended constructed-response questions in which students needed to respond with a paragraph or more in order to demonstrate the more in-depth understandings being assessed.

The short constructed-response questions were scored as demonstrating either *Acceptable* comprehension or *Unacceptable* comprehension. These questions focused on story elements such as the title's appropriateness, the story's setting, the author's use of language, qualities of the characters, and plot events. Table 8.3 displays the average percentage of students receiving an acceptable score on the short constructed-response questions for each of the seven stories at both grades.

Average Percentage of Students with Acceptable Answers on Short Constructed-Response Questions About Stories in *The NAEP Reader*, Grades 8 and 12, 1992 Reading Assessment

THE NAEP READER: GRADE 8		THE NAEP READER: GRADE 12	
Story	Average Percentage Acceptable Response	Story	Average Percentage Acceptable Response
1	35 (1.1)	1	48 (1.6)
2	30 (1.6)	2	44 (2.9)
3	35 (1.4)	3	36 (1.5)
4	33 (4.5)	4	46 (2.8)
5	27 (4.8)	5	45 (4.1)
6	29 (1.3)	6	31 (1.7)
7	25 (2.0)	7	46 (3.2)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

In general, twelfth graders appeared to have greater success in responding to the short constructed-response questions about their respective stories than did eighth graders. At grade 12, from 31 to 48 percent of the students on average provided acceptable answers to the short constructed-response questions. At grade 8, these percentages ranged from 25 to 35 percent.

As in the main assessment, the extended-response questions in *The NAEP Reader* special study were scored on a four-point scale. Responses were scored according to the level of comprehension demonstrated by the answer: *Unsatisfactory, Partial, Essential,* or *Extensive.* The first such question asked students to describe an aspect of the story that was particularly meaningful for them and to explain why. The second extended question required students to identify a major conflict in the story and to explain what the conflict was about. Finally, the third extended question asked

students to discuss how something in the story related to something that had happened to them. Table 8.4 presents the percentage of students demonstrating at least essential comprehension for each of these three extended-response questions by selected story.

Percentages of Students Demonstrating Essential or Better Comprehension on the Extended-Response Questions About Stories in *The NAEP Reader*, Grades 8 and 12, 1992 Reading Assessment

	THE NAEP READER: GRADE 8			THE NAEP READER: GRADE 12		
	First Extended Response	Second Extended Response	Third Extended Response	First Extended Response	Second Extended Response	Third Extended Response
Story #1	38 (2.8)	47 (2.7)	27 (2.3)	68 (2.5)	78 (1.8)	33 (2.7)
Story #2	32 (3.1)	45 (3.4)	24 (2.8)	50 (4.2)	61 (4.4)	30 (4.7)
Story #3	42 (3.7)	35 (4.2)	19 (2.0)	66 (3.9)	77 (2.3)	23 (3.6)
Story #4	59 (8.7)	51 (7.4)	33 (6.3)	46 (5.1)	51 (5.2)	31 (4.3)
Story #5	66 (8.2)	58 (7.2)	12 (5.2)	55 (7.6)	73 (7.4)	19 (6.5)
Story #6	36 (2.4)	42 (3.5)	27 (2.9)	37 (2.8)	55 (3.1)	16 (2.2)
Story #7	49 (4.0)	63 (3.9)	21 (4.5)	54 (5.8)	58 (6.3)	23 (6.5)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

While some variations between stories and across the two grades seem apparent from these data, it is impossible to make direct comparisons because stories were self-selected. It is clear, however, that there was a wider range of performance on these extended constructed-response questions at grade 12 than at grade 8. From 12 to 66 percent of eighth graders provided essential or better responses to the three questions, while a range of 16 to 78 percent was observed at grade 12.

At both grades, ranges of performance for these extended responses included higher levels of achievement than the ranges of performance on extended-response questions in the literary experience part of the main NAEP reading assessment. As presented in Table 8.5. the range of essential or better responses to extended questions for eighth graders on literary materials in the main assessment was from 11 to 38 percent and the range for twelfth graders was from 22 to 34 percent.

Average Percentage of Students Demonstrating Essential or Better Comprehension on the Extended Constructed-Response Questions in Main Assessment Blocks Measuring Reading for Literary Experience, Grades 8 and 12, 1992 Reading Assessment

AVERAGE PERCENTAGE ESSENTIAL OR BETTER			
8th Grade		12th Grade	
*The Flying Machine**	12 (1.1)	*The Flying Machine**	34 (1.7)
Cady's Life	11 (1.0)	*On a Mountain Trail*	22 (1.3)
Money Makes Cares	38 (1.3)	*Death of Hired Man*	34 (1.2)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details). Percentages may not total 100 percent due to rounding error. **The Flying Machine* was administered at both grades 8 and 12.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment

It would appear that some extended-response questions associated with *The NAEP Reader* elicited more in-depth demonstrations of comprehension than was observed with extended questions in the main assessment. This seemed to be particularly true with the second extended question in *The NAEP Reader*; essential or better performance ranged from 35 to 63 percent across the seven stories at grade 8 and from 51 to 78 percent at grade 12. In this question, students were asked to identify a conflict in the story and explain the nature of the conflict. In order to attain essential level understanding, students identified a conflict relevant to the story, and accurately described how the conflict was played out within the story events or between story characters. Depending on the story that was chosen, as many as one-half to three-fourths of the students were able to complete this task with at least essential level understanding.

It is important to recognize that many factors may interact to determine how well students perform in a selection task such as this one. Clearly, the nature of the passage itself will have a significant effect on how students respond to the questions. While the development committee made every attempt to ensure the comparability of story difficulty, such text characteristics as topic familiarity, identification with characters or situations, and experience with narrative structure may vary from story to story and have diverse influences on how well students understand individual stories.[64] The fact that students were given the opportunity to select their own stories may have resulted in their reading passages that fit both their past experience and their personal interests, thus, increasing the likelihood of responding successfully to the comprehension questions.

Some studies have demonstrated the positive effects that choice can have in students' literacy experiences.[65] This has been suggested by many educators as a reason for allowing more choice in students' reading programs at school.[66] In fact, many literature-based reading programs have been developed that incorporate an element of choice in students' reading assignments.[67] Although introducing choice into an assessment of reading comprehension creates some constraints on the standardization and comparability of results, it is clear that relevant information about how students perform in such situations can be achieved and used to further the discussion about the value of such literacy activities.

