Reading Assessment and Item Specifications

for the 2009 National Assessment of Educational Progress

Prepared for the

National Assessment Governing Board
In support of Contract No. ED-02-R-0007
U.S. Department of Education

American Institutes for Research
1000 Thomas Jefferson Street, N.W.
Washington, DC 20007

Developed for the National Assessment Governing Board under contract number ED-02-R-0007 by the American Institutes for Research.

For further information, contact the National Assessment Governing Board:
800 North Capitol Street, N.W.
Suite 825
Washington, DC 20002
http://www.nagb.org

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF EXHIBITS

Page

EXECUTIVE SUMMARY

As the ongoing national indicator of what American students know and can do, the National Assessment of Educational Progress (NAEP) in Reading regularly collects achievement information on representative samples of students in grades 4, 8, and 12. Through the "Nation's Report Card," the NAEP Reading Assessment reports how well students perform in reading various texts and responding to those texts by answering multiple-choice and constructed-response questions. The information that NAEP provides about student achievement helps the public, educators, and policymakers understand strengths and weaknesses in student performance and make informed decisions about education.

The 2009 NAEP Reading Assessment will measure national, regional, state, district, and subgroup achievement in reading but is not designed to report individual student or school performance. The assessment will measure the reading comprehension of students and their ability to apply vocabulary knowledge to assist them in comprehending what they read. The public will have access to performance results and released questions through NAEP reports and Web sites.

This document, the *Reading Assessment and Item Specifications for the 2009 National Assessment of Educational Progress*, provides information to guide passage selection, item development, and other aspects of test development. It accompanies the *Reading Framework for the 2009 National Assessment of Educational Progress,* which presents the conceptual base for the assessment.

The recommended 2009 NAEP Reading Framework is consistent with current *No Child Left Behind* (NCLB) legislation of 2001. In accordance with NCLB, the NAEP Reading Assessment will be administered every two years at grades 4 and 8, and the resulting data will be widely reported in a timely fashion. Because the 2009 NAEP Reading Assessment will start a new trend line, NAGB decided to delay the implementation of the new Framework from 2007 to 2009. This will enable states to obtain three years of NAEP reading data at grades 4 and 8 under NLCB—2003, 2005, and 2007—under the old Framework. In addition, NAEP will assess and report grade 12 reading results every four years.

The National Assessment Governing Board (NAGB)—the policy-making body for NAEP—has stated that the NAEP assessment will measure reading comprehension by asking students to read passages written in English and to answer questions about what they have read. The Framework "shall not endorse or advocate a particular pedagogical approach…but shall focus on important, measurable indicators of student achievement."[1] Although broad implications for instruction may be inferred from the assessment, NAEP does not specify how reading should be taught, nor does it prescribe a particular curricular approach to teaching reading.

The 2009 NAEP Reading Framework recommendations result from the work of many individuals and organizations involved in reading and reading education, including researchers, policymakers, educators, and other members of the public. Members of the committees that contributed to the development are presented in Appendix A. Their work was guided by scientifically

[1]National Assessment Governing Board. (2002, May). *National Assessment Governing Board policy on framework development*. Washington, DC: Author.

based literacy research that conceptualizes reading as a dynamic cognitive process, as reflected in the following definition of reading:

Reading is an active and complex process that involves:

- understanding written text;
- developing and interpreting meaning; and
- using meaning as appropriate to type of text, purpose, and situation.

This definition applies to the assessment of reading achievement on NAEP and is not intended to be an inclusive definition of reading or of reading instruction. Procedures for operationalizing this definition are presented in this document.

The 2009 NAEP Reading Specifications translate the NAEP Reading Framework into guidelines for selecting passages, developing items and for constructing the assessment as a whole. The primary purpose of the Specifications is to provide the test development contractor with information that will ensure that the NAEP Reading Assessment reflects the intent of NAGB. The recommendations and guidelines in this document are structured so that the test development contractor and item writers have a single document to refer to when developing the assessment. Therefore, portions of the Framework that specify characteristics of the assessment or items are repeated in these Specifications.

12th Grade NAEP

In May 2005, the National Assessment Governing Board adopted a policy statement regarding NAEP and 12th grade preparedness. The policy states that NAEP will pursue assessment and reporting on 12th grade student achievement as it relates to preparedness for postsecondary pursuits, such as college-credit coursework, training for employment, and entrance into the military, as measured by an assessment consistent with that purpose. This policy resulted from recommendations of the Governing Board's National Commission on NAEP 12th Grade Assessment and Reporting in March 2004. Subsequent studies and deliberations by the Board took place during 2004 and 2005. In reading, the Board adopted minor modifications to the 2009 NAEP Reading Framework at grade 12, based on a comprehensive analysis of the Framework conducted by Achieve, Inc. for NAGB. The current version of the Reading Framework incorporates these modifications at grade 12 to enable NAEP to measure and report on preparedness for post-secondary endeavors.

Overview of the Specifications

This document is divided into nine chapters as briefly described below. A glossary of terms used in the Exhibits and throughout the 2009 NAEP Reading Specifications is provided in Appendix B.

Chapter 1: Introduction

This chapter contains information on NAEP administration and student samples and NAEP achievement levels. This chapter provides background on the NAEP Reading Assessment, including the history of the Framework, as well as a brief overview of the 2009 NAEP Reading Framework development process and a comparison of previous Frameworks to the 2009 Frameworks. This chapter also outlines the context of the *No Child Left Behind* legislation with regard to NAEP.

Chapter 2: Assessment Specifications

This chapter provides guidelines that apply to developing the assessment as a whole. It includes information about how the entire item pool should be structured so that it is aligned with the NAEP Reading Framework.

This chapter also provides an overview of text types on the assessment, presents guidelines for selecting passages, provides exhibits about the distribution of passages across grades and text types, and introduces the vocabulary assessment in terms of passage selection. Passage mapping is also discussed in this chapter and elaborated on in Appendix C.

Chapter 3: General Item Specifications by Text Type, Cognitive Target, and Grade Level

This chapter focuses on item specifications that apply to all the grade levels assessed. It includes full discussions of the characteristics of NAEP Reading Assessment items as well as descriptions of item formats and item-writing guidelines. It also introduces the literary and informational text matrices and provides a discussion of the cognitive target matrix, guided by exhibits.

Chapter 4: Inclusion of Students With Disabilities and English Language Learners

This chapter describes accommodations and other considerations for assessing students with disabilities and students who are English language learners. It includes item-writing considerations for these students which apply to all the items written for the assessment.

Chapter 5: Vocabulary Assessment on the 2009 NAEP Reading Assessment

This chapter introduces a systematic approach to vocabulary assessment on the 2009 NAEP Reading Assessment. It discusses the measurement of meaning vocabulary and includes criteria for selecting vocabulary to be assessed.

Chapter 6: Scoring the NAEP Reading Assessment

This chapter discusses item types and score categories and provides guidelines for the development of scoring rubrics. Alignment of items and scoring rubrics is also discussed.

Chapter 7: Reviews and Item Try-Outs

This chapter discusses the review and quality control procedures built into the assessment development process. It focuses on item reviews including sensitivity reviews, classroom tryouts, and cognitive labs.

Chapter 8: Sample Items

This chapter contains sample items illustrating the concepts and item-writing principles in Chapter 2 and the specifications for text types, cognitive targets, and grade levels described in Chapter 3.

Chapter 9: Special Studies

This chapter describes three recommended special studies that examine the assessment of vocabulary in context, achievement patterns of English language learners, and gender differences in reading achievement.

CHAPTER 1

INTRODUCTION

The National Assessment of Educational Progress (NAEP) has, since 1969, been an ongoing national indicator of what American students know and can do in major academic subjects, including reading in English. NAEP reading assessments have been administered on a regular schedule to students in grades 4, 8, and 12. Under the *No Child Left Behind Act of 2001,* NAEP will assess reading in grades 4 and 8 every two years. NAEP will also measure reading in grade 12 every four years.

This *Reading Assessment and Item Specifications for the 2009 National Assessment of Educational Progress* is one of two documents that describe the assessment; it is intended for a technical audience, including the National Center for Education Statistics and the contractor that will develop the 2009 NAEP Reading Assessment. The Specifications provide the "test blueprint," that is, information about passage selection, item development, and other aspects of test development.

The second document, the *Reading Framework for the 2009 National Assessment of Educational Progress*, presents the conceptual base and content of the assessment and is intended for a more general audience.

OVERVIEW OF NAEP

Since 1969, the National Assessment of Educational Progress (NAEP) has gathered information about student achievement in reading. Results of these periodic assessments are published to inform citizens about the nature of students' achievement in this subject, to inform curriculum specialists about the level and nature of student understanding, and to provide policymakers with information about factors related to schooling and their relationship to student achievement in reading. In 1988, Congress authorized the trial state assessment program in reading and mathematics (more commonly known as the State NAEP program), which is based on the same assessment instruments as the national NAEP. Results are reported about the students in each participating state. Data are also collected that allow the comparison of students' reading achievement over long periods of time, in a separate Long-Term Trend NAEP. These assessments—at the national level only—have been administered in the same form since 1971 and provide the only available measure of extended long-term trends in reading achievement.

The National Assessment Governing Board (NAGB)—the policy-making body for NAEP created by Congress in 1988—is specifically charged with developing assessment objectives and specifications through a national approach, identifying appropriate achievement goals for each age and grade, and carrying out other NAEP policy responsibilities. NAGB has defined several parameters for the NAEP Reading Assessment. First, the NAEP Reading Assessment will measure reading comprehension in English. On the assessment, students will be asked to read passages written in English and to answer questions about what they have read. Second, because this is an assessment of reading comprehension and not listening comprehension, NAEP will not allow passages to be read aloud to students as a test accommodation. Third, under NAGB policy, the Framework "shall not endorse or advocate a particular pedagogical approach…but shall focus on

important, measurable indicators of student achievement."[2] Although broad implications for instruction may be inferred from the assessment, NAEP does not specify how reading should be taught, nor does it prescribe a particular curricular approach to teaching reading.

Reading passages to be included on the assessment are selected to be interesting to students nationwide, to represent high-quality literary and informational material, and to be free from bias. Students respond to both multiple-choice and constructed-response items. In total, the NAEP assessments at grades 4, 8, and 12 are extensive enough to ensure that results can be reported validly, but no single student participates in the entire assessment. Instead, each student reads approximately two passages and responds to questions about what he or she has read.

NAEP assessments are administered to random samples of students designed to be representative of the nation, different regions of the country, participating states, and large urban districts. As discussed in Chapter 3, NAEP results are reported for groups of students; no data are reported for individual students. Since 1992, states have been able to obtain state-level data on students' reading achievement. In 2002 and 2003, large urban school districts were able to obtain data about their students' reading achievement. Results are reported in documents such as the *NAEP Reading Highlights* and the *NAEP Reading Report Cards* which are issued following each administration of the reading assessment; through special, focused reports; and through electronic means.

In 2002, NAGB awarded a contract to the American Institutes for Research (AIR) for the purpose of developing the Framework and Specifications to guide the 2009 National Assessment of Educational Progress in Reading.[3]

Three project committees were involved in making recommendations for the NAEP Reading Framework Development project:

- A Steering Committee, consisting of representatives of the business community, national policy organizations, reading associations, and reading educators

- A Planning Committee, consisting of reading educators in K–12 public and private education and in colleges and universities, business representatives, and members of education organizations

- A Technical Advisory Panel, consisting of measurement experts from states, research organizations, and universities

Members of the Steering and Planning Committees and the Technical Advisory Panel are listed in Appendix A. In addition to the three project committees, NAGB commissioned an External Review Panel, comprising prominent reading researchers and scholars, to examine the draft document and provide an independent review of the draft 2009 NAEP Reading Framework. Members of the External Review Panel are also listed in Appendix A.

[2]National Assessment Governing Board. (2002, May). *National Assessment Governing Board policy on framework development*. Washington, DC: Author.

[3]For more information about the development process and results, see the *Reading Framework for the 2009 National Assessment of Educational Progress*.

CONTEXT FOR DEVELOPING THE ASSESSMENT AND ITEM SPECIFICATIONS

The 2009 NAEP Reading Framework and accompanying Reading Specifications were developed in a time of intense interest in the improvement of reading achievement and keen awareness of the scientific literature about reading acquisition and growth. The *No Child Left Behind* (NCLB) legislation of 2001 reflects this context and has stated specific purposes for NAEP. The NAEP legislation, as amended under NCLB and the later *National Assessment of Educational Progress Reauthorization Act (NAEPRA) of 2002*, specifies that NAEP's purpose is "to provide, in a timely manner, a fair and accurate measurement of student academic achievement and reporting of trends in such achievement in reading, mathematics, and other subjects[s]."[4]

To comply with this legislation, the NAEP reading data will measure national, regional, state, district, and subgroup trends in reading achievement but will not target the performance of individual students or schools. In further accordance with NCLB, the NAEP Reading Assessment will be administered every two years at grades 4 and 8, and the resulting data will be widely reported in a timely fashion. Finally, NAEP specifies that although the public will have full access to NAEP results and released test questions, NAEP will not seek to influence the curriculum or assessments of any state. The 2009 NAEP Reading Framework is consistent with the NCLB legislation.

To develop the 2009 NAEP Reading Framework, AIR, under contract to NAGB, engaged in a comprehensive process that involved extensive review of the scientific research literature; consultation with three committees of national and state policymakers, state assessment staff and reading educators, and others who use the information from the NAEP Reading Assessment; and wide public review of successive drafts of the Framework.

The 2009 NAEP Reading Assessment will be developed to represent the content emphasis, complexity of reading, item format guidelines, and other requirements of the NAEP Reading Framework. Item writers for the assessment will be experts in reading and reading education. Under the direction of the test development contractor, they will use the Framework and these Specifications to guide their work.

NAEP ADMINISTRATION AND STUDENT SAMPLES

As currently planned, the 2009 NAEP Reading Assessment will assess students in grades 4, 8, and 12, using three types of items: multiple-choice, short constructed-response, and extended constructed-response. The assessment will be designed to have multiple test booklets. Because the items will be distributed across these booklets using a matrix sampling design, students taking part in the assessment will not all receive the same items. In addition to the reading items, the assessment booklets will include background questionnaires, administered in separately timed sessions.[5] Each student will spend approximately one hour taking the 2009 NAEP Reading Assessment.

The NAEP Reading Assessment measures reading achievement of students in the nation's schools in grades 4, 8, and 12 and reports the results at national, regional, and state levels. To

[4]National Assessment of Educational Progress Reauthorization Act, P.L. 107–279; (section 303(b)(1)).

[5]See *The Nation's Report Card: Reading 2000* and NAEP technical reports produced by NCES for more information about administration conditions and procedures.

implement these goals, schools throughout the country are randomly selected to participate in the assessment. The sampling process is carefully planned to select schools that accurately represent the broad population of U.S. students and the populations of students in each state participating in State NAEP and students from participating large urban districts. The sample includes schools of various types and sizes from a variety of community and geographical regions, with student populations that represent different levels of economic status; racial, ethnic and cultural backgrounds; and instructional experiences. Students with disabilities and English language learners are included to the extent possible, with accommodations as necessary (see Chapter 4 for more information about inclusion criteria and accommodations).

The sophisticated matrix sampling strategy helps ensure that the NAEP program can generalize the assessment findings to the diverse student populations in the nation and participating jurisdictions. This generalizability allows the program to present information on the strengths and weaknesses in aggregate student achievement in reading; provide comparative student data according to race/ethnicity, type of community, and geographic region; describe trends in student performance over time; and report relationships among student achievement and certain background variables.

REPORTING THE RESULTS OF THE NAEP READING ASSESSMENT

Results of the NAEP Reading Assessment administrations are reported in terms of average scores for groups of students on the NAEP 0–500 scale and as percentages of students who attain each of the three achievement levels (Basic, Proficient, and Advanced), defined in Exhibit 1. The NAEP Reading Assessment is an assessment of overall achievement, not a tool for diagnosing the needs of individuals or groups of students. Reported scores are always at the aggregate level. By law, scores are not produced for individual schools or students. Results are reported for the nation as a whole, for regions of the nation, for states, and for large districts that volunteer to participate in the NAEP Trial Urban District Assessment (TUDA).

No Child Left Behind Provisions for NAEP Reporting

Under the provisions of the *No Child Left Behind* legislation, states receiving Title I grants must include assurance in their state plans that they will participate in reading and mathematics State NAEP at grades 4 and 8. Local districts that receive Title I funds must agree that they will participate in biennial NAEP administrations at grades 4 and 8 if they are selected to do so. Their results will be included in state and national reporting. Participation in NAEP will not substitute for the mandated state-level assessments in reading and mathematics at grades 3 to 8.

In 2002, NAEP initiated a Trial Urban District Assessment (TUDA) in five large urban school districts that are members of the Council of Great City Schools (Atlanta City, City of Chicago, Houston Independent School District, Los Angeles Unified School District, and New York City Public Schools). Ten large districts participated in 2003, and 10 will take part in the 2005 TUDA. Large districts that participate in the urban district assessment in the future will receive their own data, which they can use for evaluating the achievement of their own students and for comparative purposes.