---

[64] As part of the 1994 reading assessment, NAEP has enhanced *The NAEP Reader* special study to enable a disentangling of the effects of selection and story difficulty.

[65] Anderson, R.C., Mason, J., & Shirley, L., "The Reading Group: An Experimental Investigation of a Labyrinth," *Reading Research Quarterly, 20*, 6-38, 1984.

[66] Cowin, R.M., "Critical Analysis of Reading Preferences of Fifth-Graders in a Self-Selective Literature-Based Reading Program," *Dissertation Abstracts International*, 52 (University Microfilm No. 91-99, 256, 1990).

Morrow, L., "Literature: Promoting Voluntary Reading," In J. Flood, J. Jensen, D. Lapp, & L. Morrow (Eds.), *Handbook of Research in Teaching the English Language Arts*, pp. 681-690 (New York NY: Macmillan, 1991).

[67] Harris, V.J., "Literature-Based Approaches to Reading Instruction." In L. Darling-Hammond (Ed.), *Review of Research in Education*, pp. 269-297, American Educational Research Association, 1993.

Summary

The NAEP Reader special study provided a unique window into students' literacy development by allowing for a more natural type of reading experience than is usually possible in assessment situations. Eighth- and twelfth-grade students were given the chance to select a short story from among seven and to demonstrate their reading ability with a passage that had some personal significance -- one they had chosen. What has been observed is that students can and do make choices when given the opportunity. Furthermore, their choices vary widely in some instances, demonstrating that students bring unique interests and ideas to the reading situation. It was also observed that some of these variations in literature selection may have some relationship to gender at grade 12. However, there was indication that similarities or differences in the story selections made by males and females was not consistent across the two grades.

One compelling aspect of students' literature selections was the lack of clear decision-making criteria indicated at both grades 8 and 12. Over one-third (36 percent) of the eighth graders and nearly one-fifth (18 percent) of the twelfth graders were unable to express a specific criteria when they responded to the question about how they made their reading choice. This inability to describe a particular reason for one's literary choices may imply either an unfamiliarity with making reading selections or an inability to articulate the criteria for those choices. Many reading experts have pointed to the self-selection of reading materials as a critical element of literacy development and as an important element of students' educational experiences. However, the results of this special study demonstrated that many students have not yet acquired specific selection strategies or that some are unable to describe on what basis they make their literary decisions.

Students demonstrated a fair amount of success in their constructed responses to questions about self-selected stories. The range of performance on extended-response questions in this special study included higher achievement than that attained by students responding to similar questions in the literary experience portion of the main NAEP reading assessment. While direct comparisons would not be appropriate given the variations in reading materials, the results indicated that selection tasks in an assessment context may provide opportunities for increased comprehension performance. This finding was particularly evident with the question about story conflicts. Across the seven stories, from 35 to 63 percent of eighth graders, and from 51 to 78 percent of twelfth graders demonstrated essential comprehension in identifying and describing a major conflict in the story they had chosen. Comparable performance on any of the extended-response questions in the literary experience portion of the main NAEP reading assessment was attained by no more than 38 percent of the students.

performance. This finding was particularly evident with the question about story conflicts. Across the seven stories, from 35 to 63 percent of eighth-graders, and from 51 to 78 percent of twelfth-graders demonstrated essential comprehension in identifying and describing a major conflict in the story they had chosen. Comparable performance on any of the extended-response questions in the literary experience portion of the main NAEP reading assessment was attained by no more than 38 percent of the students.

Appendix A

Introduction

This appendix provides further information about the methods and procedures used in NAEP's 1992 reading assessment. The *NAEP 1992 Technical Report* and the *Technical Report for the 1992 Reading Trial State Assessment* provide more extensive information about procedures.

NAEP's Reading Assessment Content

As described earlier in the report, the framework underlying NAEP's 1992 reading assessment was newly developed under the direction of the National Assessment Governing Board through a consensus process managed by the Council of Chief State School Officers.[68] The content questions, the majority of which require students to construct their own responses, and the background questionnaires were developed through a

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

similarly broad-based process managed by Educational Testing Service. The development of the 1992 reading assessment, including the Trial State Assessment Program at grade 4, benefited from the involvement of hundreds of representatives from State Education Agencies who attended numerous NETWORK meetings; served on committees; reviewed the framework, objectives, and questions; and in general, provided important suggestions on all aspects of the program. Tables A.1 and A.2 show the approximate percentage distribution of questions for the 1992 reading assessment by reading purpose, reading stance, and grade.

Table A.1

Target and Actual Percentage Distribution of Questions by Grade and Reading Purpose, 1992 Reading Assessment

	GRADE 4		GRADE 8		GRADE 12	
Reading Purpose	**Target**	**Actual**	**Target**	**Actual**	**Target**	**Actual**
Literary	55	50	40	36	35	33
Informational	45	50	40	36	45	42
Perform a Task	N/A	N/A	20	28	20	25

Table A.2

Target and Actual Percentage Distribution of Questions by Grade and Reading Stance, 1992 Reading Assessment

	GRADE 4		GRADE 8		GRADE 12	
Reading Stance	**Target**	**Actual**	**Target**	**Actual**	**Target**	**Actual**
Initial Understanding and Developing an Interpretation	33	39	33	44	33	39
Personal Response	33	27	33	22	33	23
Critical Stance	33	34	33	34	33	38

Actual percentages are based on the classifications agreed upon by NAEP's 1992 Item Development Committee. It is recognized that making discrete classifications is difficult for these categories and that independent efforts to classify NAEP questions have led to different results.[69] Also, it had been found that developing personal response questions that are considered equitable across students' different backgrounds and experiences is difficult.

[69] *Assessing Student Achievement in the States.* The First Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: 1990 Trial State Assessment (Stanford, CA: National Academy of Education, 1992).

The Assessment Design

Each student received an assessment booklet containing a set of general background questions, reading passages and content questions, a set of subject-specific background questions, and a set of questions about his or her motivation and familiarity with the assessment materials. The same booklets were used in both the national and trial state assessments. The passages and content questions were assembled into sections or blocks, each containing a passage or passages and the corresponding questions. Students were given either two 25-minute blocks or one 50-minute block.

At grade 4, the assessment consisted of eight 25-minute blocks, each containing a passage and about 10 multiple-choice and constructed-response questions. Each block contained one extended-response question. Four of the blocks were based on literary passages and four on informational materials. The special interview study of a subsample of fourth graders was only conducted in conjunction with the national assessment. Called the Integrated Reading Performance Record (IRPR), this special study consisted of an interview with individual students in which they discussed their independent reading, read aloud, provided oral responses to several constructed-response questions included in the written portion of the assessment, and described their classroom work based on examples they brought to the interview. The findings of the special IRPR study can be found in *Interviewing Children About Their Literacy Experiences* and *Listening to Children Read Aloud*.

At grades 8 and 12, the assessment consisted of nine 25-minute blocks, each containing a passage and 10 to 15 multiple-choice and constructed-response questions. Similar to grade 4, each block contained at least one extended-response question. Three of the blocks were based on literary passages, three on informational materials, and three on materials related to performing a task. In addition, at grade 8 there were two 50-minute blocks, one literary and one informational, at grade 12 there were three such blocks, one literary and two informational. These blocks were based on more extensive texts or provided opportunities for students to compare and contrast materials, and included several extended-response questions. The 50-minute block assessing literary experience at both grades 8 and 12 was based on a compendium of short stories called "The NAEP Reader," from which students selected a story to read and then answered questions about it. Because students were given the opportunity to exercise self-selection skills, there is, of course, an interaction between these skills, the story they

selected, and their assessment performance. Therefore, these data were not included as part of the 1992 NAEP reading scale reported herein, but will be included in a future report.

At grade 4, the assessment consisted of 85 questions, of which 35 required short-constructed responses and 8 required extended-responses. At grade 8, there were 135 questions, 63 of which were short constructed-response and 16 of which were extended-response. The grade 12 assessment contained 145 questions, of which 67 were short constructed-response and 19 were extended-response.

Students received different blocks of content questions in their booklets according to a specific design. The 1992 assessment was based on an adaptation of matrix sampling called balanced incomplete block (BIB) spiraling — a design that enables broad coverage of reading content while minimizing the burden for any one student. The balanced incomplete block part of the design assigns the blocks of questions to booklets in a way that provides for position effect, complete balancing within each reading purpose, and partial balancing across reading purposes. The spiraling part of the method cycles the booklets for administration, so that typically only a few students in any assessment session receive the same booklet.

National Sampling

Sampling and data collection activities for the 1992 NAEP assessment were conducted by Westat, Inc. In 1992, the assessment was conducted from January through March, with some make-up sessions in early April.

As with all NAEP national assessments, the results for the national samples were based on a stratified, three-stage sampling plan. The first stage included defining geographic primary sampling units (PSUs), which are typically groups of contiguous counties, but sometimes a single county; classifying the PSUs into strata defined by region and community type; and randomly selecting PSUs. For each grade, the second stage included listing, classifying, and randomly selecting schools, both public and private, within each PSU selected at the first stage. The third stage involved randomly selecting students within a school for participation. Some students who were selected (about 7 to 8 percent) were excluded because of limited English proficiency or severe disability.

Table A.3 presents the student and school sample sizes and the cooperation and response rates for the national assessment.

1992 Student and School Sample Sizes, 1992 Reading Assessment

	Number of Participating Schools	Percent of Schools Participating	Number of Students	Percent of Student Completion
Grade 4	527	86	6,314	93
Grade 8	587	84	9,464	89
Grade 12	468	81	9,856	81
Total	1,582		25,634	

Although sampled schools that refused to participate were occasionally replaced, school cooperation rates were computed based on the schools originally selected for participation in the assessments. The rates, which are based on schools sampled for all subjects assessed in 1992 (reading, writing, and mathematics) are also the best estimates for the reading assessment. The student completion rates represent the percentage of students assessed of those invited to be assessed in reading, including those assessed in follow-up sessions, when necessary. Of the participating schools, 944 were public schools, and 638 were Catholic and other private schools.

Trial State Assessment Sampling

For the 44 jurisdictions participating in the 1992 Trial State Assessment Program, the basic design for each grade was to select a sample of 100 public schools from each state, with a sample of 30 students drawn from each school. For states with small numbers of schools, and no or very few small schools, all schools were included in the sample with certainty. In the fourth grade, all the eligible fourth-grade schools in the District of Columbia, Delaware, and Guam were taken into the sample with certainty.

In states where a sample of schools was drawn, schools were stratified by urbanicity, minority strata (which varied by state and urbanicity level), and median income. Special procedures were used for small schools and for identifying and including new schools in the sampling frame for each jurisdiction. To minimize the potential for nonresponse bias, substitutes for nonparticipating schools were selected on a one-by-one basis to be similar to

the original school in terms of urbanicity, percent Black enrollment, percent Hispanic enrollment, median household income, and total fourth-grade enrollment. Furthermore, the substitute school was selected from the same district whenever possible.

In Guam and the Virgin Islands, all grade-eligible students were targeted for inclusion in the assessment.[70] In the remaining jurisdictions, a systematic equal probability sample of the desired number of students (usually 30, but sometimes more) was drawn from each school, typically yielding a sample size in excess of 2,500 students at each grade for each participating state and territory. Representative samples of approximately 600 to 700 public-school fourth graders in each participating state and territory responded to each question or task. The state assessments were conducted during February.

Participation Rates for States and Territories

Information about school and student participation rates for each of the 41 participating states, the District of Columbia, and Guam is summarized in Table A.4. The table also contains comparable information for the national and regional subsamples used in this report as a basis for comparison to states and territories. More specifically, these results are based only on students attending public schools (not private schools). The guidelines for receiving notations about participation are presented below. Consistent with NCES statistical standards,[71] weighted data have been used to calculate all participation rates. A discussion of the variation in participation rates is found in the *Technical Report of the 1992 Trial State Assessment in Reading*.

Since 1989, state representatives, the National Assessment Governing Board (NAGB), several committees of external advisors to the National Assessment of Educational Progress (NAEP), and the National Center for Education Statistics (NCES) have engaged in numerous discussions about the procedures for reporting the NAEP Trial State Assessment results. As part of these discussions, it was recognized that sample participation rates

[70] In Guam, students participated in both assessments. In the Virgin Islands, half the fourth graders were assigned to the mathematics assessment and half to reading.

[71] *NCES Statistical Standards*, NCES 92-021 (Washington DC: National Center for Education Statistics, U.S. Department of Education, 1992).

across the states and territories have to be uniformly high to permit fair and valid comparisons. Unless the overall participation rate is high for a state or territory, there is a risk that the assessment results for that jurisdiction are subject to appreciable nonresponse bias. Moreover, even if the overall participation rate is high, there may be significant nonresponse bias if the nonparticipation that does occur is heavily concentrated among certain classes of schools or students. Therefore, NCES established four guidelines for school and student participation in the 1990 Trial State Assessment Program.