Achievement Levels

Since 1990, NAGB has used student achievement levels for reporting results on NAEP assessments. The achievement levels represent an informed judgment of "how good is good enough" in the various subjects that are assessed. Generic policy definitions for achievement at the Basic, Proficient, and Advanced levels describe in very general terms what students at each grade level should know and be able to do on the assessment. Preliminary reading achievement levels have been developed that are specific to the 2009 NAEP Reading Framework as a whole and to the vocabulary component of the assessment. The new reading-specific achievement level descriptors will replace those aligned to the previous framework.[6] These preliminary achievement levels will guide item development and initial stages of standard setting for the 2009 NAEP Reading Assessment; they may be refined as a result of the achievement-level-setting process.

Exhibit 1 presents the generic achievement level descriptors. Preliminary achievement levels for the 2009 NAEP Reading Assessment are presented in Chapter 2.

EXHIBIT 1
Generic Achievement Levels for the National Assessment of Educational Progress

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

Reporting NAEP Results

NAEP Reading Assessment results are reported in terms of average scores for groups of students on the NAEP 0–500 scale and as percentages of students who attain each of the three achievement levels; Basic, Proficient, and Advanced. Information is also provided about students who score below Basic on the assessment. These students are not necessarily nonreaders; many can complete some tasks on the assessment but are not able to attain the minimum score on the NAEP scale to be designated at the Basic level.

Data are reported on subgroups of students by gender, race/ethnicity, eligibility for free or reduced-price lunch, region of the country, type of community, public or nonpublic school, and other variables of interest. Data are never provided for individual schools or students. Subscores will be

[6]National Assessment Governing Board. (2003). *Reading framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

provided for literary and informational texts. Information will also be provided about students' responses to the vocabulary items.

It is recommended that the 2009 NAEP Reading Assessment results use a 0–500 cross-grade scale. Such a scale affirms that reading is a development process, in that students' reading skills mature throughout their school years as they read increasingly diverse and sophisticated texts.

The primary vehicles for reporting NAEP reading results are the *Reading Highlights* and *Reading Report Cards* that are issued after each assessment administration. These reports provide detailed information on the assessments, the students who participated, and the assessment results. Results are disaggregated by specific groups and are also presented for states that participate in the State NAEP. Among the focal groups are males and females, students from various racial/ethnic backgrounds, and students who took the assessment with and without accommodations.

NAEP data and information about the assessments are also available electronically through the NAGB (www.nagb.org) and the National Center for Education Statistics/NAEP (nces.ed.gov/nationsreportcard) Web sites. Further, the *NAEP Report Generator* tool can be used by interested education administrators, researchers, and other stakeholders to develop focused reports. The *NAEP e-Library* (nces.ed.gov) provides other information; access to NAEP reports, sample assessment passages, items, scoring rubrics with student-constructed responses; and data sources for more in-depth analysis of student achievement results or of the assessments themselves.

Reporting State NAEP Results

As discussed previously, states receiving Title I funding must participate in the NAEP Reading Assessment at grades 4 and 8. Results are reported in the aggregate for participating students and are also disaggregated for specific reference groups of students. Individual state reports are generated in addition to reports that contrast results from participating states and from the nation as a whole. The *NAEP Report Generator,* mentioned above, allows state and local administrators and others to customize reports and to investigate specific aspects of student reading achievement.

Reporting Trend Data

According to NAEP law and NAGB policy, long-term trend assessments are conducted as part of NAEP to continue the national trend reports, which, in reading, have been administered since 1971. The long-term trend reports provide the only continuous measures of student achievement over an extended period of time. Passages and accompanying test items administered as part of the long-term trend assessments have remained unchanged from their initial administration in 1971.

Because NAEP reports provide trend results over time, they are useful for informing decisions, allocating resources, and framing policy about reading. NAEP addresses the following questions:

- Are students improving in reading achievement over time?
- Are percentages of students at the upper achievement levels increasing, decreasing, or remaining the same?
- Are the gaps in achievement among various groups narrowing?

As discussed later in this chapter and illustrated in Exhibit 3, the 2009 NAEP Reading Framework represents several important changes from the Framework that has guided the assessment since 1992. These changes are significant enough that the reading trend line from the 1992 assessment will be broken; a new trend line will be instituted to reflect student achievement in reading throughout the use of the 2009 NAEP Reading Framework. Assessments aligned to the 1992 Framework and its subsequent versions will have yielded trend data from seven national and six state administrations, as shown in Exhibit 2.

EXHIBIT 2
Years of Administration of NAEP Reading Assessments Aligned to the 1992 Framework

Year	Grades for National Administration	Grades for State Administration
1992	4, 8, 12	4
1994	4, 8, 12	4
1998	4, 8, 12	4, 8
2000	4	
2003	4, 8, 12	4, 8
2005	4, 8, 12	4, 8
2007	4, 8	4, 8

Background Variables

Students participating in the NAEP assessments respond to background questionnaires that gather information on variables that contribute to an understanding of reading achievement nationwide. Teachers and school administrators also complete background questionnaires that collect relevant data. To the extent possible, information is also gathered from non-NAEP sources, such as state, district, or school records, to minimize the burden on those who are asked to complete the questionnaires.

As stated in NAGB policy, background data on students, teachers, and schools are needed to fulfill the statutory requirement that NAEP include information, whenever feasible, disaggregated by race or ethnicity, socioeconomic status, gender, disability, and limited English proficiency. Background information serves the additional purpose of enriching the reporting of NAEP results by examining factors related to academic achievement in the specific subjects that are assessed.

To satisfy the goal of enriching reports on student achievement in reading, background variables are selected to be of topical interest, to be timely, and to be directly related to academic achievement. The selection of variables about which questions will be developed may reflect current trends in the field, such as the use of technology in reading instruction or the extent to which students use the Internet as a reference tool. Questions are nonintrusive; free from bias; and secular, neutral, and nonideological. The questions do not elicit personal feelings, values, or attitudes.

COMPARISON OF THE 1992–2007 AND THE 2009 READING FRAMEWORK

The Framework for the 2009 NAEP Reading Assessment replaces a Framework that was first developed for the 1992 assessment. The previous Framework was refined during its use to reflect more clearly the goal of precisely measuring students' reading skills and strategies and was reissued for the 2003 NAEP Reading Assessment. The 2009 NAEP Reading Framework honors many aspects of the previous Framework but also introduces some changes that can lead to better measurement and more precise reporting of assessment results. Important changes featured in the 2009 NAEP Reading Framework follow:

- An assessment design based on current scientific reading research
- Consistency with the *No Child Left Behind* legislation
- Use of international reading assessments to inform the NAEP Framework
- A more focused measurement of vocabulary
- Measurement of reading behaviors (cognitive targets) in a more objective manner
- Distinction of cognitive targets relevant to literary and informational text
- Use of expert judgment, augmented by readability formulas, for passage selection
- Testing of poetry at grade 4, in addition to grades 8 and 12
- A special study of vocabulary to inform the development of the 2009 NAEP Reading Assessment

Key similarities and differences between the past and the 2009 NAEP Reading Frameworks are presented in Exhibit 3. Chapter 2 explains the proposed content and design of the 2009 NAEP Reading Assessment. The content and cognitive targets, as operationalized to reflect the definition of reading presented above, will yield passages and items that reflect the complex interaction of the reader, the text, and the context of the assessment.

EXHIBIT 3
Similarities and Differences:
1992–2007 and 2009 NAEP Reading Frameworks

<table>
<tr><th></th><th colspan="2">Previous Reading Framework</th><th colspan="3">2009 NAEP Reading Framework</th></tr>
<tr><td>CONTENT</td><td>Content of Assessment:
• Literary
• Informational
• Document</td><td>Contexts for Reading:
• For literary experience
• For information
• To perform a task</td><td>Literary Text
• Fiction
• Literary Nonfiction
• Poetry</td><td colspan="2">Informational Text
• Exposition
• Argumentation and Persuasive Text
• Procedural Text and Documents</td></tr>
<tr><td>COGNITIVE PROCESSES</td><td colspan="2">Stances/Aspects of Reading:
• Forming a general understanding
• Developing interpretation
• Making reader/text connections
• Examining content and structure</td><td colspan="3">Cognitive Targets, Distinguished by Text Type
Locate/Recall | Integrate/Interpret | Critique/Evaluate</td></tr>
<tr><td>VOCABULARY</td><td colspan="2">Vocabulary as a “target” of item development, with no information reported on students’ use of vocabulary knowledge in comprehending what students read</td><td colspan="3">Systematic approach to vocabulary assessment, with potential for a vocabulary subscore</td></tr>
<tr><td>POETRY</td><td colspan="2">Poetry included as stimulus material at grades 8 and 12</td><td colspan="3">Poetry included as stimulus material at all grades</td></tr>
<tr><td>PASSAGE SOURCE</td><td colspan="2">Use of intact, authentic stimulus material</td><td colspan="3">Use of authentic stimulus material, plus some flexibility in excerpting stimulus material</td></tr>
<tr><td>PASSAGE LENGTH</td><td colspan="2">Grade 4: 250–800
Grade 8: 400–1,000
Grade 12: 500–1,500</td><td colspan="3">Grade 4: 200–800
Grade 8: 400–1,000
Grade 12: 500–1,500</td></tr>
<tr><td>PASSAGE SELECTION</td><td colspan="2">Expert judgment as criterion for passage selection</td><td colspan="3">Expert judgment and use of at least two research-based readability formulas for passage selection</td></tr>
<tr><td>ITEM TYPE</td><td colspan="2">Multiple-choice and constructed-response items included at all grades</td><td colspan="3">Multiple-choice and constructed-response items included at all grades</td></tr>
</table>

CHAPTER 2

ASSESSMENT SPECIFICATIONS

This chapter provides guidelines for developing the assessment as a whole, including information about passage selection. It offers a brief overview of item types and the cognitive targets toward which items should be developed. Chapter 3 contains detailed information about developing assessment items. Chapter 4 describes accommodations and other considerations for assessing students with disabilities and students who are English language learners; Chapter 5 discusses the development of vocabulary items; and Chapter 6 provides information on scoring.

The 2009 NAEP Reading Assessment will be developed so that it is aligned with the content and skills defined by the 2009 NAEP Reading Framework. The assessment will include two distinct types of text at grades 4, 8, and 12—literary and informational. Doing so will allow the development of items that measure students' comprehension of the different kinds of text they encounter in their in-school and out-of-school reading experiences. Literary and informational text should be included as described in the text-type matrices in this chapter and in accompanying descriptions. The assessment will also include items that assess students' ability to apply their knowledge of vocabulary as an aid in comprehension.

THE DEFINITION OF READING FOR THE 2009 NAEP READING ASSESSMENT

The 2009 NAEP Reading Assessment is guided by a definition of reading that reflects scientific research, draws on multiple sources, and conceptualizes reading as a dynamic cognitive process. The definition applies to the assessment of reading achievement on NAEP and is not intended to be an inclusive definition of reading or of reading instruction. The definition for the 2009 NAEP Reading Assessment states:

Reading is an active and complex process that involves:

- understanding written text;
- developing and interpreting meaning; and
- using meaning as appropriate to type of text, purpose, and situation.

This definition of reading is derived from research on reading acquisition and growth and reflects the definitions that guide the development of two international assessments of literacy, the Progress in International Reading Literacy Study (PIRLS)[7] and the Programme for Student Assessment (PISA).[8]

[7] Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001, March). *Framework and specifications for PIRLS Assessment 2001.* Chestnut Hill, MA: Boston College, Lynch School of Education, PIRLS International Study Center.

[8] Organisation for Economic Co-operation and Development. (2000). *Measuring student knowledge and skill: The PISA 2000 assessment of reading, mathematical and scientific literacy.* Paris: Author.

Terms used in the definition are further explained as follows:

Understanding written text—Readers attend to ideas and content in a text by locating and recalling information and by making inferences needed for literal comprehension of the text. In doing so, readers draw on their fundamental skills for decoding printed words and accessing their vocabulary knowledge.

Developing and interpreting meaning—Readers use more complex inferencing skills to comprehend information implied by a text. They integrate the sense they have made of the text with their knowledge of other texts and of outside experiences. At times, they revise their sense of the text as they encounter additional information or ideas.

Using meaning—Readers draw on the ideas and information they have acquired from text to meet a particular purpose or situational need. The "use" of text may be as simple as knowing the time when a train will leave a particular station or may involve more complex behaviors such as analyzing how an author developed a character's motivation or evaluating the quality of evidence presented in an argument.

Text—As used in the assessment, the term reflects the breadth of components in typical reading materials. Thus, text on the assessment will include literary or informational passages and may contain noncontinuous print material such as charts. Texts selected for inclusion on the assessment represent practical, academic, and other contexts and are drawn from grade-appropriate sources spanning the content areas.

Purpose—Students' purpose for reading the passages presented on NAEP is determined by the assessment context; thus, the influence of purpose on readers' comprehension is somewhat limited.

Situation—The situation for reading often determines the way that readers prepare for and approach their task. They consider why they are reading (e.g., to study, to relax), how much they know about the topic, and other concerns that shape the time they will spend reading.

Other terms used in the Exhibits and throughout the 2009 NAEP Specifications are defined in the glossary in Appendix B. Terms are defined according to their use in this document.

ACCESSIBILITY

It is important that the NAEP Reading Assessment be accessible to as many students as possible. Accessibility in an educational assessment context refers to the degree to which the assessment provides all students in the targeted population the opportunity to demonstrate their achievement in relation to the construct of interest, in this case reading as defined by the NAEP Reading Framework. The NAEP Reading Assessment will measure reading comprehension in English. On the assessment, students will be asked to read passages and items written in English and to answer questions about what they have read. Because this is an assessment of reading comprehension and not listening comprehension, NAEP does not allow passages to be read aloud to students as a test accommodation.

The NAEP Reading Assessment is designed to measure the achievement of students across the nation. Therefore, it should allow all students to demonstrate their knowledge and skills, including those who have learned to read in a variety of ways, following different curricula and using different instructional materials; students who have varying degrees of reading competency; students with disabilities; and students who are English language learners. The question to ask in developing such an assessment is, what is a reasonable way to measure the same intended constructs for students who come to the assessment with different experiences, strengths, and challenges; who approach the constructs from different perspectives; and who have different ways of displaying their knowledge and skill?

The central requirement for the assessment is that the reading constructs that are assessed be the same for all students who take the test, regardless of their individual differences. To this end, the assessment should maintain the rigor of the reading expectations in the Framework, while providing the means for all tested students to demonstrate their levels of knowledge and skills.

Two methods that NAEP uses to design an accessible assessment program are (1) developing the standard assessment so that it is accessible to the widest number of students, and (2) providing accommodations for students with special needs.

TEXT TYPES FOR THE 2009 NAEP READING ASSESSMENT

As previously stated, the 2009 NAEP Reading Assessment will include two types of texts—literary and informational. These text types are explained in the following section and further discussed in Chapter 3. The distinction of text types is grounded in research on textual differences and is reflected in practice in the teaching of reading and of English literature.

Literary Texts

There are three categories of literary text: **fiction**, **literary nonfiction**, and **poetry**. Students in elementary and middle schools read many different examples of these texts for enrichment and enjoyment. These texts represent the developing conceptual understandings formed by students during this period. In higher grades, more complex literary structures are common.

For purposes of the 2009 NAEP Reading Assessment, these literary texts—especially fiction or literary nonfiction—may be either intact passages or passages excerpted from longer, more complex forms such as novels. Material that is excerpted from longer pieces will be carefully analyzed to ensure that it has the structural integrity and cohesion necessary to sustain item development.

The first category of literary text is **fiction**. Fiction is characterized by a setting or settings; a simple or complex plot consisting of one or more episodes, a problem to be solved or a conflict that requires characters to change; a solution; and a reaction that expresses the protagonist's feelings about attaining the goal.

The second category of literary text is **literary nonfiction**. Stylistically, literary nonfiction frequently blends narrative forms of writing with factual information for the dual purpose of informing and offering reading satisfaction. The reader must be able to distinguish increasingly

subtle weaving of factual material in the narrative and must be able to distinguish among bias, opinion, and fact.

The third category of literary text in the 2009 NAEP Reading Assessment is **poetry**. Poetry has distinctive forms, functions, structures, and textual features. It is possible that two poems may be used together in an intertextual item set to allow students to perform complex reading tasks such as comparing thematic treatment in the two poems or contrasting two poets' choices of literary devices.

Exhibit 4 presents examples of the kinds of literary text that are appropriate for inclusion on the 2009 NAEP Reading Assessment at grades 4, 8, and 12.

EXHIBIT 4
Stimulus Material: Literary

		Fiction	Literary Nonfiction	Poetry
LITERARY	GRADE 4	Adventure Stories Historical Fiction Contemporary Realistic Fiction Folktales Legends Fables Tall Tales Myths Fantasy	Personal Essay Autobiographical/Biographical Sketches Speech	Narrative Poem Free Verse Lyrical Poem Humorous Poem
	GRADE 8	Science Fiction Plus Increasingly Complex Application of Grade 4	Character Sketch Memoir Plus Increasingly Complex Application of Grade 4	Ode Song (including ballad) Epic Plus Increasingly Complex Application of Grade 4
	GRADE 12	Satire Parody Allegory Monologue Plus Increasingly Complex Application of Grades 4 and 8	Classical Essay Plus Increasingly Complex Application of Grades 4 and 8	Sonnet Elegy Plus Increasingly Complex Application of Grades 4 and 8

Informational Texts

There are also three categories of informational text: **exposition**; **argumentation** and **persuasive text**; and **procedural text** and **documents**. As they progress beyond the early grades, students read informational text with increasing frequency both in and out of school.[9] The primary goals of informational text for school-age readers are to communicate information and to advance learning.