For the 1992 Trial State Assessment, NCES decided to continue to use those four guidelines, two relating to school participation — one for overall sample participation and the other for classes of students — and two relating to student participation — one for overall sample participation and the other for classes of students. The guidelines are based on the standards for sample surveys that are set forth in the *NCES Statistical Standards*. Three of the guidelines for the 1992 program are identical to those used in 1990, while the guideline for overall school participation has been modified.

Those states receiving notations for not satisfying the guideline about overall school participation rates included Maine, Nebraska, New Hampshire, New Jersey, and New York. These five states as well as Delaware failed to meet the guideline about minimum participation rates for classes of schools with similar characteristics. Therefore, these six states are designated with asterisks in the tables and figures containing state-by-state results. All participants met or exceeded the two student participation guidelines about overall student participation rates and minimum participation rates for classes of students with similar characteristics.

The results of further study of participation rates for entities that failed to meet the sample participation guidelines are presented in the *Technical Report of the 1992 Trial State Assessment in Reading*. Evidence of significant nonresponse bias was not detected for any state. However, the participation rate data are presented so that readers of the report can accurately assess the quality of the data being presented.

The Sample Participation Guidelines

The following notations concerning school and student participation rates in the Trial State Assessment Program were established to address four significant ways in which nonresponse bias could be introduced into the jurisdiction sample estimates. The four conditions that will result in a state

or territory receiving a notation in the 1992 reports are presented below. Note that in order to receive no notations, a state or territory must satisfy all four guidelines.

A jurisdiction will receive a notation if:

1. **Both the state's weighted participation rate for the initial sample of schools was below 85 percent AND the weighted school participation rate after substitution was below 90 percent; OR the weighted school participation rate of the initial sample of schools was below 70 percent (regardless of the participation rate after substitution).**

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

To help ensure adequate sample representation for each jurisdiction participating in the 1992 Trial State Assessment Program, NAEP provided substitutes for nonparticipating schools. When possible, a substitute school was provided for each initially selected school that declined participation before November 15, 1991. For states or territories that used substitute schools, the assessment results will be based on the student data from all participating schools from both the original sample and the list of substitutes (unless both an initial school and its substitute eventually participated, in which case only the data from the initial school was used).

The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that decide not to participate in the assessment. However, considerable technical consideration was given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate bias due to the nonparticipation of initially selected schools. Thus, for the weighted school participation rates including substitute schools, the guideline was set at 90 percent.

Finally, if the jurisdiction's school participation rate for the initial sample of schools is below 70 percent, even if the rate after substitution exceeds 90 percent, there is a substantial possibility that, in aggregate, the substitute schools are not sufficiently similar to the schools that they replaced to assure that there is negligible bias in the assessment results. The last part of this guideline takes this into consideration.

A jurisdiction will receive a notation if:

2. **The nonparticipating schools included a class of schools with similar characteristics, which together accounted for more than five percent of the state's total fourth-grade weighted sample of public schools. The classes of schools from each of which a state needed minimum school participation levels were determined by urbanicity, minority enrollment, and median household income of the area in which the school is located.**

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

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

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

A jurisdiction will receive a notation if:

3. **The weighted student response rate within participating schools was below 85 percent.**

Discussion: This guideline follows the NCES standard of 85 percent for overall student participation rates. The weighted student participation rate is based on all eligible students from initially selected or substitute schools

who participated in the assessment in either an initial session or a make-up session. If the rate falls below 85 percent, then the potential for bias due to students' nonresponse is too great.

A jurisdiction will receive a notation if:

4. **The nonresponding students within participating schools included a class of students with similar characteristics, who together comprised more than five percent of the state's weighted assessable student sample. Student groups from which a state needed minimum levels of participation were determined by age of student and type of assessment session (unmonitored or monitored), as well as school urbanicity, minority enrollment, and median household income of the area in which the school is located.**

Discussion: This notation addresses the fact that if nonparticipating students are concentrated within a particular class of students, the potential for substantial bias remains, even if the overall student participation level appears to be satisfactory. Student nonresponse adjustment cells have been formed using the school-level nonresponse adjustment cells, together with the student's age and the nature of the assessment session (unmonitored or monitored). If more than five percent (weighted) of the invited students who do not participate in the assessment are from a single adjustment cell, then the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific student nonresponse rates.

Summary of School and Student Participation, Grade 4, 1992 Trial State Reading Assessment

	Weighted Percentage School Participation Before Substitution	Weighted Percentage School Participation After Substitution	Notation Number 1	Weighted Percentage Student Participation After Make-ups	Notation Number 3	Weighted Overall Rate
Nation	86	87		94		82
Northeast	80	80		95		76
Southeast	92	93		94		87
Central	92	92		95		87
West	82	83		93		77
States						
Alabama	76	97		96		93
Arizona	99	99		95		95
Arkansas	87	96		96		93
California	92	97		94		92
Colorado	100	100		95		95
Connecticut	99	99		95		94
Delaware*	92	92		95		88
District of Columbia	99	99		94		94
Florida	100	100		95		95
Georgia	100	100		96		96
Hawaii	100	100		95		95
Idaho	82	96		96		92
Indiana	77	92		96		88
Iowa	100	100		96		96
Kentucky	94	97		96		93
Louisiana	100	100		96		96
Maine*	58	71	***	95		67
Maryland	99	99		95		95
Massachusetts	87	97		96		92
Michigan	83	90		94		84
Minnesota	81	94		96		90
Mississippi	98	100		97		97
Missouri	90	97		95		93
Nebraska*	76	87	***	96		83
New Hampshire*	68	81	***	96		77
New Jersey*	76	82	***	96		79
New Mexico	76	91		95		86
New York*	78	84	***	95		79
North Carolina	95	99		96		95
North Dakota	70	91		97		89
Ohio	78	91		96		87
Oklahoma	86	98		85		83
Pennsylvania	85	95		95		91
Rhode Island	83	96		95		92
South Carolina	98	99		96		96
Tennessee	93	94		95		89
Texas	92	97		96		93
Utah	99	99		96		95
Virginia	99	99		96		95
West Virginia	100	100		96		96
Wisconsin	99	99		96		95
Wyoming	97	97		96		93
Territory						
Guam	100	100		94		94

See explanations of the notations and guidelines about sample representativeness and for the derivation of weighted participation. Notation Number 1 = Both the state's weighted participation rate for the initial sample of schools was below 85% AND the weighted school participation rate after substitution was below 90%; OR the weighted school participation rate of the initial sample of schools was below 70% (regardless of the participation rate after substitution.) Notation number 3 = The weighted student response rate within participating schools was below 85 percent.

SOURCE: National Assessment of Educational Progress (NAEP), 1992 Reading Assessment.

LEP and IEP Students

It is NAEP's intent to assess all selected students. Therefore, all selected students who are capable of participating in the assessment should be assessed. However, some students sampled for participation in NAEP can be excused from the sample according to carefully defined criteria. Specifically, some of the students identified as having Limited English Proficiency (LEP) or having an Individualized Education Plan (IEP) may be incapable of participating meaningfully in the assessment. These students are identified as follows:

LEP students may be excluded if:

- The student is a native speaker of a language other than English; AND
- He or she has been enrolled in an English-speaking school for less than two years; AND
- The student is judged to be incapable of taking part in the assessment.

IEP students may be excluded if:

- The student is mainstreamed less than 50 percent of the time in academic subjects and is judged to be incapable of taking part in the assessment, OR
- The IEP team has determined that the student is incapable of taking part meaningfully in the assessment.

When there is doubt, the student is included in the assessment.

For each student excused from the assessment, including those in the 1992 Trial State Assessment Programs, school personnel complete a questionnaire about the characteristics of that student and the reason for exclusion. Approximately 7 to 8 percent of the students nationally were excluded from the assessment. Across the participating states and territories, the percentages ranged from 2 to 12 percent at grade 4.

Data Collection

As with all NAEP assessments, data collection for the 1992 assessment was conducted by a trained field staff. For the national assessment, this was accomplished by Westat staff. However, in keeping with the legislative requirements of the Trial State Assessment Program, the state reading assessments involving approximately 110,000 fourth graders in about 4,300 schools were conducted by personnel from each of the participating states. NAEP's responsibilities included selecting the sample of schools and students for each participating state, developing the administration procedures and manuals, training the personnel who would conduct the assessments, and conducting an extensive quality assurance program.

Each participating state and territory was asked to appoint a State Coordinator to be the liaison between NAEP and participating schools. The State Coordinator was asked to gain cooperation of the selected schools, assist in scheduling, provide information necessary for sampling, and notify personnel about training. At the local school level, the administrators, usually school or district staff, were responsible for attending training, identifying excluded students, distributing school and teacher questionnaires, notifying sampled students and their teachers, administering the assessment session, completing the necessary paperwork, and preparing the materials for shipment.

Westat staff trained assessment administrators within the states in three and one-half hour sessions that included a videotape and practice exercises to provide uniformity in procedures. For the 1992 Trial State Assessment Program, which also included mathematics at grades 4 and 8, nearly 10,000 persons were trained in NAEP data collection procedures in about 500 training sessions around the nation.

To provide quality control across states, a randomly selected 50 percent of the state assessment sessions were monitored by approximately 400 quality control monitors, who were also trained Westat staff. The identity of the schools to be monitored was not revealed to state, district, or school personnel until shortly before the assessment was to commence. The analysis of the results for the unmonitored schools as compared to the monitored schools yielded no systematic differences that would suggest different procedures were used. See the *Technical Report of the 1992 Trial State Assessment in Reading* for details and results of this analysis.

Scoring

Materials from the 1992 assessment, including the Trial State Assessment Program, were shipped to National Computer Systems in Iowa City for processing. Receipt and quality control were managed through a sophisticated bar-coding and tracking system. After all appropriate materials were received from a school, they were forwarded to the professional scoring area, where the responses to the open-ended items were evaluated by trained staff using guidelines prepared by NAEP. Each open-ended question had a unique scoring guide that defined the criteria to be used in evaluating students' responses. The extended constructed-response questions were evaluated on a scale of 1 to 4, permitting degrees of partial credit to be given.

Primary-trait scoring rubrics were developed for each short and extended constructed-response question in the assessment. These rubrics were first written during the initial item development stage and were further refined during the field test of the 1992 NAEP reading assessment to reflect students' responses to and interpretations of the questions. This process was directed by the Instrument Development Committee that met in Iowa City, Iowa during the field test to review students' responses to all the questions in the assessment.

For the national reading assessment and the Trial State Assessment Program approximately 2 million student responses were scored, including a 25 percent reliability sample. The overall percentage of agreement between readers for the national reliability samples at each of the three grades assessed was 89 percent at grade 4, 86 percent at grade 8, and 88 percent at grade 12. For the Trial State Assessment Program at grade 4, the percentage of agreement across questions and states averaged 91 percent. In general, scoring reliabilities for the questions rarely dropped below 85 percent and often exceeded 90 percent exact agreement. Table A.5 contains the reliability results for the extended responses, eight of which were administered at two different grades.

Subsequent to the professional scoring, the booklets were scanned, and all information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control.

Percentages of Exact Agreement for Scoring Reliability Samples for Extended-Response Questions, 1992 Reading Assessment

	National	States	Overall
Grade 4 — Extended Questions			
Watch Out for Wombats	94	91	92
Blue Crabs	91	89	89
Spider and Turtle	90	88	88
Box in Barn	95	93	93
Sybil Sounds the Alarm	94	90	90
Amanda Clements	88	85	86
Money Makes Cares	93	93	93
Ellis Island	96	94	94
Grade 8 — Extended Questions			
Money Makes Cares	90		
Ellis Island	90		
Dorothea Dix	87		
Oregon Trail-1	87		
Oregon Trail-2	92		
Cady's Life	91		
Time Capsule	88		
Gift of Phan-1	86		
Gift of Phan-2	94		
Flying Machine	89		
Write Your Senator-1	96		
Write Your Senator-2	88		
Bus Schedule	92		
Grade 12 — Extended Questions			
On A Mountain Trail	97		
Garbage Glut	91		
Hired Man	96		
Battle of Lexington	91		
Battle of Shiloh-1	90		
Battle of Shiloh-2	90		
Battle of Shiloh-3	85		
Cali me Gentle-1	88		
Call me Gentle-2	93		
Gift of Phan-1	85		
Gift of Phan-2	92		
Flying Machine	85		
Write Your Senator-1	94		
Write Your Senator-2	87		
Bus Schedule	91		
Tax Form	87		

* Scoring extended-response questions was based on five categories: Extended, Essential, Partial, Unsatisfactory, and Not Rateable. At grades 8 and 12, the reading assessment was conducted only for the nation.