[9]Broer, N.A., Aarnoutse, C.A.J., Kieviet, F.K., & Van Leeuwe, J.F.J. (2002). The effect of instructing the structural aspect of texts. *Educational Studies, 28*(3), 213–238.

The first category of informational text, **exposition**, presents information, provides explanations and definitions, and compares and contrasts. The complexity of the exposition that students read increases as they progress through school, as illustrated in Exhibit 5.

The second category of informational text includes **argumentation** and **persuasive text**. These texts pose an argument or attempt to persuade readers toward a particular viewpoint. Argumentation and persuasive text present information to support or prove a point, to express an opinion, and to try to convince readers that a specific viewpoint is correct or justifiable. Authors of argumentation and persuasive text often reveal their own biases through their prose.

The third category of informational text includes **procedural text** and **documents**. Procedural text is primarily prose structured to show specific steps toward accomplishing a goal, or it may combine both textual and graphic elements to communicate to the user. Documents, in contrast, use text sparingly, in a telescopic way that minimizes the continuous prose that readers must process to gain the information they need. Documents on the 2009 NAEP Reading Assessment must be embedded within continuous text at grades 4 and 8, but stand-alone documents may be used at grade 12.

Exhibit 5 presents examples of the kinds of informational text that are appropriate for inclusion on the 2009 NAEP Reading Assessment at grades 4, 8, and 12. Stand-alone documents will be included on the assessment *only* at grade 12.

EXHIBIT 5
Stimulus Material: Informational

		Exposition	Argumentation and Persuasive Text	Procedural Text and Documents
INFORMATIONAL	GRADE 4	Informational Trade Book Textbook News Article Feature Article Encyclopedia Entry Book Review	Informational Trade Book Journal Speech Simple Persuasive Essay	Embedded in Text • Directions • Map • Time Line • Graph • Table • Chart
	GRADE 8	Historical Document Essay (e.g., informational, persuasive, analytical) Research Report Plus Increasingly Complex Application of Grade 4	Letter to the Editor Argumentative Essay More Complex Persuasive Essay Editorial Plus Increasingly Complex Application of Grade 4	Embedded in Text • Recipe • Schedules Plus Increasingly Complex Application of Grade 4
	GRADE 12	Essay (e.g., political, social, historical, scientific, natural history) Literary Analysis Plus Increasingly Complex Application of Grades 4 and 8	Essay (e.g., political, social) Historical Account Position Paper (e.g., persuasive brochure, campaign literature, advertisements) Plus Increasingly Complex Application of Grades 4 and 8	Stand-Alone Material • Manual • Contract • Application • Product Support Material Plus Increasingly Complex Application of Grades 4 and 8

Passage Length

For several reasons, material on the assessment will range in length from fairly short to fairly long texts as shown in Exhibit 6. First, to gain the most valid information about students' reading, stimulus material should be as similar as possible to what students actually encounter in their in-school and out-of-school reading. Using passages of varying lengths helps accomplish this goal because students read text of different lengths in school and out of school. Second, longer material challenges students to use their strategic reading skills in ways that reflect the kinds of reading they do in nontest situations.[10] Third, short passages usually will not yield approximately 10 distinct items, the required minimum number for each NAEP item set. Longer passages, with clear structural patterns, can support the development of multiple, distinct, nontrivial items that cover the range of content included in the literary and informational text matrices presented in the next chapter. These items will also allow broad coverage of the cognitive targets, which are discussed in Chapter 3. Exhibit 6 presents the range of passage lengths by grade.

[10]Paris, S.G., Wasik, B.A., & Turner, C.J. (1991). The development of strategic readers. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *The handbook of reading research* (Vol. II, pp. 609–640). Mahwah, NJ: Erlbaum.

EXHIBIT 6
Passage Lengths for Grades 4, 8, and 12

Grade	Range of Passage Lengths (Number of Words)
4	200–800
8	400–1,000
12	500–1,500

It is expected that in some cases, two poems will be used together to assess students' ability to compare them in terms of their themes and stylistic features. Prose passages used in intertextual item sets will also be fairly short. Likewise, it is possible that two documents might be included as intertextual stimuli at grade 12. Again, details are provided in Chapter 3.

Selecting Literary and Informational Prose

Several methods of evaluating passages will be used to ensure that the best possible stimulus material is included on the 2009 NAEP Reading Assessment. Expert judgment will be the primary method for evaluating and selecting passages for inclusion on the assessment. Additional methods will be passage mapping and vocabulary mapping. Passages will be thoroughly reviewed for potential bias and sensitivity issues. At least two research-based readability formulas will also be used to gather additional information about passage difficulty.[11]

Stimulus material must be of the highest quality, and it must come from authentic sources such as those students would encounter in their in-school and out-of school reading.[12] Texts will reflect our literary heritage by including significant works from varied historical periods. Material must be coherent and allow items that assess domain-specific knowledge. Additionally, systematic efforts will be made to ensure that texts selected for inclusion on the 2009 NAEP Reading Assessment will be interesting to the widest number of students. Readers become more engaged in text and consequently comprehend a selection better when they find the material interesting.

Passages selected for inclusion on the assessment will be well written, interesting to read, and "considerate." That is, they will be easily comprehensible because they are well organized, have appropriate vocabulary, and, where needed, have useful supplemental explanatory features such as definitions of technical terms or topographical features. Ideas marked by topographical features such as italics, bold print, and signal words and phrases tend to be processed more easily and recalled longer than unmarked information. In selecting passages, attention will be paid to written clues

[11] Klare, G.R. (1984). Readability. In P.D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 681–744). Mahwah, NJ: Erlbaum; White, S., & Clement, J. (2001,August). *Assessing the Lexile framework: Results of a panel meeting,* Working Paper No. 2001-08**.** Washington, DC: National Center for Education Statistics.

[12] Baumann, J. (1986). Effect of rewritten textbook passes on middle-grade students' comprehension of main ideas: Making the inconsiderate considerate. *Journal of Reading Behavior, 18,* 1–22; Wade, S., Buxton, W., & Kelly, M. (1999). Using think-alouds to examine reader-text interest. *Reading Research Quarterly, 34*(2), 194–213; Wade, S., & Moje, E. (2000). The role of text in classroom learning. Classroom language and literacy learning. In M. Kamil, P. Mosenthan, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 609–627. Mahwah, NJ: Earlbaum; Wade, S., Schraw, G., Buxton W., & Hayes, M. (1993). Seduction of the strategic reader: Effects of interest on strategy and recall. *Reading Research Quarterly, 28*(2), 92–114.

within text that can help readers understand structure, guide the development of main ideas, and influence the recall of information. For example, readers tend to organize and remember the emphasized information better when authors lead them with signal words indicating main ideas (for example, *the most important point here*), with phrases indicating sequencing (such as the words *first, second, third*), and with statements cross-referencing disparate parts of text.[13]

Especially in the selection of informational text, the degree of content elaboration will be an important criterion for passage selection. Sufficient elaboration of new concepts is needed if students are to gain sufficient information to respond to questions. Tersely written informational text tends to be more difficult for students to comprehend than text written with more elaborated explanations. Whether text is tersely written or presents fully elaborated content is particularly important with topics that may be beyond the background knowledge of some students.

An inviting writing style can also enhance interest and thereby increase comprehension. Material may be interesting not because of *what* is said but because of *how* it is said. For example, writers can increase interest by using active rather than passive verbs, by including examples that make the writing less abstract, and by using vivid and unusual words. An inviting writing style also influences voice. Voice, the qualities that help a reader view text as communication between an author and a reader, can have a positive effect on recall.[14]

Passage mapping[15] procedures should continue to be used to identify appropriate passages for the 2009 NAEP Reading Assessments. Methods used in previous assessments have been expanded for the new assessment. Mapping procedures result in a graphic representation of a possible stimulus selection that clearly highlights the hierarchical structure and the interrelatedness of the components of the passages. Story mapping, for example, shows how the setting of a story is related to and contributes to the development of plot and theme. Concept mapping shows the structure of informational text, along with the concepts presented and the relational links among concepts. Organizing information hierarchically within a passage allows identifying the various levels of information within a text so that items can target the most important aspects of what students read. As NAEP begins to assess vocabulary in a systematic way, the passage mapping procedures will be modified to ensure that the words selected for item development are appropriate. These procedures are detailed in Appendix C.

Selecting Poetry

In selecting poetry for the 2009 NAEP Reading Assessment, careful attention must be paid both to the language components of possible poems. The language should be rich and may have rhythm, rhyme, unusual juxtapositions of sound, and appeal to the senses, with metaphor and imagery.[16] Words and phrases should be used with economy to support and amplify the meaning inherent in the text; the style should be distinguished by author's craft and project the poet's feelings about his or her topic or theme.

---

[13]Armbruster, B.B. (1984). The problem of "inconsiderate text." In Duffy, G.G., Roehler, I.R., & Mason, J. (Eds.), *Comprehension instruction* (pp. 202–217). New York: Longman.

[14]Beck, I., McKeown, M., & Worthy, J. (1995). Giving a text voice can improve students' understanding. *Reading Research Quarterly, 30,* 220–238.

[15]Wixson, K.K., & Peters, C.W. (1987). Comprehension assessment: Implementing an interactive view of reading. *Educational Psychologist, 22*, 333–356.

[16]Sloan, G.D. (1978). *The child as critic.* New York: Teachers College Press.

Poems selected for the assessment should also present a theme. Especially at grades 4 and 8, the theme should be implicitly presented in terms that are accessible to students. Themes of poems used on the grade 12 assessment may be more abstract. Poems should "speak to" students at their own level of understanding while also broadening their experience and stretching their imaginations.[17]

Selecting Noncontinuous Text and Documents

In addition to prose and poetry of continuous text, the 2009 NAEP Reading Assessment will include prose that is augmented by noncontinuous textual elements such as embedded tables or charts; it will also include stand-alone documents at grade 12. In selecting materials that contain these noncontinuous textual elements, it is important to analyze layout to ensure that embedded information is used appropriately, in a way that is well integrated into the prose text and is not gratuitously distracting.

Stand-alone documents must be rich with appropriate information about which questions can be developed. The number of categories of information presented graphically and the clarity of the layout of documents will be essential criteria for selecting documents to be included on the assessment. The vocabulary and concept load of multimedia elements and of documents will also be considered.

Summary of Text Selection Considerations

Exhibit 7 summarizes the considerations for selecting passages and documents for the 2009 NAEP Reading Assessment. The first two columns present considerations for literary and informational continuous text. The third column presents considerations that must be made in selecting noncontinuous text that is embedded within continuous text or documents that will be used as stand-alone stimulus material at grade 12. Certain considerations are considered essential for each kind of stimulus material and represent the fundamental characteristics that make a text or document appropriate for inclusion on NAEP. All potential stimulus material must also be grade-appropriate to ensure that students will be able to understand the concepts presented and have had familiarity with the stylistic features of the material. Finally, balance must be considered so that the assessment as a whole reflects the full range of print and noncontinuous text that students encounter in their in-school and out-of-school reading.

[17]Sloan (1978), *op. cit.*

EXHIBIT 7
Considerations for Selecting Stimulus Material for the 2009 NAEP Reading Assessment

Literary Text	Informational Text	Graphical Displays of Information
Essential Characteristics • Ability to engage readers • Well-written, rich text • Recognized literary merit • Theme/topic appropriateness by grade level	Essential Characteristics • Ability to engage readers • Well-written, considerate text • Coherence • Theme/topic appropriateness by grade level	Essential Characteristics • Coherence • Clarity • Relevance (when embedded)
Grade Appropriateness • Complexity of characters • Number of characters • Vocabulary • Sophistication in use of literary devices • Complexity of dialogue • Point of view • Complexity of theme • Multiple themes (major/minor) • Use of time (flashbacks, progressive/digressive) • Illustrations	Grade Appropriateness • Topic • Vocabulary • Concepts (number, familiarity, abstractness) • Curricular appropriateness at grade level • Integrity of structure • Types of adjunct aids • Explicitness of perspective • Style	Grade Appropriateness • Structural complexity • Topic • Vocabulary • Concepts (number, familiarity, abstractness) • Number of categories of information presented • Amount of information within categories
Balance • Reflective of our literary heritage • Style • Variety of sentence and vocabulary complexity • Appropriateness of mode (prose vs. poetry) • Traditional as well as contemporary • Representative of varied historical periods, cultures, socio-economic backgrounds, etc. • Genre	Balance • Varied content areas • Style • Genre • Variety of sentence and vocabulary complexity • Appropriateness of mode	Balance • Embedded documents balanced with stand-alone documents (at grade 12) • Format

The search for stimulus material for the 2009 NAEP Reading Assessment should also be guided by the percentages of text types by grade level presented in Exhibit 8. The change across the grade levels reflects changes in the kinds of reading that students do as they progress through school. Note that these figures represent the balance of text types by grade level on the operational assessment and that more passages must be located and used in preparing for the pre-operational pilot test. The third column in Exhibit 8 also suggests a range of passages that might include documents or other graphics. As noted, in grades 4 and 8, these must be embedded within procedural texts. At grade 12, they may be used as stand-alone stimuli.

EXHIBIT 8
Percentage of Passages by Text Type and Grade Level

Grade	Literary	Informational
4	(50%) 30% Fiction 10% Literary Nonfiction 10% Poetry	(50%) 40% Exposition 10% Argumentation/Persuasive* (2–4 embedded within Procedural texts)
8	(45%) 20% Fiction 15% Literary Nonfiction 10% Poetry	(55%) 30% Exposition 25% Argumentation/Persuasive (2–3 embedded within Procedural texts)
12	(30%) 20% Fiction 5% Literary Nonfiction 5% Poetry	(70%) 30% Exposition 30% Argumentation/Persuasive (2–3 embedded within Procedural texts) and/or 10% Stand-Alone Procedural

*Note: In 2009 argumentation and persuasive texts will not be included in the grade 4 assessment.

Cognitive Targets for the 2009 NAEP Reading Assessment

In developing the 2009 NAEP Reading Framework, careful attention was given to the kinds of thinking that students do as they read literary and informational texts and the kinds of thinking that are applicable to all texts. These kinds of thinking have been conceptualized as "cognitive targets" for the assessment, that is, the targets toward which item writing should be directed. Exhibit 16, in Chapter 3, presents the kinds of thinking that are generalizable to all kinds of text as well as the genre-specific cognitive targets toward which item writing should be directed.

Item Formats

The assessment will use multiple-choice, short constructed-response, and extended constructed-response items. Multiple-choice items are scored as right (1) or wrong (0). Short constructed-response items are scored according to scoring rubrics with two (0, 1) or three (0–2) categories, and extended constructed-response items may be scored according to scoring rubrics with up to four categories. Depending on grade level, 40–50% of students' testing time should be allotted to multiple-choice items, approximately 40–45% of testing time should be devoted to short constructed-response items, and approximately 10–15% of testing time should be devoted to extended constructed-response items. Chapters 3 and 6 contain guidelines for writing multiple-choice items and constructed-response items and scoring rubrics.

ACHIEVEMENT LEVELS FOR THE 2009 NAEP READING ASSESSMENT

As discussed in Chapter 1, NAEP results are reported in terms of three achievement levels: Basic, Proficient, and Advanced. The generic policy definitions of these levels of achievement are presented in Exhibit 1 in Chapter 1. In developing the 2009 NAEP Reading Framework, preliminary reading-specific achievement levels were created to guide item writing and to be used in initial stages

of standard setting for the 2009 NAEP Reading Assessment. The preliminary achievement-level descriptions will be refined as a result of the achievement-level-setting process.

The reading-specific achievement level descriptors consider both the difficulty of the texts that students read at different grades and also the quality of student performance on tasks that themselves have particular requirements. Text difficulty is influenced by several factors, some of which are measured by typical readability indices, primarily vocabulary familiarity and sentence complexity. Additionally, the explicitness of concepts and relations within a text affects difficulty, as when an informational text incorporates obvious topic sentences or a story lays out a character's motivation clearly. The tasks that students will be asked to perform in the 2009 NAEP Reading Assessment are encompassed by the cognitive targets, which are explained fully in Chapter 3.

Preliminary achievement-level descriptors for literary and informational texts were developed to reflect the inclusion of both kinds of text on the assessment. These are presented in Exhibit 9. Distinct achievement-level descriptors are provided for grades 4, 8, and 12.

EXHIBIT 9
Preliminary Achievement Levels for the 2009 NAEP Reading Assessment

GRADE 4		
Achievement Levels	Literary	Informational
ADVANCED	Grade 4 students at the *Advanced* level should be able to: • Interpret figurative language • Make complex inferences • Identify point of view • Evaluate character motivation • Describe thematic connections across literary texts	Grade 4 students at the *Advanced* level should be able to: • Make complex inferences • Evaluate the coherence of a text • Explain author's point of view • Compare ideas across texts • Identify evidence for or against an argument
PROFICIENT	Grade 4 students at the *Proficient* level should be able to: • Infer character motivation • Interpret mood or tone • Explain theme • Identify similarities across texts • Identify elements of author's craft	Grade 4 students at the *Proficient* level should be able to: • Identify implicitly stated author's purpose • Summarize major ideas • Find evidence in support of an argument • Distinguish between fact and opinion • Draw conclusions
BASIC	Grade 4 students at the *Basic* level should be able to: • Locate textually explicit information, such as plot, setting, and character • Make simple inferences • Identify supporting details • Describe character's motivation • Describe the problem • Identify mood • Identify simple causal relations	Grade 4 students at the *Basic* level should be able to: • Find the topic sentence or main idea • Identify supporting details • Identify author's explicitly stated purpose • Make simple inferences • Identify simple causal relations

EXHIBIT 9 (Continued)
Preliminary Achievement Levels for the 2009 NAEP Reading Assessment

GRADE 8		
Achievement Levels	Literary	Informational
ADVANCED	Grade 8 students at the *Advanced* level should be able to: • Make complex inferences • Critique point of view • Evaluate character motivation • Describe thematic connections across literary texts • Evaluate how an author uses literary devices to convey meaning	Grade 8 students at the *Advanced* level should be able to: • Make complex inferences • Evaluate author's purpose • Evaluate strength and quality of supporting evidence • Compare and contrast ideas across texts • Critique causal relations
PROFICIENT	Grade 8 students at the *Proficient* level should be able to: • Make inferences that describe problem and solution, cause and effect • Analyze character motivation • Interpret mood or tone • Explain theme • Identify similarities and differences across texts • Analyze how an author uses literary devices to convey meaning • Interpret figurative language	Grade 8 students at the *Proficient* level should be able to: • Summarize major ideas • Draw conclusions • Provide evidence in support of an argument • Describe author's purpose • Analyze and interpret implicit causal relations
BASIC	Grade 8 students at the *Basic* level should be able to: • Interpret textually explicit information • Make inferences • Identify supporting details • Identify character's motivation • Describe the problem • Identify mood	Grade 8 students at the *Basic* level should be able to: • Locate the main idea • Distinguish between fact and opinion • Make inferences • Identify explicitly stated author's purpose • Recognize explicit causal relations

EXHIBIT 9 (Continued)
Preliminary Achievement Levels for the 2009 NAEP Reading Assessment

GRADE 12		
Achievement Levels	Literary	Informational
ADVANCED	Grade 12 students at the *Advanced* level should be able to: • Make complex inferences • Critique point of view • Evaluate character motivation • Explain thematic connections across literary texts • Analyze and evaluate how an author uses literary devices to convey meaning	Grade 12 students at the *Advanced* level should be able to: • Evaluate the quality of supporting evidence • Critique point of view • Analyze causal relations • Critique the presentation of information • Evaluate the quality of counterarguments within and across texts
PROFICIENT	Grade 12 students at the *Proficient* level should be able to: • Examine relations among theme, setting, and character • Make inferences that describe problem and solution, cause and effect • Analyze character motivation • Interpret mood or tone • Integrate ideas to determine theme • Analyze how an author uses literary devices to convey meaning	Grade 12 students at the *Proficient* level should be able to: • Find evidence in support of an argument • Integrate information from a variety of sources • Determine unstated assumptions • Analyze the point of view • Judge the logic, coherence, or credibility of an argument
BASIC	Grade 12 students at the *Basic* level should be able to: • Interpret textually explicit information • Make inferences • Describe character's motivation • Recognize alternative interpretations or point of view • Explain the theme • Explain how the message is affected by the genre • Identify elements of author's style	Grade 12 students at the *Basic* level should be able to: • Summarize the main idea • Identify key details • Identify author's purpose • Identify causal relations • Draw conclusions

Each passage on the 2009 NAEP Reading Assessment will include at least two items that are specifically designed to measure students' vocabulary knowledge. For that reason, preliminary achievement-level descriptors were developed to reflect basic, proficient, and advanced levels of vocabulary knowledge. Several factors distinguish vocabulary understanding and use. These include the depth and breadth of individuals' vocabulary, that is, the number of words they identify and comprehend. Also important are the extent to which known words represent abstract ideas and concepts, and the numbers of words that do so. An additional critical factor is the flexibility with which individuals can extend the nuances of words to fit new contexts in their speech, writing, or reading. Chapter 5 provides details on writing vocabulary items.

Exhibit 10 presents the preliminary achievement-level descriptions for vocabulary. The descriptions are not presented by grade level but instead refer to reading achievement at basic, proficient, and advanced levels when students encounter grade-appropriate text. Students at grades 4, 8, and 12 will differ in the number of words they know and must apply their vocabulary skills to increasingly sophisticated texts at each grade.

EXHIBIT 10
Preliminary Achievement Levels: Vocabulary

Achievement Level	Description
Advanced	Advanced readers have outstanding vocabularies, with a sound knowledge of words and terms well beyond the expectations of reading material generally ascribed to their particular grade level. In addition, they have an excellent grasp of the multiple meanings of an extensive set of words and complex networks of associations to the words they know. They also have a strong base of words that identify complex and abstract ideas and concepts. Finally, their sophistication with words and word meanings enables them to be highly flexible in extending the senses of words they know to appropriately fit different contexts.
Proficient	Proficient readers have sizable meaning vocabularies, including knowledge of many words and terms above that of reading material generally ascribed to their grade level. They also have greater depth of knowledge of words, beyond the most common meaning. Proficient readers are flexible with word meanings and able to extend the senses of words whose meanings they know in order to appropriately fit different contexts and understand passage meaning.
Basic	Readers at the Basic level generally have limited vocabularies that consist primarily of concrete words at and below that of reading material generally ascribed to their grade level. Knowledge of these words is limited to the most familiar definition, making it difficult for these readers to identify the appropriate meaning of a word among the distractors.

CHAPTER 3

GENERAL ITEM SPECIFICATIONS BY TEXT TYPE, COGNITIVE TARGET, AND GRADE LEVEL

This chapter discusses specifications that apply to all grade levels assessed by the NAEP Reading Assessment. It begins with general information on writing high-quality items and then presents details on writing items that are aligned with the 2009 NAEP Reading Framework. NAGB's NAEP Item Development and Review Policy Statement also provides detailed information about the development and review of items. See Appendix D for the full policy statement. Additional item-writing considerations that apply to all the items written for the assessment are in Chapter 4. Chapter 5 presents information about writing items that assess students' meaning vocabulary. And Chapter 6 provides information on scoring multiple-choice and constructed-response items for the 2009 NAEP Reading Assessment.

PRINCIPLES OF GOOD ITEM WRITING

Principles of clear measurement intent and use of plain language should guide the development of items for the NAEP Reading Assessment.

Clear Measurement Intent

A critical step in good item writing is making sure that the measurement intent of the item is clear and that students understand what is being measured and what type of response is expected. Guidelines for ensuring clear measurement intent follow:

- What is being measured in each item should be clear to the students who take the test. Writers should be careful not to make assumptions about how students will interpret an item's implicit requirements.

- Item response requirements should not be designed to be dependent on one another. Passages can be related, as in the intertextual categories, but the items should be constructed as independent entities and not require correct answers in initial questions to correctly respond to subsequent items.

- Constructed-response items should contain clear directions to students about how they should respond. For example, if a correct response requires that students explain their thinking with references from the text, the item should state this requirement very clearly.

- Item writers should provide a clear description of what each item is intended to measure. This will help classify items according to assessment specifications, help develop clear scoring rubrics and scoring materials, reduce confusion in reviews, and provide evidence of the degree of alignment of the assessment to the framework.

Plain Language

Items should be written in plain language, that is, they should clearly convey what they are intended to measure. Plain language guidelines often increase access and minimize confusion for students. They can be summarized as follows:

- Write questions using brief simple sentences or stems, while also attending to the cohesion that is presented.
- When used, clauses should appear at the beginning of the sentence.
- Use present tense and active voice to the extent possible.
- Pronoun use should be limited; when used, pronouns should be close to the word to which they refer.
- Use high-frequency words as much as possible.
- Avoid colloquialisms or regionalisms.
- When using words with multiple meanings, ensure that the intended meaning is clear.
- Avoid using unnecessary descriptive information.
- Avoid unnecessary wordiness.

ITEM FORMATS

As stated in Chapter 2, the NAEP Reading Assessment has three types of items: multiple-choice, short constructed-response, and extended constructed-response items.

- **Multiple-choice items** require students to select one correct or best answer to a given question.

  — These items are scored as either correct (1) or incorrect (0).
  — Multiple-choice items should take approximately 1 minute to complete.

- **Short constructed-response items** require students to give a short answer, usually a phrase or a sentence or two.

  — Short constructed-response items are scored according to scoring rubrics with two or three categories.
  — Short constructed-response items should take approximately 1–2 minutes to complete.

- **Extended constructed-response items** require students to consider a situation that demands more than a short written response and provide a paragraph or two.

  — Extended constructed-response items are scored according to scoring rubrics with up to four categories.
  — Extended constructed-response items should take approximately 5 minutes to complete.

Item writers should carefully consider the knowledge and skills they intend to assess when deciding whether to write a multiple-choice or a constructed-response item. As detailed in this chapter, each cognitive target for the assessment focuses on specific knowledge and skills that can be measured using each of the three item formats. Constructed-response items must assess aspects of reading that can most appropriately be measured by having students actually write about their thinking.

Exhibit 11 shows the distribution of time among the item types by grade level. Less time is allocated to constructed-response items at grade 4 to reflect the developmental differences across the three grades that are assessed. Students at grade 4 may not be as familiar with written responses to reading questions as older students are.

EXHIBIT 11
Distribution of Time to Be Spent on Specific Item Type by Grade

Grade	Multiple Choice	Short Constructed Response	Extended Constructed Response
4	50%	40%	10%
8	40%	45%	15%
12	40%	45%	15%

The time students take to complete any item will depend on several factors in addition to format; for example, the length of the passage(s), the difficulty of the item, and the cognitive targets to be assessed by the item all contribute to the amount of time students need to respond. If the developer does not have data about how long it takes students to complete the items, informed judgment based on each item's characteristics should be used to assemble the pilot tests so that they will fit time specifications. Once items have been pilot-tested, information about the actual time needed to complete the items should be used in developing the tests. In initial item development, prior to the pilot test, approximately twice as many items should be written so that those with the best performance can be selected for use on the actual assessment.

Students will read and respond to two "item sets" or "blocks" consisting of a passage (or in the case of intertextual sets, two passages) and 10–12 items. Each block will contain one constructed-response item. Thus, the range in the number of items to be included on each block can be estimated as shown in Exhibit 12.

EXHIBIT 12
Number of Items by Item Type by Grade

Grade	Multiple Choice	Short Constructed Response	Extended Constructed Response	Total Number of Items per Block
4	4–5	4–5	1	10–12
8	3–5	5–8	1	10–12
12	3–5	5–8	1	10–12

Intertextual Item Sets

The 2009 NAEP Reading Assessment will contain intertextual item sets at all grade levels. These sets include two short passages that treat the same topic or theme in different ways. The purpose of these pairings is to prompt students to think across the two texts, that is, to compare and contrast ideas, themes, arguments, styles, or other aspects of text. Pairings may consist of passages within and/or across genres—that is, within the categories of either literary or informational text. The following are some examples of possible text pairings:

- Pro and con in argumentation and persuasive text
- Differing perspectives on the same topic
- A text passage paired with a commentary on that text
- A poem and a short story with a similar theme

Exhibit 13 presents the suggested distribution of intertextual blocks by grade level.

EXHIBIT 13
Percentage of Passages by Text Type and Grade*

Grade	Literary	Informational	Intertextual
4	(50%) 30% Literary Narrative 10% Literary Nonfiction 10% Poetry	(50%) 40% Exposition 10% Argumentation/Persuasive** (2–4 embedded within Procedural text)	1 of 10 blocks
8	(45%) 20% Literary Narrative 15% Literary Nonfiction 10% Poetry	(55%) 30% Exposition 25% Argumentation/Persuasive (2–3 embedded within Procedural text)	2 of 10 blocks
12	(30%) 20% Literary Narrative 5% Literary Nonfiction 5% Poetry	(70%) 30% Exposition 30% Argumentation/Persuasive (2–3 embedded within Procedural text) and/or 10% Stand Alone Procedural (2–3 embedded)	2–3 of 10 blocks

*Ten blocks is used for illustrative purposes only. The number of intertextual blocks should increase proportionally with the total number of blocks on the exam, if greater than 10 per grade. **In 2009 argumentation and persuasive texts will not be included in the grade 4 assessment due to difficulty in locating high quality texts appropriate for this grade.

DEVELOPING AND SCORING MULTIPLE-CHOICE ITEMS

Multiple-choice items are an efficient way to assess knowledge and skills, and they can be developed to measure each of the cognitive targets. In a well-designed multiple-choice item, the stem

clearly presents the question to the student. The stem may be in the form of a question, a phrase, or an expression, as long as it conveys what is expected of the student. The stem is followed by four answer choices, or options, only one of which is correct. Good multiple-choice items have the following characteristics:

- The stem includes only the information needed to make the student's task clear.
- Options are as short as possible and are parallel in length.
- Options are parallel in structure, syntax, and complexity.
- Options do not contain inadvertent cues to the correct answer, such as repeating a word from the stem in the correct answer or using specific determiners (e.g., *all, never*) in the distractors (incorrect options).
- Distractors are plausible, but not so plausible as to be possible correct answers.
- Distractors are designed to reflect the measurement intent of the item, not to trick students into choices that are not central to the idea being assessed.

Multiple-choice items are scored dichotomously; the rubric defines the following two categories:

1 = Correct
0 = Incorrect

DEVELOPING CONSTRUCTED-RESPONSE ITEMS AND SCORING RUBRICS

The type of constructed-response item, short or extended, that is written should depend on the construct that is being assessed. Item writers should draft the rubric that will guide scoring at the same time they are developing the item so that both the item and the rubric reflect the construct being measured.

All constructed-response items should communicate clearly to the student how the response to the item will be evaluated, for example whether they must justify their response with reference to the text. The corresponding scoring rubric should evaluate students' responses appropriately. For this reason, it is important to use passage maps in developing scoring rubrics (see Appendix C).

Short Constructed-Response Items

Some short constructed-response items are written to be scored dichotomously. Short constructed-response items with two scoring categories should measure knowledge and skills in a way that multiple-choice items cannot or provide greater evidence of the depth of students' understanding. They are also useful when there is more than one possible correct answer, when there are different ways to explain an answer, or when a brief justification is required. Item writers should take care that short constructed-response items would not be better or more efficiently structured as

multiple-choice items—there should be real value in having students actually constructing a response, rather than selecting the right answer from among wrong answers.

Other short constructed-response items are written to be scored on a three-category scale. Short constructed-response items with three scoring categories should measure knowledge and skills that require students to go beyond giving a simple acceptable answer that can obviously be scored right or wrong. Items scored with a 3-point rubric allow degrees of accuracy in a response so that a student can receive some credit for demonstrating partial understanding of a concept or skill.

Item writers must draft a scoring rubric for each short constructed-response item. For dichotomous items, the rubrics should define the following two categories:

1 = Correct
0 = Incorrect

For items with three score categories, the rubrics should define the following categories:

2 = Correct
1 = Partial
0 = Incorrect

Extended Constructed-Response Items

In general, extended constructed-response items ask students to think deeply about what they have read, to integrate concepts, to analyze a situation, or to explain a concept. Extended constructed-response items may have up to four scoring categories:

3 = Extensive
2 = Essential
1 = Partial
0 = Incorrect

In developing the scoring rubric for an extended constructed-response item, writers should think about the kind of student responses that would show increasing degrees of knowledge and understanding. Writers should sketch condensed sample responses for each score category.

Item writers must develop a draft scoring rubric specific to each extended constructed-response item. The rubric should clearly reflect the measurement intent of the item. Item writers also should include a justification or explanation for each rubric category description. Doing so will allow the writer to document the scoring rubric, as well as provide guidance for scoring the item.

Chapter 6 describes guidelines and requirements for developing scoring rubrics for all item types in greater detail.

DEVELOPING ITEMS TO ASSESS COMPREHENSION OF LITERARY AND INFORMATIONAL TEXTS

In addition to varying by format (multiple-choice or constructed-response), each item written for the NAEP Reading Assessment reflects two major dimensions: text type and cognitive target. Exhibits 4 and 5 in Chapter 2 suggest the kinds of literary and informational texts that are appropriate for the 2009 NAEP Reading Assessment and briefly discuss their characteristics. This chapter provides more information on the different categories of literary and informational text that are to be included on the assessment. The matrices in Exhibits 14 and 15 provide a reference for item writers. More specifically, they show the following aspects of literary and informational text at grades 4, 8, and 12, about which items can be written:

- Genres and types of text to be assessed
- Text structures and features about which items may be asked
- Aspects of author's craft about which items may be asked

These terms are defined as follows:

- **Genre and text type** refer to the kinds of literary and informational text and documents that can be included on the assessment. These are the idealized norms of a genre,[18] not the source of the stimulus material per se.