Data Analysis and IRT Scaling

After the assessment information had been compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse. Through poststratification, the weighting assured that the representation of certain subpopulations corresponded to figures from the U.S. Census and the Current Population Survey.[72]

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining the percentages of students who gave the various responses to the NAEP cognitive items, a distinction was made between missing responses at the end of each block (i.e., missing responses subsequent to the last item the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the block were considered "not reached," and treated as if they had not been presented to the student. In calculating percentages for each item, only students classified as having been presented the item were included in the denominator of the statistic.

It is standard practice at ETS to treat all nonrespondents to the last item as if they had not reached the item. For multiple-choice and short constructed-response items, the use of such a convention most often produces a reasonable pattern of results in that the proportion reaching the last item is not dramatically smaller than the proportion reaching the next-to-last item. However, for the blocks that ended with extended-response questions, use of the standard ETS convention resulted in an extremely large drop in the proportion of students attempting the final item. A drop of such magnitude seemed somewhat implausible. Therefore, for blocks ending with an extended-response question, students who answered the next-to-last item but did not respond to the extended-response question were classified as having intentionally omitted the last item.

Item response theory (IRT) was used to estimate average scale-score proficiency for the nation, various subgroups of interest within the nation, and for the states and territories. IRT models the probability of answering an item in a certain way as a mathematical function of proficiency or skill. The

[72] For additional information about the use of weighting procedures in NAEP, see Eugene G. Johnson, "Considerations and Techniques for the Analysis of NAEP Data" in *Journal of Educational Statistics* (December 1989).

main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups, such as those defined by grades, and subgroups, such as those defined by race/ethnicity or gender. Because of the BIB-spiraling design used by NAEP, students do not receive enough questions about a specific topic to provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would lead to misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain proficiency level. Instead, NAEP constructs sets of plausible values designed to represent the distribution of proficiency in the population. A plausible value for an individual is not a scale score for that individual but may be regarded as a representative value from the distribution of potential scale scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the NAEP proficiency scale are based on these plausible values. They estimate values that would have been obtained had individual proficiencies been observed — that is, had each student responded to a sufficient number of cognitive items so that proficiency could be precisely estimated.[73]

For the 1992 assessment, a scale ranging from 0 to 500 was created to report performance for each reading purpose — Literary and Informational at grade 4 and Literary, Informational, and to Perform a Task at grades 8 and 12. The scales summarize examinee performance across all three question types used in the assessment (multiple-choice, short constructed-response, and extended-response). In producing the scales, three distinct IRT models were used. Multiple-choice items were scaled using the three-parameter logistic (3PL) model; short constructed-response questions were scaled using the two-parameter logistic (2PL) model; and the extended-response tasks were scaled using a generalized partial-credit (GPC) model.[74] Recently developed by ETS and first used in 1992, the generalized partial-credit model permits the scaling of questions scored according to multi-point rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex performance tasks.

---

[73] For theoretical justification of the procedures employed, see Robert J. Mislevy, "Randomization-Based Inferences About Latent Variables from Complex Samples," *Psychometrika*, 56(2), 177-196, 1988.

For computational details, see *Focusing the New Design: NAEP 1988 Technical Report* (Princeton, NJ: Educational Testing Service, National Assessment of Education Progress, 1990) and the *1990 NAEP Technical Report*.

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

Each scale was based on the distribution of student performance across all three grades assessed in the national assessment (grades 4, 8, and 12) and had a mean of 250 and a standard deviation of 50. A composite scale was created as an overall measure of students' reading proficiency. The composite scale was a weighted average of the separate scales for the reading purposes, where the weight for each reading purpose was proportional to the relative importance assigned to the reading purpose the specifications developed through the consensus planning process as shown previously in Table A.1.

The separate reading scales do not share any items in common, and are not explicitly linked to one another. Therefore, the scores across the reading scales are, in general, not comparable. Such comparisons may be meaningful, however, in a restricted sense. Comparisons across reading scales rely on a norm referenced explanation, based on an implicit comparison to the performance of students at the other grades used in the IRT scaling. Thus, by "higher" we mean that 4th graders are closer to 8th and 12th graders on the Literary scale than they are on the Information subscale. This interpretation requires the following conditions: a) the scales compared were constructed using *cross-grade* scaling, allowing the above interpretation of comparisons; b) equivalent groups (e.g., two random samples from the same population) were used to construct the scales; and c) equivalent groups are being compared.

Linking the Trial State Results to the National Results

Although the assessment booklets used in the Trial State Assessment Program were identical to those used in the national assessment, the various differences between the national and trial state assessments, including those in administration procedures, required that careful and complex equating procedures based on a special design be used to create an appropriate basis for comparison between the national and state results.

Two separate sets of IRT-based scales (one set based on data from the trial state assessment and one set based on national assessment data) were established for the 1992 assessment. The scales from the trial state assessment were linked to those from the national assessment through a linking function determined by comparing the results for the aggregate of students assessed in the trial state assessment (except those in Guam and the Virgin Islands) with the results for students in the State Aggregate Comparison subsample of the national assessment. This subsample is

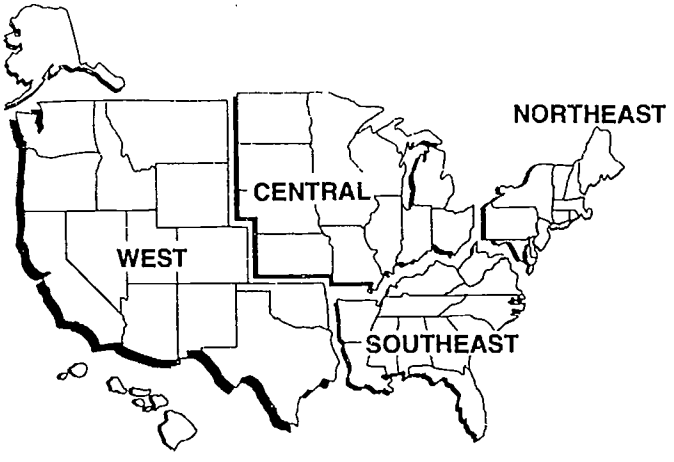
representative of the population of all grade-eligible public-school students within the aggregate of the 41 participating states and the District of Columbia who were assessed as part of the national assessment.