- **Text structures** refer to the organization of text, the ways ideas are arranged and are connected to one another.

- **Text features** refer to visual and structural elements that support and enhance the reader's ability to understand the text.

- **Author's craft** pertains to the specific techniques that an author chooses to relay an intended message.

Items will assess students' application of knowledge about text types, text features and structures, and author's craft, not their recognition of specific terminology in isolation. The designation of entries in the matrices by grade level reflects the levels at which these components of text are presented in state English language arts standards. The matrices have further been confirmed by experienced teachers and teacher educators.

LITERARY TEXT

The Literary Text matrix consists of three sections, one for each category of text: **fiction**, **literary nonfiction**, and **poetry**. Each category of literary text is explained below.

[18]Fludernik, M. (2000). Genres, text types, or discourse modes? Narrative modalities and generic categorization. *Style, 34*(2), 274–292.

Fiction

The first category of literary text is **fiction**. Students in elementary and middle schools read many different kinds of stories for enrichment and enjoyment. These texts are representative of the developing conceptual understandings formed by students during this period. At grades 8 and 12, more complex structures are common, including satires, parodies, science fiction, allegories, and monologues. For purposes of the NAEP Reading Assessment, these complex fictional texts may be either intact passages or passages excerpted from longer, more complex narrative forms such as novels. Material that is excerpted from longer pieces will be carefully analyzed to ensure that it has the structural integrity and cohesion necessary to sustain item development.

The matrix in Exhibit 14 shows the aspects of text structures and features and author's craft that may be assessed. These components, as well as the purposes for reading, become increasingly complex and sophisticated in the texts that students read as they move through the elementary, middle, and high school grades. For example, themes may be more abstract; conflicts may be internal as well as external; characterization may develop with antagonists, protagonists, and narrators with motives, beliefs, traits, and attitudes that are intertwined; the theme and setting may be more integral to each other; and the plot may consist of a series of rising and falling actions within episodes. Additionally, point of view, a complex component of narrative, becomes a component of the text structure. Generally, the point of view is not explicit; rather, the reader infers the point of view through subtle clues within the text. In material appropriate for grade 12 readers, theme and point of view are more complex, often including interior monologues, unreliable narrators, and multiple points of view.

Authors select from a range of stylistic devices to enhance their presentation of fictional text. In the matrix, these are referred to as author's craft. At grade 4, author's craft includes symbolism, simile, metaphor, and diction and word choice, such as dialect and exaggeration. More abstract elements are part of author's craft at grade 8 such as flashback and imagery, in addition to more complex applications of author's craft listed for grade 4. The fictional passages for grade 12 are complex and include the following literary devices: dramatic irony, character foils, comic relief, and unconventional use of language, in addition to, the devices under author's craft at grades 4 and 8.

EXHIBIT 14
Literary Text Matrix: Fiction

		Genre/Type of Text	Text Structures and Features	Author's Craft
FICTION	GRADE 4	Adventure Stories Historical Fiction Contemporary Realistic Fiction Folktales Legends Fables Tall Tales Myths Fantasy	Themes Morals Lessons Organization • Plot—Sequence of Events • Conflict • Solution • Resolution Elements • Setting • Characterization	Diction and Word Choice • Dialogue • Exaggeration • Figurative Language —Symbolism —Simile and Metaphor
	GRADE 8	Science Fiction Plus Increasingly Complex Application of Grade 4	Organization • Parallel Plots • Circular Plots Elements • Point of View • Contradictions • Internal vs. External Conflict Plus Increasingly Complex Application of Grade 4	Diction and Word Choice • Mood • Imagery • Flashback • Foreshadowing • Personification Plus Increasingly Complex Application of Grade 4
	GRADE 12	Satire Parody Allegory Monologue Plus Increasingly Complex Application of Grades 4 and 8	Organization • Differentiation of Plot Structures for Different Purposes and Audiences Elements • Interior Monologue • Unreliable Narrators • Multiple Points of View Plus Increasingly Complex Application of Grades 4 and 8	Diction and Word Choice • Dramatic Irony • Character Foils • Comic Relief • Unconventional Use of Language Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft.

Literary Nonfiction

The second category of literary text is **literary nonfiction**; it may include elements of narration and exposition and is often referred to as "mixed text."[19] Literary nonfiction is an example of mixed text because it uses literary techniques usually associated with stories, but also presents information or factual material. Stylistically, it frequently blends narrative forms of writing and factual information with the dual purpose of informing and offering reading satisfaction. The reader must be able to distinguish increasingly subtle weaving of factual material in the narrative and must be able to distinguish among bias, opinions, and facts. The text types for literary nonfiction at grade 4 include autobiographical and biographical sketches, personal essays, and speeches. At grade 8, additional forms of literary nonfiction are character sketches and memoirs. Classical essays are introduced as literary nonfiction at grade 12. Autobiographical and biographical works are also classified as literary nonfiction. Unlike texts that can be categorized as informational because of their sequential, chronological, or causal structure, literary nonfiction uses a storylike or narrative structure. Often organized around a thesis, literary nonfiction may interweave personal examples and ideas with factual information to attain the purpose of explaining, presenting a perspective, or describing a situation or an event. This organization around a thesis invites the kinds of questions that will be developed for fictional texts included on the 2009 NAEP Reading Assessment.

Although ostensibly a hybrid genre, the literary nonfiction selected for inclusion on NAEP will conform to the highest standards of literary quality. The structural elements listed in the matrix for literary nonfiction combine structures from both narrative and informational texts. Literary nonfiction is multidimensional and contains an interplay of text characteristics, which signals the complexity of this genre. At grade 4, text structures and features in literary nonfiction include description, cause and effect, comparison, chronology, point of view, themes or central ideas, and supporting ideas. Text features such as logical connective devices and transitional devices are listed in the matrix at grade 4. Increasingly complex examples of these structures listed above are found in literary nonfiction at grades 8 and 12.

A range of literary devices and techniques termed author's craft are present in literary nonfiction. Examples of author's craft at grade 4 include diction and word choice, various ways to introduce characters, exaggeration, and figurative language. At grade 8, increasingly complex techniques are listed for author's craft, such as voice, tone, imagery, and metaphoric language. Denotation, connotation, and irony are listed at grade 12 for author's craft. Grades 8 and 12 will include more complex forms of the text features, text structure, and author's craft listed at grade 4.

[19]Alexander, P.A., & Jetton, T.J. (2000). Learning from text: A multidimensional and developmental perspective. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 285–310). Mahwah, NJ: Erlbaum.

EXHIBIT 14 (Continued)
Literary Text Matrix: Literary Nonfiction

		Genre/Type of Text	Text Structures and Features	Author's Craft
LITERARY NONFICTION	GRADE 4	Personal Essay Speech Autobiographical and Biographical Sketches	Organization • Description • Cause and Effect • Comparison • Chronology Elements • Point of View • Themes or Central Ideas • Supporting Ideas • Logical Connections • Transitions	Diction and Word Choice • Use of Exposition, Action, or Dialogue to Introduce Characters • Exaggeration • Figurative Language — Symbolism — Simile and Metaphor
	GRADE 8	Character Sketch Memoir Plus Increasingly Complex Application of Grade 4	Plus Increasingly Complex Application of Grade 4	Diction and Word Choice • Voice • Tone • Imagery • Metaphoric Language Plus Increasingly Complex Application of Grade 4
	GRADE 12	Classical Essay Plus Increasingly Complex Application of Grades 4 and 8	Plus Increasingly Complex Application of Grade 4	Diction and Word Choice • Denotation • Connotation • Irony Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft.

Poetry

The third category of literary text included in the 2009 NAEP Reading Assessment is **poetry**. Like narratives, poetry has distinctive forms, functions, and structures that are further guided by literary structures and textual features. The matrix lays out the kinds of poetry that students encounter at different grade levels. Thus, basic poetic forms at grade 4 are narrative, lyrical, and humorous poems and free verse. Additionally at grade 8, odes, songs, and epics are included in the matrix for possible item development. More complex poetic forms are included at grade 12, such as sonnets and elegies. It is possible that two poems may be used together in intertextual item sets to allow students to perform complex reading tasks, such as comparing thematic treatment in the two poems or contrasting two poets' choices of literary devices.

Readers use the structure of poetry to aid in comprehension. Poetic structures range from simple to complex. Students at grade 4 can be expected to be familiar with simple organizational patterns such as verse and stanza, along with the basic elements of rhyme scheme, rhythm, mood, and theme and intent. At grades 8 and 12, increasingly complex poetic organizational patterns and elements will be included for assessment.

Understanding a poet's choices also aids in understanding poetry. Language choice is of particular importance because the message in poetry is distilled to as few words as possible. Poets choose from among a range of rhetorical structures and figurative language, using, for example, repetition, dialogue, line organization and shape, patterns, and many forms of figurative language. Increasingly complex application of figurative language, rhetorical devices, and complex poetry arrangements are included at grades 8 and 12.

EXHIBIT 14 (Continued)
Literary Text Matrix: Poetry

		Genre/Type of Text	Text Structures and Features	Author's Craft
POETRY	GRADE 4	Narrative Poem Lyrical Poem Humorous Poem Free Verse	Organization • Verse • Stanza Text Features • Repetition • Omission • Dialogue • Line Organization • Patterns Elements • Rhyme Scheme • Rhythm • Mood • Theme and Intent	Diction and Word Choice (including the decision to omit words which may leave the reader with much to infer) • Exaggeration • Use of Imagery to Provide Detail • Figurative Language —Simile —Metaphor —Imagery —Alliteration —Onomatopoeia Choice of Different Forms of Poetry to Accomplish Different Purposes
	GRADE 8	Ode Song (including ballad) Epic Plus Increasingly Complex Application of Grade 4	Elements • Abstract Theme • Rhythm Pattern • Point of View Plus Increasingly Complex Application of Grade 4	Diction and Word Choice • Figurative Language —Symbolism —Personification Plus Increasingly Complex Application of Grade 4
	GRADE 12	Sonnet Elegy Plus Increasingly Complex Application of Grades 4 and 8	Elements • Complex Themes • Multiple Points of View • Interior Monologue • Soliloquy • Iambic Pentameter Plus Increasingly Complex Application of Grades 4 and 8	Diction and Word Choice • Denotation • Connotation • Irony • Tone • Complex Symbolism • Extended Metaphor and Analogy Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft.

INFORMATIONAL TEXTS

The Informational Text matrix also consists of three sections, one for each category of informational text: **exposition**, **argumentation** and **persuasive text**, and **procedural text** and **documents**. The matrices included in Exhibit 15 provides information on the text structures and features and elements of author's craft that will be covered on the assessment.

Exposition

The first category of informational text is **exposition**. As they progress beyond the early grades, students read expository text with increasing frequency both in and out of school.[20] The primary goals of expository text for school-age readers are to communicate information and to advance learning. Forms that may be assessed at grade 4 are informational trade books, textbook passages, news stories, feature stories, encyclopedia entries, and book reviews. At grade 8, expository text genres include historical documents, various grade-appropriate essays, and research reports. More complex essay formats will be introduced to the assessment at grade 12; these include political, social, historical, or scientific essays that seek to communicate information, along with essays that present literary analysis.

Expository texts are characterized by internal structural patterns that are designed to move the exposition forward and to help the reader comprehend the text. As shown in the matrix in Exhibit 15, the major organizational structures of exposition are description, sequence, cause and effect, problem and solution, and comparison and contrast.[21] Exposition may also include lists as a structural component, with lists of descriptions, causes, problems, solutions, and views presented within the text. Commonly, exposition does not contain just one structural format, but rather combines several structures embedded in the text.

Specific elements within these organization structures signal meaning to the reader. Sequence, point of view, topics or central ideas, and supporting ideas and evidence are listed at grade 4; at grades 8 and 12, the structural organization and elements will be assessed at increasingly complex levels and with increasingly sophisticated texts. Some surface-level features support the text structures of exposition and guide the reader through the text. Other textual features can be categorized as reflecting author's craft; these features guide the reader through the use of transitional words, signal words, voice, figurative language, and rhetorical structures. At grades 8 and 12, increasingly complex use of these features and of the author's craft is included for assessment.

[20]Broer, N.A., Aarnoutse, C.A.J., Kieviet, F.K., & Van Leeuwe, J.F.J. (2002). The effect of instructing the structural aspect of texts. *Educational Studies, 28*(3), 213–238.

[21]Meyer, B.J.F. (2003). Text coherence and readability. Topics in *Language Disorders, 23*, 204–224.

EXHIBIT 15
Informational Text Matrix: Exposition

		Genre/Type of Text	Text Structures and Features	Author's Craft
EXPOSITION	GRADE 4	Informational Trade Book Textbook News Article Feature Article Encyclopedia Entry Book Review	Organization • Description • Sequence (e.g., enumeration, chronology) • Cause and Effect • Problem and Solution • Comparison and Contrast Content Features • Point of View • Topics or Central Ideas • Supporting Ideas and Evidence Graphic Features • Titles • Subheadings • Italics • Captions • Sidebars • Photos and Illustrations • Charts and Tables	Transitional Words Signal Words Voice Figurative Language and Rhetorical Structures • Parallel Structure • Quotations • Examples • Repetition • Logical Arguments
	GRADE 8	Historical Document Essay (e.g., informational, persuasive, analytical) Research Report Plus Increasingly Complex Application of Grade 4	Plus Increasingly Complex Application of Grade 4	Irony Sarcasm Plus Increasingly Complex Application of Grade 4
	GRADE 12	Essay (e.g., political, social, historical, scientific, natural science) Literary Analysis Plus Increasingly Complex Application of Grades 4 and 8	Plus Increasingly Complex Application of Grade 4 and 8	Denotation Connotation Complex Symbolism Extended Metaphor and Analogy Paradox Contradictions/Incongruities Ambiguity Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft.

Argumentation and Persuasive Text

The second category of informational text is **argumentation** and **persuasive text**. Many forms of informational text pose an argument or attempt to persuade readers toward a particular viewpoint. These texts present information to support or prove a point, to express an opinion, and to try to convince readers that a specific viewpoint is correct or justifiable. Various logical fallacies and forms of bias may be found in argumentation and persuasive text. As the matrix shows, there is considerable similarity in structure and literary features and elements among exposition, argumentation, and persuasive text. However, the real distinction lies in the purpose for which an author writes these particular kinds of informational text; as stated, exposition seeks to inform and educate, whereas argumentation and persuasive texts seek to influence their readers' thinking in other, often subtle but significant ways.

At grade 4, argumentation and persuasive texts listed in the matrix are informational trade books that specifically argue a position or persuade the reader toward a stance, journals, speeches, and simple persuasive essays. [Note: In 2009 NAEP will not assess argumentation and persuasive texts at grade 4 due to difficulty in locating high quality texts appropriate for this grade level.] At grade 8, there are more complex forms of argumentation and persuasive texts, such as letters to the editor and editorials, argumentative and grade-appropriate persuasive essays, and editorials. At grade 12, argumentation and persuasive texts become increasingly more complex with a variety of types of essays, such as political and social commentary essays; historical accounts; and position papers, such as persuasive brochures, campaign literature, and advertisements.

Particular organization techniques and elements are used to create a clear argument or to form a persuasive stand. The differences between exposition and argumentation and persuasive text lie not in the structural organization, but in the way the texts are elaborated through the use of contrasting viewpoints, shaping of arguments, appeals to emotions, and other manipulations of the elements of text and language. The organizational structures at all levels are the same as in exposition: description, sequence, cause and effect, problem and solution, and compare and contrast; they are represented in grades 8 and 12 with increasing complexity.

Elements within these organizational structures include the author's craft; topics or central ideas; supporting ideas; contrasting viewpoints or perspectives; and the presentation of the argument (e.g., issue definition, issue choice, stance, and relevance). These elements appear at all grade levels, with increasing complexity at higher grade levels. In addition, at grade 12 students may be asked about the structure of a given argument; connections among evidence, inferences, and claims; and the structure of a deductive vs. inductive argument. Twelfth grade students may also be asked questions about the range and quality of evidence; and logical fallacies, false assumptions/ premises, loaded terms, caricature, leading questions, and faulty reasoning in argumentation and persuasive texts.

EXHIBIT 15 (Continued)
Informational Text Matrix: Argumentation and Persuasive Text

		Genre/Type of Text	Text Structures and Features	Author's Craft
ARGUMENTATION AND PERSUASIVE TEXT	GRADE 4	Informational Trade Book Journal Speech Simple Persuasive Essay	Organization • Description • Sequence (e.g., enumeration, chronology) • Cause and Effect • Problem and Solution • Comparison and Contrast Content Features • Author's Perspective or Position • Topics or Central Ideas • Supporting Ideas and Evidence • Contrasting Viewpoints and Perspectives • Presentation of the Argument (e.g., issue definition, issue choice, stance, relevance) Graphic Features • Titles • Subheadings • Italics • Captions • Sidebars • Photos and Illustrations • Charts and Tables	Transitional Words Signal Words Voice Figurative Language and Rhetorical Structure • Parallel Structure • Quotations • Examples • Repetition • Exaggeration • Emotional Appeal • Tone • Logical Fallacies
	GRADE 8	Letter to the Editor Argumentative Essay More Complex Persuasive Essay Editorial Plus Increasingly Complex Application of Grade 4	Plus Increasingly Complex Application of Grade 4	Irony Sarcasm Plus Increasingly Complex Application of Grade 4
	GRADE 12	Essay (e.g., political, social) Historical Account Position Paper (e.g., persuasive brochure, campaign literature, advertisement) Plus Increasingly Complex Application of Grades 4 and 8	Plus Increasingly Complex Application of Grade 4 and 8	Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft. Note: In 2009 NAEP will not assess argumentation and persuasive texts at grade 4 due to difficulty in locating high quality texts appropriate for this grade level.