The linking was accomplished for each subscale by matching the mean and standard deviation of the subscale proficiencies across all students in the Trial State Assessment (excluding Guam and the Virgin Islands) to the corresponding subscale mean and standard deviation across all students in the State Aggregate Comparison subsample.

NAEP Reporting Groups

This report contains results for the nation, participating states, and groups of students within the nation and the states defined by shared characteristics. The definitions for subgroups as defined by region, race/ethnicity, gender, and type of school follow.

Region. The United States has been divided into four regions: Northeast, Southeast, Central, and West. States in each region are shown on the following map.



Race/Ethnicity. Results are presented for students of different racial/ethnic groups based on the students' self-identification of race/ethnicity according to the following mutually exclusive categories: White, Black, Hispanic, Asian/Pacific Islander, and American Indian (including Alaskan Native). Based on statistically determined criteria, at least 62 students in a particular subpopulation must participate in order for the results for that

subpopulation to be considered reliable. However, the data for all students, regardless of whether their racial/ethnic group was reported separately, were included in computing the overall national or state level results.

Gender. Results are reported separately for males and females. Gender was reported by the student.

Type of School. For the nation, results are presented separately for public-school students and for private-school students, including those attending Catholic schools and other types of private schools.

Minimum Subgroup Sampling Size

As described earlier, results for reading proficiency and background variables were tabulated and reported for groups defined by race/ethnicity and type of community, as well as by gender and parents' education level. However, in many states or territories and for some regions of the country, the number of students in some these population subgroups was not sufficiently high to permit accurate estimation of proficiency and/or background variable results. As a result, data are not provided for the subgroups with very small sample sizes. For results to be reported for any subgroup, a minimum sample size of 62 students was required. This number was determined by computing the sample size required to detect an effect size of .2 at the 5 percent significance level, with a probability of .8 or greater.

Estimating Variability

Because the statistics presented in this report are estimates of group and subgroup performance based on samples of students, rather than the values that could be calculated if every student in the nation answered every question, it is important to have measures of the degree of uncertainty of the estimates. Two components of uncertainty are accounted for in the variability of statistics based on proficiency: the uncertainty due to sampling only a relatively small number of students and the uncertainty due to sampling only a relatively small number of reading questions. The variability of estimates of percentages of students having certain background characteristics or answering a certain cognitive question correctly is accounted for by the first component alone.

In addition to providing estimates of percentages of students and their average proficiency, this report also provides information about the uncertainty of each statistic. Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate and NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any information about students that can be observed without error, but each student typically responds to so few items within any content area that the proficiency measurement for any single student would be imprecise. In this case, using plausible values technology makes it possible to describe the performance of groups and subgroups of students, but the underlying imprecision that makes this step necessary adds an additional component of variability to statistics based on NAEP proficiencies.[75]

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

Drawing Inferences from the Results

The use of *confidence intervals*, based on the standard errors, provides a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample mean proficiency ± 2 standard errors represents a 95 percent confidence interval for the corresponding population quantity.

---

[75] For further details, see Eugene G. Johnson, "Considerations and Techniques for the Analysis of NAEP Data" in *Journal of Educational Statistics* (December 1989)

This means that with approximately 95 percent certainty, the average performance of the entire population of interest is within ± 2 standard errors of the sample mean.

As an example, suppose that the average reading proficiency of students in a particular group was 256, with a standard error of 1.2. A 95 percent confidence interval for the population quantity would be as follows:

$$\text{Mean} \pm 2 \text{ standard errors} = 256 \pm 2 \cdot (1.2) = 256 \pm 2.4 = 256 - 2.4 \text{ and } 256 + 2.4 = 253.6, 258.4$$

Thus, one can conclude with 95 percent certainty that the average proficiency for the entire population of students in that group is between 253.6 and 258.4.

Similar confidence intervals can be constructed for percentages, provided that the percentages are not extremely large (greater than 90) or extremely small (less than 10). For extreme percentages, confidence intervals constructed in the above manner may not be appropriate. However, procedures for obtaining accurate confidence intervals are quite complicated. Thus, comparisons involving extreme percentages should be interpreted with this in mind.

To determine whether there is a real difference between the mean proficiency (or proportion of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the proficiency means or proportions of these groups for the sample. This estimate of the degree of uncertainty — called the standard error of the difference between the groups — is obtained by taking the square of each group's standard error, summing these squared standard errors, and then taking the square root of this sum.

Similar to the manner in which the standard error for an individual group mean or proportion is used, the standard error of the difference can be used to help determine whether differences between groups in the population are real. The difference between the mean proficiency or proportion of the two groups ± 2 standard errors of the difference represents an approximate 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between groups in the population. If the interval does not contain zero, the difference between groups is statistically significant (different) at the .05 level.

The procedures described in this 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. When one considers sets of confidence intervals, like those for the average proficiency of all participating states and territories, 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. If one wants to hold the certainty level for a specific set of comparisons at a particular level (e.g., .95), adjustments (called multiple-comparisons procedures) need to be made.

The standard errors for means and proportions reported by NAEP are statistics and subject to a certain degree of uncertainty. In certain cases, typically when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the standard errors may be quite large. Throughout this report, estimates of standard errors subject to a large degree of uncertainty are designated by the symbol "!". In such cases, the standard errors — and any confidence intervals or significance tests involving these standard errors — should be interpreted cautiously.

ACKNOWLEDGMENTS

The work presented herein represents the efforts of the thousands of individuals who are necessary to implement a project of this size and scope. Many persons have contributed the expertise, energy, and dedication necessary to develop, conduct, analyze, and report NAEP's 1992 reading assessment. Most importantly, NAEP is grateful to students and school staff who made the assessment possible.

NAEP's 1992 reading assessment, including the Trial State Assessment Program, was truly a collaborative effort among staff from State Education Agencies, the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). The program benefited from the contributions of hundreds of individuals at the state and local levels — Governors, Chief State School Officers, State and District Test Directors, State Coordinators, and district administrators — who tirelessly provided their wisdom, experience, and hard work.

The assessment was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. Emerson Elliott, Commissioner, provided consistent support and guidance. The staff — particularly Gary Phillips, Peggy Carr, Sharif Shakrani, Steve Gorman, Shi Chang Wu, Sheida White, Susan Ahmed, Eugene Owen, and Maureen Treacy — worked closely and collegially with ETS, Westat, and NCS staff and played a crucial role in all aspects of the program.