Procedural Texts and Documents

The third category of informational text consists of **procedural texts** and **documents**. Research indicates that adults spend considerably more time reading documents (i.e., information in matrix or graphic form) than they do reading prose materials.[22] Documents and procedural texts are indeed common in our society; for example, we interpret bus schedules, assemble simple devices, order goods from a catalog, or follow directions to set the VCR clock. Such texts are used frequently in elementary and secondary schools, where students encounter textbooks that are replete with graphs, tables, and illustrations to accompany and expand traditional continuous text.

Procedural text may be primarily prose, arranged to show specific steps toward accomplishing a goal, or may combine both textual and graphic elements to communicate to the reader. Documents, in contrast, use text sparingly and in a telescopic way that minimizes the continuous prose that readers must process to gain the information they need.

As the matrix shows, document texts on the 2009 NAEP Reading Assessment may include, but are not limited to, tables and charts. Stand-alone procedural text or documents will not be included at grades 4 and 8; such text will be embedded in or ancillary to continuous text. They may appear as stand-alone stimuli at grade 12, but their use will account for only a small amount of the stimuli in the entire assessment. It is likely that many of the documents may be used as part of intertextual item sets. For example, a student might encounter a bar graph and a timeline with items that relate to both texts.

Documents and procedural text features act as necessary clues to the organization of the text. As textual supports, these features guide the reader through the text. For the purposes of the 2009 NAEP Reading Assessment, textual features include titles, labels, headings, subheadings, sidebars, photos and illustrations, charts and graphs, and legends at grades 4, 8, and 12. More complex examples of these will be included at each successive grade.

[22]Guthrie, J.T., & Mosenthal, P. (1987). Literacy as multidimensional: Learning information on reading comprehension. *Educational Psychologist, 22,* 279–297; Kirsch, I.S., & Mosenthal, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25,* 5–30; Mosenthal, P.B. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Education Psychology, 88,* 314–332; Mosenthal, P.B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Education Research Journal, 35,* 269–307.

EXHIBIT 15 (Continued)
Informational Text Matrix: Procedural Texts and Documents

		Genre/Type of Text	Text Structures and Text Features
PROCEDURAL TEXTS AND /DOCUMENTS	GRADE 4	Embedded in Text • Directions • Map • Time Line • Graph • Table • Chart	Organization • Description • Procedures • Sequence (e.g., enumeration, chronology) Graphic Features • Titles • Labels • Headings • Subheadings • Sidebars • Photos and Illustrations • Charts and Graphs • Legends
	GRADE 8	Embedded in Text • Recipe • Schedule Plus Increasingly Complex Application of Grade 4	Plus Increasingly Complex Application of Grade 4
	GRADE 12	Stand-Alone Material • Application • Manual • Product Support Material • Contract Plus Increasingly Complex Application of Grades 4 and 8	Plus Increasingly Complex Application of Grades 4 and 8

The entries listed within each cell of the matrices should be construed as neither definitive nor inclusive of all text structures and features or techniques of author's craft.

COGNITIVE TARGETS FOR THE 2009 NAEP READING ASSESSMENT

As discussed briefly in Chapter 2, careful attention has been given to the kinds of thinking that students do as they read literary and informational texts and the kinds of thinking that are applicable to all texts. The term *cognitive targets* refers to the mental processes or kinds of thinking that underlie reading comprehension; the targets serve to guide development of test items. There are three categories of cognitive targets:

- Locate and recall information from text
- Integrate and interpret information and ideas presented in text
- Critique and evaluate information and ideas in text and the ways in which authors present text

The matrix in Exhibit 16 shows both the cognitive targets that are generalizable to all kinds of text and also the genre-specific cognitive targets for literary and informational text. These distinctions recognize that readers often adopt their strategies and skills to the demands of different kinds of text. The matrix will serve as a reference for item writers as they write items that measure the ways students at all grades think about the text structures and features of literary and informational text and the aspects of author's craft evident in the passages they read on the assessment. Depending on the kind of text, item writers can write items aligned to the cognitive targets that are applicable to both literary and informational text or to the appropriate genre-specific targets. The cognitive targets are explained in the text following Exhibit 16.

EXHIBIT 16
Sample Cognitive Targets for Literary and Informational Texts

	Locate/Recall	Integrate/Interpret	Critique/Evaluate
Both Literary and Informational Text	Identify textually explicit information and make simple inferences within and across texts, such as: • Definitions • Facts • Supporting details	Make complex inferences within and across texts to: • Describe problem and solution, cause and effect • Compare or connect ideas, problems, or situations s • Determine unstated assumptions in an argument • Describe how an author uses literary devices and text features	Consider text(s) critically to: • Judge author's craft and technique • Evaluate the author's perspective or point of view within or across texts • Take different perspectives in relation to a text
Specific to Literary Text	Identify textually explicit information within and across texts, such as: • Character traits • Sequence of events or actions • Setting Identify figurative language	Make complex inferences within and across texts to: • Infer mood or tone • Integrate ideas to determine theme • Identify or interpret a character's motivations and decisions • Examine relations between theme and setting or characters Explain how rhythm, rhyme, or form in poetry contribute to meaning	Consider text(s) critically to: • Evaluate the role of literary devices in conveying meaning • Determine the degree to which literary devices enhance a literary work • Evaluate a character's motivations and decisions • Analyze the point of view used by the author
Specific to Informational Text	Identify textually explicit information within and across texts, such as: • Topic sentence or main idea • Author's purpose • Causal relations Locate specific information in text or graphics	Make complex inferences within and across texts to: • Summarize major ideas • Draw conclusions and provide supporting information • Find evidence in support of an argument • Distinguish facts from opinions • Determine the importance of the information within and across texts	Consider text(s) critically to: • Analyze the presentation of information • Evaluate the way the author selects language to influence readers • Evaluate the strength and quality of evidence used by the author to support his or her position • Determine the quality of counterarguments within and across texts • Judge the coherence, logic, or credibility of an argument

Literary texts include fiction, literary nonfiction and poetry.
Informational texts include exposition, argumentation and persuasive text, and document and procedural materials.

Locate and Recall Information in Text

The first category of cognitive targets is **locate** and **recall.** As students locate or recall information from what they read, they may identify textually explicit information, including main ideas or supporting details, or they may find essential elements of a story, such as characters, time, or setting. Answering assessment items often involves matching information given in the item to either literal or synonymous information in the text before they can then use the textual information to develop a response. As readers engage in these cognitive behaviors, they monitor their reading in order to understand whether or not they are comprehending. When they realize that the text is not making sense, they employ specific strategies to ensure that they begin to comprehend again.

Items assessing this component of reading usually focus on information contained in relatively small amounts of text—a sentence, a paragraph, or two or more adjacent paragraphs or parts of a document. These items provide information about the most basic comprehension skills, those that ultimately form the foundation for a more elaborated understanding of what is read. At the same time, these items address the kinds of reading that occur routinely during in-school and out-of-school reading activities.

In addition, at grade 12 students may be asked to identify the thesis of a speech and the essential elements that elaborate it, or they may need to follow instructions in informational or procedural texts to perform specific tasks or answer questions.

Integrate and Interpret What Is Read

The next category of cognitive targets refers to what readers do as they **integrate** new information into their initial sense of what a passage says and often **interpret** what they read in the process. When readers engage in behaviors involving integrating and interpreting, they compare and contrast information or character actions, examine relations across aspects of the text, or consider alternatives to what is presented in the text. This aspect of the reading is critical to comprehension and can be considered the stage in which readers really move beyond the discrete information, ideas, details, themes, and so forth presented in text and extend their initial impressions by processing information logically and completely. As readers integrate information and interpret what they read, they frequently form questions, use mental images, and make connections that draw on larger sections of text, often at an abstract level. They are also drawing on their knowledge of the elements and structure of literary and informational text.

In applying these cognitive behaviors, readers think across large portions of text, across the text as a whole, or even across multiple texts; they relate textual information to knowledge from other sources, such as their previous content learning, or to internalized criteria and logic. Thus, readers might ask themselves whether something they are reading makes sense to them within the realm of their own experiences or when considered against what they have read in other sources. They examine the text in terms of their specific reading goals or the needs they have for the information that the text can provide them. In certain reading situations, readers may apply what they know to what they are reading, for example, determining a real-world application of suggestions in a text on bicycle safety. They also apply information gained from reading, in following instructions for repairing a bicycle or reading a map to determine where bike routes have been designated in a city.

Items assessing these cognitive behaviors might ask students to form generalizations about a piece of informational text or make statements about how the setting of a story contributes to the communication of theme. Other items might require interpretation of a character's motivations or of an author's reasons for attempting to persuade readers about an issue. Other questions might ask for alternative actions that a character might have taken or an interpretation of an implied message or moral from a story.

In more complex texts at grade 12, students may be asked to distinguish evidence from inferences, synthesize information from multiple sources, interpret and use graphical information embedded in texts, and identify interrelationships between and among ideas and concepts.

Critique and Evaluate Text

The final category of cognitive targets**—critiquing** and **evaluating text**—requires readers to stand back from what they read and view the text objectively. The focus remains on the text itself, but the readers' purpose is to consider the text critically. Readers assess the text from numerous perspectives and synthesize what is read with other texts and other experiences. Items targeted toward these behaviors may ask students to evaluate the quality of the text as a whole, to determine what is most significant in a passage, or to judge the effectiveness of specific textual features to accomplish the purpose of the text (for example, the effectiveness of details selected to support a persuasive argument). Items might ask for the likelihood that an event could actually have taken place, the plausibility of an argument, or the adequacy of an explanation for an event. Items can ask students to focus at the level of language choices (for example, nuances expressed in a metaphor) or at the broader level of the entire text (for example, evaluating the effectiveness of an author's craft to accomplish his or her overall goals). To answer these questions, students draw on what they know about text, language, and the ways authors manipulate language and ideas to achieve their goals.

At the twelfth grade, students may be asked to critique the presentation of information including the structure of a given argument; connections among evidence, inferences, and claims; how two or more authors reach similar or different conclusions; the ways in which the style and organizational structures of texts support or confound their meaning or purpose; and the clarity, simplicity, and coherence of texts and the appropriateness of their graphics and visual appeal. Additional cognitive targets under critique and evaluate for grade 12 include analysis of false premises/assumptions, faulty reasoning, loaded terms, caricatures, and/or sarcasm.

Writing Items Aligned to the Cognitive Targets

Although it is impossible to cover every cognitive target outlined in the 2009 NAEP Reading Framework on one assessment, appropriate alignment between the assessment and the NAEP Reading Framework at each grade should be maintained in the item pools. The assessment should be built so that the constructs represented by the cognitive targets for each text type are adequately represented. The breadth and relative emphasis of knowledge and skills covered in each area must be represented on the assessment as a whole. The developer should avoid under- or over-emphasizing particular cognitive targets to ensure broad coverage in any given year's item pool and coverage of all cognitive targets over time. To help align the assessment with the content framework, the following guidelines should be used:

- Each item on the assessment must measure part or all of one or more cognitive targets.

- For multiple-choice items, incorrect options should be related to inadequate or incomplete knowledge for the cognitive target(s) assessed.

- For constructed-response items, all criteria in the scoring rubrics should be related to the cognitive target(s) assessed.

- The items and tasks should not require students to use knowledge and skills irrelevant to the cognitive target(s) assessed.

- The cognitive targets toward which items are developed should be appropriate to the text type, unless the cognitive target is one that has been identified as applicable to both literary and informational texts.

The three categories of cognitive targets reflect different kinds of thought processes about the different kinds of text that students read. Even though the matrix represents a progression in terms of the depth of analysis of text from locate/recall through integrate/interpret to critique/evaluate, there is a range of difficulty within each cell of the cognitive targets matrix. In thinking about the three categories of cognitive targets, item writers should not consider locate/recall items as globally unchallenging, even though the readers' task involves considering text-explicit information, ideas, themes, and textual elements. Further, locate/recall items may not necessarily be *easier* for students than items targeted at other forms of thinking. Item writers should note that if students incorrectly respond to a locate/recall question, they might have no trouble responding to an integrate/interpret item correctly. Less skilled readers sometimes find close examination of the text (locate/recall) very difficult. It is not uncommon for less skilled readers to return to correct a locate/recall item after considering the text in response to an integrate/interpret item.

Items that embody the locate/recall process ask about central themes and secondary elements that support these ideas within the passage. The notion of analysis is key to the further distinction between integrate/interpret and critique/evaluate. The process within integrate/interpret involves contemplating primarily text-based information. The critique/evaluate process draws upon much more information from students' general stores of knowledge to analyze text-based information. This application of prior knowledge, which students' apply in constructing their response, is the subsequent, evaluative step. The difference between integrate/interpret and critique/evaluate is characterized by items that prompt students to make a judgment in their response. The focus in critique/evaluate also starts to shift away from the author to the individual student as the reader.

CHAPTER 4

INCLUSION OF STUDENTS WITH DISABILITIES AND ENGLISH LANGUAGE LEARNERS

The NAEP Reading Assessment is designed to measure the academic achievement of all test takers at a given grade level. The assessment is administered to English language learners and students with disabilities who, based on inclusion criteria provided by NAEP, are capable of participating. Special care is taken in designing and developing the assessment to ensure that these students, along with all others, find the passages and items accessible. Items are written in plain language, without jargon or unnecessarily complex syntactical structures.

As noted in Chapter 2, there are two ways that NAEP addresses the issue of accessibility. One is to follow careful item and assessment development procedures to build accessibility into the standard assessment. The other is to provide accommodations for students with disabilities and for English language learners.[23]

Students from diverse populations may need accommodations to be able to participate in the NAEP Reading Assessment. NAEP attempts to provide accommodations to students that match the way they are tested in school, as long as those accommodations do not alter the reading construct being measured. For example, large-print versions are made available for students with visual impairments; students with disabilities may be given one-on-one or small-group testing situations or extended time to complete the assessment. Some students, for example, those who are learning English, may have the test directions (but not the passages or items) read orally to them. Other students may benefit from having a trained aide transcribe dictated responses for them. Accommodations may be provided in combination, such as, extended testing time and individual administration of the assessment.

Test Accessibility Components

Multiple access points appropriate for the diverse population of students should be available throughout the assessment. Ways to strengthen access include the following:

- Paying careful attention to how items are presented to students in the assessment (e.g., plain language editing procedures, use of graphics, and item format considerations);
- Designing constructed-response items so that they allow multiple ways of responding, as appropriate to the knowledge and skill assessed;

[23]For more information about research that has guided the increased inclusion of students in NAEP, see Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progres*s (NCES Publication No. 97–482). Washington, DC: National Center for Education Statistics. Also see Mazzeo, J., Carlson, J.E., Voelkl, K.E., & Lutkus, A.D. (1999). *Increasing the participation of special needs students in NAEP: A report on 1996 research activitie*s (NCES Publication No. 2000–473). Washington, DC: National Center for Education Statistics.

- Developing scoring rubrics so that the targeted knowledge and skills are evaluated at all score levels;

- Formatting assessment booklets to allow enough space between items, using boxes and lines judiciously;

- Providing proper training procedures and materials for scorers;

- Providing adapted forms for students with disabilities, such as large-print versions; and

- Providing administration accommodations for English language learners and students with disabilities, as allowed by NAEP.

Accommodations

For many students with disabilities and students with a native language other than English, the standard administration of the NAEP assessment will be most appropriate. However, for some students with disabilities and some English language learners, the use of one or more administration accommodations will be more suitable. How to select and provide appropriate accommodations is an active area of research, and new insights are emerging on how to best apply accommodation guidelines to meet the needs of individual students. The NAEP Reading Accommodations Policy allows a variety of accommodations, depending on the needs of each student. Most accommodations that schools routinely provide in their own testing programs are allowed in the reading assessment, as long as they do not affect the construct tested. For example, it would NOT be appropriate to read aloud passages or test questions on the assessment. Accommodations are offered in combination as needed; for example, students who receive one-on-one testing generally also use extended time. Accommodations include, but are not limited to:

- one-on-one testing,
- small-group testing,
- extended time,
- oral reading of directions,
- large-print booklets, and
- use of an aide to transcribe responses.

In a very small number of cases, students will not be able to participate in the assessment even with the accommodations offered by NAEP. Examples of these situations include:

- students with disabilities whose Individualized Education Plan (IEP) teams or equivalent groups have determined that they cannot participate, or whose cognitive functioning is so severely impaired that they cannot participate, or whose IEP requires an accommodation that NAEP does not allow; and

- students with limited English proficiency who have received reading instruction primarily in English for less than three school years and who cannot participate in the assessment when it is administered in English, with or without an accommodation (according to NAGB policy as of 2004).

Item-Writing Considerations for English Language Learners

Many students who are learning English as their second language will take the standard, English-only version of the test. These students are diverse both across and within their language groups. This is particularly the case with Spanish language speakers who come from various countries in Latin America and the Caribbean. Among the Spanish-speaking population are linguistic differences (mainly in vocabulary), cultural differences, and differences in socio-economic backgrounds. English language learners may have difficulty understanding what items are asking for on assessment forms administered in English.[24]

Item writers should keep the following issues in mind to minimize the possibility of students' misinterpreting items:

- Rules of syntax or word order often differ in English and in other languages. While students may know the basic differences between their primary language and English, subtle differences that can lead to confusion based on word choice, syntax, or order should be avoided. Thus, items should be written in as straightforward a way as possible.

- The same word in English and in the student's native language can have different meanings; when possible, avoid using such words in items and related materials and in overall assessment instructions.

- Frequent dual meanings of English words can be unnecessarily confusing for English language learners and can be avoided by using a different word, defining the word, or providing a fuller context.

Item-Writing Considerations for Students With Disabilities

Most students with disabilities will take the standard assessment without accommodations, and those who take the assessment with accommodations will use the standard version of the test. Item writers and the assessment developer should minimize item characteristics that could hinder accurately measuring the reading achievement of students with disabilities. In particular:

- avoid layout and design features that could interfere with the ability of the student to understand the requirements and expectations of the item;

- use plain language, as described in Chapter 2; and

- develop items so that they can be used with allowed accommodations, such as text enlargement.

The item-writing considerations discussed in this chapter are not intended to imply that items should be written differently for these student populations, but that certain item-writing techniques can increase accessibility for all students, including English language learners and students with

[24]For more information about designing assessments that are accessible to English language learners, see Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State School Officers.

disabilities. Chapter 3 contains additional item-writing techniques that can increase accessibility for English language learners and students with disabilities, as well as for other students.

Scoring Responses From English Language Learners

Literacy issues and students' varied background experiences have an impact on how well scorers can properly read, understand, and evaluate the responses of English language learners to constructed-response items.[25] Responses sometimes can be difficult to read because of confusion between the students' native language and English. Although this is developmentally appropriate in terms of language acquisition, many scorers are not trained to interpret these types of systematic errors. The following procedures should be used to score these responses properly:

- Scoring leaders should have additional training in recognizing and properly interpreting responses from English language learners.

- Experts in reading responses of English language learners should be available to scorers in a consultancy role, if needed, during the scoring process.

- Responses showing systematic errors should be included in training materials for scorers so that scorers can more accurately identify and evaluate such responses. Major systematic language errors include the following:

  — intermittent use of the student's native language and English, called code switching;

  — use of native language phonetics in attempting to write English or beginning-stage English phonetic spelling;

  — use of writing conventions from the native language when students are responding in English;

  — word mergers (the condensing of words into one mega-word), transposition of words, and omission of tense markers, articles, plurals, prepositions, or other words;

  — substitution of common words for more precise terminology (e.g., it may be acceptable for students to substitute the word *fattest* for *greatest* when the intent of the item is not to evaluate students' understanding of terminology, but, if the intent is to measure students' knowledge and ability to use such terminology in an application setting, then this substitution would be incorrect.);

  — confusion about the meaning of words (e.g., *left* as opposed to *right* vs. *left* defined as that which is *remaining*);

  — inappropriate use of unfamiliar words;

[25]For more information about scoring responses from English language learners, see Kopriva, R., & Saez, S. (1997). *Guide to scoring LEP student responses to open-ended reading items*. Washington, DC: Council of Chief State School Officers.

— poorly-developed sentence and paragraph structures; and

— over reliance on nonverbal forms of communication, such as charts or pictures, that are embedded within written responses.

Novel interpretations and responses are common for English language learners and often reflect lack of familiarity with certain aspects of texts and with the vocabulary. It is important for scorers to evaluate responses based on the measurement intent of the item and recognize when an unusual response is actually addressing that intent.

It is not unusual for scoring rubrics to seem to favor writing styles that mirror what is taught in language arts curricula in U.S. schools. However, some cultures encourage more circular, deductive, and abbreviated writing styles; these are the models of writing with which students may be most familiar. Scorers should be sensitive to these types of responses so that they can be scored appropriately; they need to understand the nature, conventions, and approaches shown in these writing styles and to able to separate the structure of the written response from the substantive content being evaluated.

The scoring procedures discussed in this chapter are not intended to imply that items should be scored differently for these student populations, but that certain scoring procedures can be used to score student responses properly, including responses from English language learners and students with disabilities. Chapter 6 contains additional scoring rubric development and item-scoring procedures that apply to English language learners and students with disabilities, as well as to other students.

CHAPTER 5

VOCABULARY ASSESSMENT ON THE 2009 NAEP READING ASSESSMENT

The intent of the vocabulary assessment on the 2009 NAEP Reading Assessment is to determine whether readers know and understand the meanings of the words that writers use to convey new information or meaning, not to measure readers' ability to learn new terms or words. Hence, the assessment will focus on words that characterize the vocabulary of mature language users and characterize written rather than oral language. These will be words that convey concepts, ideas, actions, or feelings that the readers most likely know. In general, the words selected as targets for item development characterize the language of mature readers and are used in texts from a variety of content domains.[26]

The NAEP vocabulary items should be designed to assess readers' ability to connect an appropriate meaning to the candidate words to gain comprehension. Whereas, passage comprehension items are intended to measure readers' learning from text, vocabulary items should measure readers' knowledge of certain important words the author uses to impart this meaning. In sum, NAEP vocabulary items should elicit readers' sense of a word's meaning as it relates to passage comprehension, rather than precise definitions of terms.

Selecting Words to Assess

For meaning vocabulary items, item writers should select candidate words in the text that meet the following criteria:

- The word is unlikely to be part of the expressive (speaking and writing) vocabulary of grade-level readers.
- The word is of general use across contexts and domains.
- The word is not a technical term or jargon; it has broad application.
- The word is representative of the vocabulary found in challenging texts at the students' grade level.
- The word is central to constructing an understanding of a local part of the context of the passage and may be linked to central ideas, but it does not represent the key concept (e.g., the word *emancipation* would not be tested in an article dealing with the Emancipation Proclamation).

[26]Beck, McKeown, and Kucan refer to these as tier 2 words. This term distinguishes them from tier 1 words—common, everyday words basic to the speech and writing of most students—and from tier 3 words—rarely used words or technical terminology. See Beck, McKeown, and Kucan (2002), as cited in Appendix D.

- The text does not explicitly define the word, nor are context clues strong enough to allow a reader who is unfamiliar with the word to gain its meaning; however, enough context exists around the word to allow the reader's knowledge of the word to help the reader understand its role in the text.

Bound by these criteria, some passages that are otherwise of high quality and utility on the 2009 NAEP Reading Assessment may offer few or no candidate words for meaning vocabulary items. Conversely, a single passage may contain an abundance of candidate words. In these cases, items may be developed for the words that are more central to the passage. This selection of central words strongly demonstrates the connection between vocabulary and comprehension. In addition, it illustrates principles for the most effective kind of vocabulary instruction, as instruction based on centrally significant words is most productive.

Exhibit 17 summarizes the considerations for selecting words about which items are written and for words selected as distractors.

EXHIBIT 17
Considerations for Selecting Vocabulary Items and Distractors for the 2009 NAEP Reading Assessment

Vocabulary Words to Be Tested	Vocabulary Words Excluded From Testing	Considerations for Distractors
• Characterize the vocabulary of mature language users and characterize written rather than oral language • Label generally familiar and broadly understood concepts, even though the words themselves may not be familiar to younger learners • Are necessary for understanding at least a local part of the context and are linked to central ideas such that lack of understanding may disrupt comprehension • Are found in grade-level reading material	• Are narrowly defined and are not widely useful, such as those related to specific content domains (e.g., *photosynthesis*, *fiduciary*) or words with limited application (e.g., *deserter*, *hamlet*) • Label or name the main idea of the passage (e.g., the word *emancipation* would not be tested in an article dealing with the Emancipation Proclamation) • Are those already likely to be part of students' everyday speaking vocabulary at the grade level • Are those whose meanings are readily derived from language context (e.g., appositives, parenthetic definitions, idiomatic expressions)	• Present a different common meaning of the target vocabulary word, which must be ignored in favor of the meaning in context • May present correct information or content from the text that is NOT what is meant by the target word • May be an alternative interpretation of the context in which the target word occurs • May be the meaning of another word that looks or sounds similar to the target word • May present a common, but inaccurate association with the target word.

The process for reviewing passages, selecting vocabulary words, and developing items and distractors is a multi-step process. For each passage, item writers will need to proceed through the following steps.

- **Step 1:** Read the passage carefully to both fully understand it and to judge its relative difficulty for the grade it will be used to assess.

- **Step 2:** Reread the passage and identify possible words that are challenging within the relative difficulty of the passage. List the words in the order that they appear in the passage. Challenging words are those that:

  - are not likely to be part of the oral language of students for whom this passage is challenging but comprehensible
  - may be in the student's oral vocabulary but whose meaning in the text is different from or less common than the meaning of the word used orally (e.g., "fine," "suppose"); that is, there are potential alternative interpretations that readers might make

- **Step 3**:Review the list and cross out any words that :

  - could be omitted or substituted by the word "it" or an auxiliary verb without any loss of meaning to the context
  - label or name the main idea of the passage
  - are defined by the context
  - are narrowly defined

- **Step 4**: Engage in group discussion. Develop a list of candidate words by considering those words that are and are not common to all group members. In deciding to keep or omit words from the list, consider whether the target readers might draw an alternative interpretation of the word's meaning from the text that could be captured in item distractors. Eliminate words that do not meet the selection criteria.

- **Step 5**: Review the list of candidate words to ensure that each of the vocabulary words meets the criteria from column 1 of Exhibit 17. Eliminate words that are the least challenging. Select the words that are most centrally related to the content of the passage. These are the words around which items should be developed

- **Step 6**: Construct items and relevant distractors using general item development principles and the criteria for distractors in column 3 of Exhibit 17.

Developing Vocabulary Items

Meaning vocabulary items for the 2009 NAEP Reading Assessment should be presented only in a multiple-choice format. The items should avoid such terms as *define*, *definition*, and *meaning* to clarify that the items are asking for meaning vocabulary, not precise definitions. To reinforce this, the item stem should provide text references by quoting phrases or sentences as relevant to the candidate word, bolding the candidate word, and noting the location of the word within the passage

by paragraph or page number. At grades 8 and 12, some items will reflect more depth of processing in terms of a word's connectedness to other words and access to multiple meanings.

Use of these guidelines to select target words to assess and for developing items are illustrated in Appendix C on passage mapping.

CHAPTER 6

SCORING THE NAEP READING ASSESSMENT

The 2009 NAEP Reading Assessment will include multiple-choice and constructed-response items. Both item types yield valuable information about students' reading and allow a rich, full description of how the nation's students approach different kinds of texts. The inclusion of both types of items on the NAEP Reading Assessment affirms the complex nature of the reading process because it recognizes that different kinds of information can be gained from each item type. It also acknowledges the real-world skill of being able to write about what one has read.

This chapter elaborates on information presented in Chapter 3 about the development of scoring rubrics for multiple-choice and constructed-response items to be included on the 2009 NAEP Reading Assessment.

Multiple-Choice Items

The stem of a multiple-choice item is followed by four answer choices, or options, only one of which is correct. A generic scoring rubric can be used for all multiple-choice items, as they are scored dichotomously, either correct (1 point) or incorrect (0 points).

Exhibit 18 further defines the score categories for multiple-choice items.

EXHIBIT 18
Multiple-Choice Score Categories

Score Category and Description
Correct These responses represent the one correct option.
Incorrect These responses represent one of the three incorrect options.

Constructed-Response Items

In a constructed-response item, student answers are scored on a range of score categories from 2–4, depending on the item type (short constructed-response versus extended constructed-response) and the complexity of the item.

All constructed-response items will be scored using rubrics unique to each item. General principles that apply to these rubrics follow:

- Students will not receive credit for incorrect responses.
- All scoring criteria will be text based; students must support statements with information from the reading passage.

- Partial credit will be given for responses that answer a portion of the item but do not provide adequate support from the passage.

- Student responses will be coded to distinguish between blank items and items answered incorrectly.

- Responses will be scored on the basis of the response as it pertains to the item and the passage, not on the quality of writing.

- As part of the item review, the testing contractor will ensure a match between each item and the accompanying scoring guide.

Constructed-response items will be scored using primary trait scoring rubrics, where the traits are the three categories of cognitive behaviors—locate/recall, integrate/interpret, and critique/evaluate—as they apply to the two text types—literary and informational texts. Standard procedures for training scoring personnel and calibrating draft rubrics ensure accurate scoring. A general rubric will be developed for each open-ended item; after items are field tested, the draft rubrics will be anchored using a sample of student responses and made more specific.

Short Constructed-Response Items and Score Categories

Some short constructed-response items are written to be scored dichotomously, that is, either acceptable or unacceptable. Short constructed-response items with two scoring categories should measure knowledge and skills in a way that multiple-choice items cannot or provide greater evidence of the depth of students' understanding. Some short constructed-response items might be appropriate for measuring the same skills that multiple-choice items could measure when the intent is to eliminate student guessing, which could be a factor if a multiple-choice item were used. Short constructed-response items are also useful when more than one correct answer is possible, when there are different ways to display an answer, or when a brief explanation is required. Item writers should take care that short constructed-response items would not be better or more efficiently structured as multiple-choice items—they should not be simply multiple-choice items without the options.

Item writers must draft a scoring rubric for each short constructed-response item. For dichotomous items, the rubrics should define the following two categories:

1 = Acceptable
0 = Unacceptable

Exhibit 19 further defines the score categories for dichotomous short constructed-response items.

EXHIBIT 19
Dichotomous Short Constructed-Response Score Categories

Score Category and Description
Acceptable These responses represent an understanding of the text and a correct response to the item.
Unacceptable These responses represent a lack of understanding and an incorrect response to the item.

Some short constructed-response items are written to be scored on a three-category scale. Short constructed-response items with three scoring categories should measure knowledge and skills that require students to go beyond giving an acceptable answer. These items allow degrees of accuracy in a response so that a student can receive some credit for demonstrating partial understanding of a concept or skill.

For items with three score categories, the rubrics should define the following categories:

2 = Correct
1 = Partial
0 = Incorrect

Exhibit 20 further defines the score categories for short constructed-response items.

EXHIBIT 20
Short Constructed-Response Score Categories

Score Category and Description
Correct These responses represent an understanding of the text and a correct response to the item.
Partial These responses represent a partial understanding of the text and a partially correct response.
Incorrect These responses represent little or no understanding of the text and an incorrect response.

Extended Constructed-Response Items and Score Categories

In general, extended constructed-response items ask students to respond to a question by interpreting and analyzing information from the text. Extended constructed-response will usually have four scoring categories (with the possibility for additional score categories as appropriate):

3 = Extensive
2 = Essential
1 = Partial
0 = Incorrect

Exhibit 21 further defines the score categories for extended constructed-response items.

EXHIBIT 21
Extended Constructed-Response Score Categories

Score and Description
Extensive These responses represent an in-depth, rich understanding of the text and a correct response supported by multiple pieces of information from the passage.
Essential These responses represent a solid understanding of the text and a correct response supported by some information from the passage.
Partial These responses represent some understanding of the text and little or no information from the text as part of the response.
Incorrect These responses represent little or no understanding of the text and an incorrect response.

As they are developing items, writers should draft a corresponding scoring rubric so that both the item and the rubric reflect the construct being measured. Item writers must develop a draft scoring rubric specific to each extended constructed-response item. The rubric should clearly reflect the measurement intent of the item. The rubric should also contain the appropriate number of score categories. The number of score categories should represent an appropriate gradation of responses for each item and rubric. Item writers should anticipate the reasonable range of student responses when deciding on the number of score categories to include in the rubric for each item. The next section describes some requirements for writing scoring rubrics.

In developing the scoring rubric for an item, writers should think about what kind of student responses would show increasing degrees of knowledge and understanding. Writers should refer to the passage map for each item and sketch sample responses for each score category. Item writers also should include a justification or an explanation for each rubric category description. Doing so will assist the writer in developing a clear scoring rubric as well as provide guidance for scoring the item. Although specific scoring rubrics must be developed for each item, sample generic scoring rubrics for the different items types should be used as a basis and guide for developing the specific rubrics.

Alignment of Items and Rubrics

Item writers should refer to the measurement intent of the item when developing the scoring rubric. The number of categories used in the rubric should be based on the demand of the item.