The members of the National Assessment Governing Board (NAGB) and the NAGB staff provided advice and guidance throughout. Their contractor for NAEP's reading consensus project, the Council of Chief State School Officers (CCSSO), worked diligently under tight time constraints to create the forward-looking framework underlying the assessment.

NAEP owes a debt of gratitude to the numerous panelists and consultants who provided their expertise and worked so conscientiously on developing the assessment and providing a frame for interpreting the results, including those who helped create the *NAEP Reading Framework*, and develop the assessment instruments.

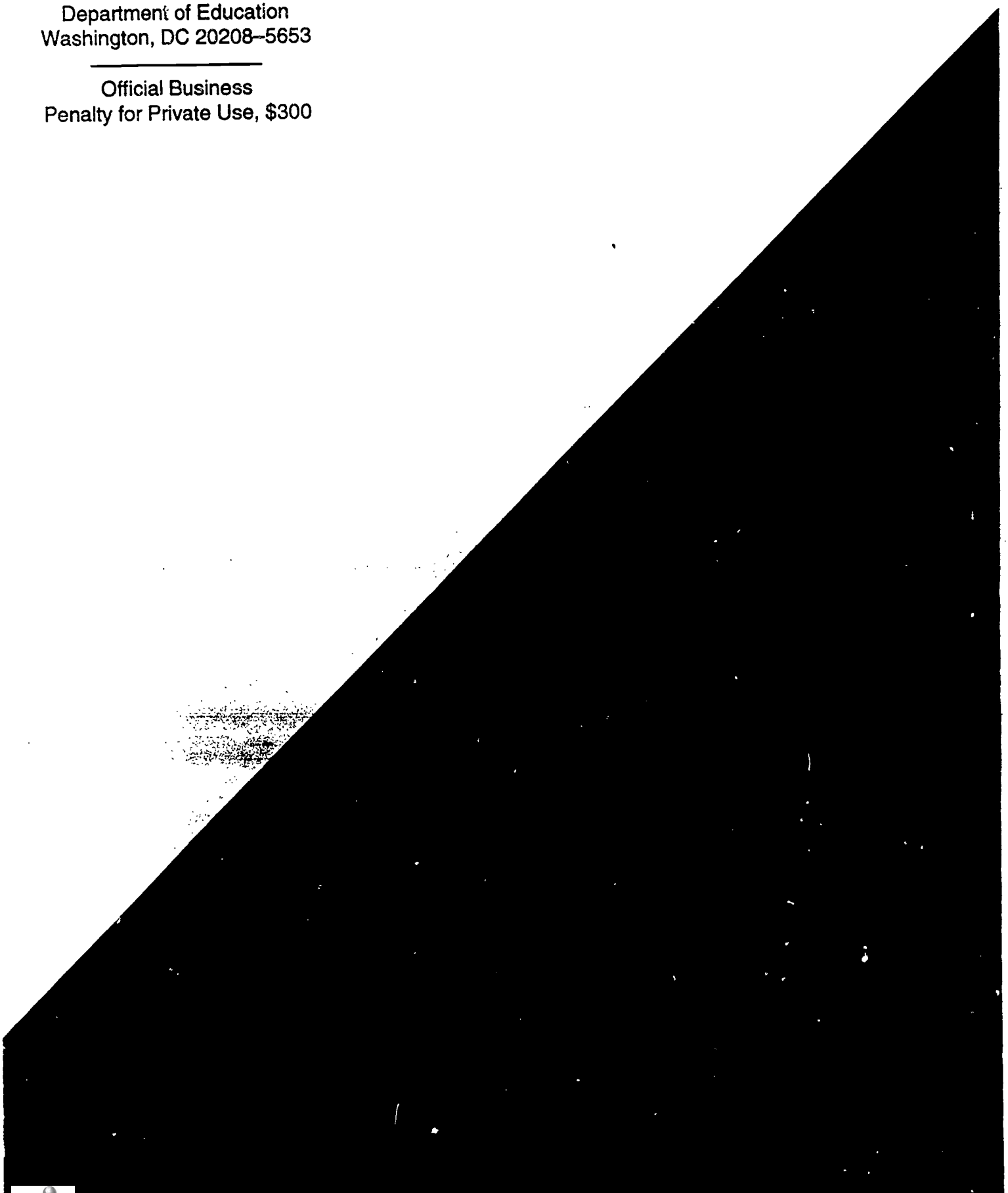
The NAEP project at ETS resides in the Center for the Assessment of Educational Progress (CAEP) managed by Archie Lapointe. Under the NAEP contract to ETS, Ina Mullis served as the Project Director. Stephen Koffler managed test development activities, and he and John Olson coordinated state services. Jay Campbell and Mary Foertsch worked with the Reading Item Development committee to develop the assessment

instruments. Jules Goodison managed the operational aspects together with John Olson, and sampling and data collection activities were carried out by Westat under the direction of Renee Slobasky, Nancy Caldwell, and Keith Rust. Printing, distribution, scoring, and processing activities were conducted by NCS, under the supervision of John O'Neill, Judy Moyer, Diane Smrdel, Lavonne Mohn, Brad Thayer, and Mathilde Kennel.

Reading and state assessment design, statistical, psychometric procedures and implementation were led by Nancy Allen and John Donoghue under the direction of Eugene Johnson and John Mazzeo. Major contributions were made by James Carlson, Hua Hua Chang, Angela Grima, Frank Jenkins, Jo-lin Liang, Eiji Muraki, Spencer Swinton, and Ming-mei Wang. Steve Isham performed the reading analysis, assisted by Drew Bowker and Yim Fai Fong. To enable accurate state-by-state reporting, Robert Patrick and David Freund designed and implemented the table generation system, and Jennifer Nelson the graphics. Judith Langer, Jay Campbell, Susan Neuman, Ina Mullis, Hilary Persky, and Patricia Donahue wrote the report, with considerable production help from Kent Ashworth, who also coordinated the cover design. The authors of this report are indebted to the numerous reviewers who suggested improvements to successive drafts. The comments and critical feedback provided by the following reviewers are reflected in the final version of this report: Marilyn Binkley, Mary Lyn Bourque, John Burkett, John Donoghue, Larry Fineberg, Andrew Kolstad, Mary Naifeh, P. David Pearson, Sheida White, and Shi Chang Wu. Sharon Davis-Johnson, with the assistance of Karen Damiano, provided the excellent desk-top publishing skills essential to the project.

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300



NCES 95–727