1. ***Defining the Score Categories***

   Each score category must be distinct from the others; descriptions of the score categories should clearly reflect increasing understanding and skill in the reading constructs of interest. Distinctions among the categories should suggest the differences in student responses that would fall into each category; the definitions must be clear enough to use in training scorers. Each score level should be supported by the intent of the item. Factors unrelated to the measurement intent of the item should not be evaluated in the rubric. For example, because items are not meant to measure writing skills, the scoring rubric should be clear that the demonstration of reading in the response does not need to be tied to how well the response is written. However, if an explanation is part of the item requirement, the rubric should reflect that the explanations must be clear and understandable.

2. ***Measuring More than One Concept***

   If an item is measuring more than one skill or concept, the description of the score categories in the rubric should clearly reflect increasing understanding and achievement in each area. For instance, if the item is measuring both students' understanding of a skill and developing an appropriate approach, then the description of each category in the rubric should explain how students' understanding and skill are evaluated. If an item requires both an acceptable answer and an explanation, the rubric should show how these two requirements are addressed in each score category.

3. ***Specifying Response Formats***

   Unless the item is measuring whether a student can use a specified approach to a question, each score category should allow various approaches to the item. It should be clear in the rubric that different approaches to the item are allowed.

Chapter 8 presents sample passages and items, along with scoring rubrics.

CHAPTER 7

REVIEWS AND ITEM TRYOUTS

The test development contractor should build careful review and quality control procedures into the assessment development process. The NAGB NAEP Item Development and Review Policy Statement also provides detailed information about the development and review of items (see Appendix D). Although large-scale pilot testing provides critical item-level information for test development, other useful information about the items should be collected before and after pilot testing. Before pilot testing, items and scoring rubrics should be reviewed by experts in reading and measurement, including reading teachers and representatives of state education agencies, and by reviewers trained in sensitivity review procedures. After pilot testing, the items and the assessment as a whole should be reviewed to make sure that they are as free as possible from irrelevant variables that could interfere with allowing students' to demonstrate their knowledge and skills.

Sensitivity reviews are a particularly important part of the assessment development process. These reviews focus on identifying material that has the potential to be offensive to some test takers, stereotypes of individuals or situations, or material that may introduce construct-irrelevant factors in the assessment. They provide valuable guidance about the context, wording, and structure of items, and they identify flaws in items that confound the validity of the inferences for the groups of students they represent.

Two particularly useful procedures for collecting information about how items are working are described below. The information collected is valuable for determining whether items are measuring the construct as intended and for refining the items and scoring procedures before pilot testing. The information that the test development contractor garners from classroom tryouts and cognitive laboratory interviews should be provided to item writers to help them develop new items and revise existing items before field testing, and it can be used to enhance item-writing training and reference materials.

Classroom Tryouts

Classroom tryouts are an efficient and cost-effective way to collect information from students and teachers about how items and directions are working. Tryouts allow the test developer to troubleshoot the items and scoring rubrics. Classroom tryouts usually involve a nonrandom, small-scale sample; the students should reflect the range of student achievement in the target population as well as represent the diversity of examinees. The more the sample represents the various groups in the testing population, the more likely the tryout will identify areas that can be improved in the items.

In addition to providing student response data, tryouts can provide various kinds of information about the items, including what students and their teachers think the items are measuring, the appropriateness of the associated test materials, and the clarity of the instructions. Students can be asked to edit the items, for example, by circling words, phrases, or sentences they find confusing and suggesting improvements. Teachers can ask students what they thought each item was asking them to do and why they answered as they did and provide the information to the test developer. Teachers can also be asked to edit items and associated test materials. Item tryouts also are an efficient way to test how accommodations work and to try out other materials.

Student responses to the items should be reviewed by content and measurement experts to detect any problems in the items and should be used along with the other information gathered to refine the items and scoring rubrics. Using a sample that includes important groups in the population will allow reviewers to look for issues that might be specific to these groups. Responses also are useful in developing training materials and notes for scorers.

Cognitive Laboratory Interviews

In cognitive laboratory (cognitive lab) interviews, students are interviewed individually while they are taking, or shortly after they have completed, a set of items. Cognitive labs are used when it is necessary to gather more in-depth information about how an item functions than can be gathered through classroom tryouts. Students engage in a "think aloud" procedure that reveals their thought processes as they interact with text and formulate responses to items. Because cognitive labs highlight measurement considerations in a more in-depth fashion than other administrations can, they can provide important information for item development and revision. For example, cognitive labs can identify whether and why an item is not providing meaningful information about student achievement, provide information about how new formats are working, or verify the amount of time needed to read a passage and respond to a series of items.

The student samples used in cognitive labs are much smaller than those used in classroom tryouts. Students should be selected purposefully to allow an in-depth understanding of how an item is working and to provide information that will be helpful in revising items or in developing a particular type of item. For example, students from a range of backgrounds and with a range of reading abilities increase the usefulness of the information gained from cognitive labs.

CHAPTER 8

SAMPLE PASSAGES AND VOCABULARY ITEMS

This section illustrates the approach to vocabulary assessment recommended in the Framework by presenting the following:

1. The passage about which items were developed
2. A listing of words that were identified as likely candidates for item development from released NAEP passages
3. Two sample multiple-choice items for each passage

GRADE 4

HOW THE BRAZILIAN BEETLES GOT THEIR COATS

RETOLD BY ELSIE EELLS

In Brazil the beetles have such beautifully colored, hard-shelled coats upon their backs that they are sometimes set in pins and necklaces like precious stones. Once upon a time, years and years ago, they had ordinary plain brown coats. This is how it happened that the Brazilian beetle earned a new coat.

One day a little brown beetle was crawling along a wall when a big gray rat ran out of a hole in the wall and looked down **scornfully** at the little beetle. "O ho!" he said to the beetle, "how slowly you crawl along. You'll never get anywhere in the world. Just look at me and see how fast I can run."

The big gray rat ran to the end of the wall, wheeled around, and came back to the place where the little beetle was slowly crawling along at only a tiny distance from where the rat had left her.

"Don't you wish that you could run like that?" said the big gray rat to the little brown beetle.

"You are surely a fast runner," replied the little brown beetle politely. Her mother had taught her always to be polite and had often said to her that a really polite beetle never **boasts** about her own **accomplishments**. The little brown beetle never boasted a single boast about the things she could do. She just went on slowly crawling along the wall.

A bright green and gold parrot in the mango tree over the wall had heard the conversation. "How would you like to race with the beetle?" he asked the big gray rat. "I live next door to the tailor bird," he added, "and just to make the race exciting I'll offer a brightly colored coat as a prize to the one who wins the race. You may choose for it any color you like and I'll have it made to order."

"I'd like a yellow coat with stripes like the tiger's," said the big gray rat, looking over his shoulder at his **gaunt** gray sides as if he were already admiring his new coat.

"I'd like a beautiful, brightly colored new coat, too," said the little brown beetle.

The big gray rat laughed long and loud until his gaunt gray sides were shaking. "Why, you talk just as if you thought you had a chance to win the race," he said, when he could speak.

The bright green and gold parrot set the royal palm tree at the top of the cliff as the goal of the race. He gave the signal to start and then he flew away to the royal palm tree to watch for the end of the race.

The big gray rat ran as fast as he could. Then he thought how very tired he was getting. "What's the use of hurrying?" he said to himself. "The little brown beetle cannot possibly win. If I were racing with somebody who could really run it would be very different." Then he started to run more slowly, but every time his heart beat it said, "Hurry up! Hurry up!" The big gray rat decided that it was best to obey the little voice in his heart so he hurried just as fast as he could.

When he reached the royal palm tree at the top of the cliff he could hardly believe his eyes. He thought he must be having a bad dream. There was the little brown beetle sitting quietly beside the bright green and gold parrot. The big gray rat had never been so surprised in all his life. "How did you ever manage to run fast enough to get here so soon?" he asked the little brown beetle as soon as he could catch his breath.

The little brown beetle drew out the tiny wings from her sides. "Nobody said anything about having to run to win the race," she replied, "so I flew instead."

"I did not know that you could fly," said the big gray rat in a **subdued** little voice.

"After this," said the bright green and gold parrot, "never judge anyone by his looks alone. You never can tell how often or where you may find **concealed** wings. You have lost the prize."

Then the parrot turned to the little brown beetle who was waiting quietly at his side. "What color do you want your new coat to be?" he asked.

The little brown beetle looked up at the bright green and gold parrot, at the green and gold palm trees above their heads, at the green mangoes with golden flushes on their cheeks lying on the ground under the mango trees, at the golden sunshine upon the distant green hills. "I choose a coat of green and gold," she said.

From that day to this the Brazilian beetle has worn a coat of green with golden lights upon it.

And until this day, even in Brazil, where the flowers and birds and beasts and insects have such gorgeous coloring, the rat wears a dull gray coat.

Candidate Words for Item Development

scornfully

boasts

accomplishments

gaunt

subdued

concealed

Grade 4 Sample Items

When the rat says "I did not know that you could fly" in a subdued voice, this means the rat

(A) sounded very angry

*** (B) spoke very quietly**

(C) felt tired from running

(D) thought he had been tricked

When the parrot says that you can never tell "where you may find concealed wings," he is talking about wings that

*** (A) cannot be seen**

(B) have magical powers

(C) do not look like wings

(D) have dull colored feathers

GRADE 8

DOROTHEA DIX: QUIET CRUSADER
BY LUCIE GERMER

Dorothea Dix was so shy and quiet that it is hard to believe she had such a tremendous **impact** on nineteenth-century America. Yet almost single-handedly, she transformed the way people with mental illness were treated.

Dorothea was born in Maine in 1802 to a neglectful father and a mother who had trouble **coping** with daily activities. She ran away at the age of twelve to live with her grandmother, a cold, **inflexible** woman who nevertheless taught her the importance of doing her duty, as well as the **organizational** skills to help her do it.

Dorothea grew into an attractive woman, with blue-gray eyes, wavy brown hair, and a rich, low speaking voice. As a young adult, she spent her time teaching, writing books for children, and fighting the effects of tuberculosis. Despite her poor health, by age thirty-nine, she had saved enough money so that she had no financial worries. Afraid that her health was too poor for her to continue teaching, she looked forward to a lonely, unfulfilling life.

Then a friend suggested that she teach a Sunday school class for women in a Massachusetts jail. It would be useful without overtaxing her. On her first day, she discovered that among the inmates were several mentally ill women. They were anxious to hear what she had to say, but she found it impossible to teach them because the room was unheated. Dix, angry at this **neglect** on the part of the authorities, asked noted humanitarian Samuel Howe for his help in taking the case to court. The court ordered the authorities to install a wood stove.

This sparked Dix's interest in the ways mentally ill people were treated. Encouraged by Howe and education reformer Horace Mann, she spent two years visiting every asylum, almshouse, and jail in Massachusetts, quietly taking notes on the conditions. Her grandmother had trained her to be thorough and the training paid off.

Dix put her findings into a memorial (a report) that Howe presented to the Massachusetts legislature: "I tell what I have seen. . . . [I]nsane persons confined . . . in cages, closets, cellars, stalls, pens; chained, naked, beaten with rods and lashed into obedience."

The memorial caused an **uproar**: What kind of woman would be interested in such a subject and insist on discussing it in public? Gradually, the personal attacks **abated**, primarily because Dix's research had been so thorough and her results were so complete that no one could argue with them. Howe was able to push a bill through the Massachusetts legislature to enlarge the state asylum.

Dix spent the next few years systematically studying conditions and getting legislation passed in other states. Her health did not keep her from putting in long hours of hard work and travel. First, she studied the psychological and legal views of mental illness and its treatment. Before she went into a state, she examined local laws and previous proposals for change. Then she visited every institution, small or large, and met with administrators, politicians, and private citizens. She put all this information together in a memorial that was presented to the legislature. She also wrote newspaper articles to inform the public of her findings. By this time, she knew what kind of

opposition to expect, and she could help **deflect** it by appealing to the citizens' sense of pride or desire for economy. She also met privately with small groups of politicians to answer their questions and try to persuade them to come around to her point of view. She was usually successful, and public institutions to house and treat people with mental illness were established.

Unfortunately, that success did not carry over to her next goal: national legislation to improve the living conditions for people with mental illness. In the l850s, Congress passed a bill setting aside land for the establishment of national hospitals for those with mental illness, but President Franklin Pierce vetoed the bill on constitutional grounds.

Dix was shattered. Her health, which had been surprisingly good during her struggles, took a turn for the worse, and doctors recommended she take a long voyage. Dix was unable to relax, however, and her vacation turned into a marathon journey through Europe, as she examined the living conditions of mentally ill people in each place she visited. She spoke with doctors, government officials, and even the pope, pleading for humanitarian treatment for those who were mentally ill. She went as far east as Constantinople (now Istanbul) in Turkey and as far north as St. Petersburg (now Leningrad) in Russia. She was greeted respectfully everywhere she went, and many of her recommendations were followed.

She returned to the United States in 1857 and was appointed superintendent of women nurses during the Civil War. Dix was the only woman to hold an official position in the U.S. government during the war.

After the war, Dix continued her work on behalf of mentally ill people both in the United States and abroad. She died in 1887 at the age of eighty-five. Between 1841, when she began her crusade, and the year she died, thirty-two new hospitals for those who were mentally ill were built, most of them directly because she had brought the problem to the attention of people in power. Several other institutions in Canada and Europe, and even two in Japan, were established because of her influence. She also left a **legacy** of concern: No longer was mental illness treated as a crime, and her enlightened and tireless work led to more humane living conditions for people with mental illness.

Candidate Words for Item Development

impact	neglect	legacy
coping	uproar	
inflexible	abated	
organizational	deflect	

Grade 8 Sample Items

When the author says that personal attacks on Dorothea **abated,** the author means that

(A) the attacks became violent

* (B) there were fewer attacks

(C) people said rude things about her

(D) the police began to protect her

According to the text, when Dorothea knew what kind of opposition to expect she could **deflect** it. This means that Dorothea could

(A) avoid people who did not support her views

* (B) create arguments to convince people to help her

(C) write articles that all people could read

(D) be very polite to people who argued with her

GRADE 12

NEWTON MINOW
ADDRESS TO THE BROADCASTING INDUSTRY

I invite you to sit down in front of your television set. . .and keep your eyes glued to that set until the station signs off. I can assure you that you will observe a vast wasteland.

Newton Minow (1926–) was appointed by President John Kennedy as chairman of the Federal Communications Commission, the agency responsible for regulating the use of the public airwaves. On May 9, 1961, he spoke to 2,000 members of the National Association of Broadcasters and told them that the daily fare on television was "a vast wasteland." Minow's indictment of commercial television launched a national debate about the quality of programming. After Minow's speech, the television critic for *The New York Times* wrote: "Tonight some broadcasters were trying to find dark explanations for Mr. Minow's attitude. In this matter the viewer possibly can be a little helpful; Mr. Minow has been watching television."

. . . Your industry possesses the most powerful voice in America. It has an inescapable duty to make that voice ring with intelligence and with leadership. In a few years this exciting industry has grown from a novelty to an instrument of overwhelming impact on the American people. It should be making ready for the kind of leadership that newspapers and magazines assumed years ago, to make our people aware of their world.

Ours has been called the jet age, the atomic age, the space age. It is also, I submit, the television age. And just as history will decide whether the leaders of today's world employed the atom to destroy the world or rebuild it for mankind's benefit, so will history decide whether today's broadcasters employed their powerful voice to enrich the people or **debase** them. . . .

Like everybody, I wear more than one hat. I am the chairman of the FCC. I am also a television viewer and the husband and father of other television viewers. I have seen a great many television programs that seemed to me eminently worthwhile, and I am not talking about the much-bemoaned good old days of "Playhouse 90" and "Studio One."

I am talking about this past season. Some were wonderfully entertaining, such as "The Fabulous Fifties," the "Fred Astaire Show" and the "Bing Crosby Special"; some were dramatic and moving, such as Conrad's "Victory" and "Twilight Zone"; some were marvelously informative, such as "The Nation's Future," "CBS Reports," and "The Valiant Years." I could list many more—programs that I am sure everyone here felt enriched his own life and that of his family. When television is good, nothing—not the theater, not the magazines or newspapers—nothing is better.

But when television is bad, nothing is worse. I invite you to sit down in front of your television set when your station goes on the air and stay there without a book, magazine, newspaper, profit-and-loss sheet, or rating book to distract you—and keep your eyes glued to that set until the station signs off. I can assure you that you will observe a vast wasteland.

You will see a procession of game shows, violence, audience participation shows, formula

comedies about totally unbelievable families, blood and thunder, mayhem, violence, sadism, murder, Western badmen, Western good men, private eyes, gangsters, more violence and cartoons. And, endlessly, commercials—many screaming, **cajoling**, and offending. And, most of all, boredom. True, you will see a few things you will enjoy. But they will be very, very few. And if you think I exaggerate, try it.

Is there one person in this room who claims that broadcasting can't do better?. . .

Why is so much of television so bad? I have heard many answers: demands of your advertisers; competition for ever higher ratings; the need always to attract a mass audience; the high cost of television programs; the insatiable appetite for programming material—these are some of them. Unquestionably these are tough problems not **susceptible** to easy answers.

But I am not convinced that you have tried hard enough to solve them. I do not accept the idea that the present overall programming is aimed accurately at the public taste. The ratings tell us only that some people have their television sets turned on, and, of that number, so many are tuned to one channel and so many to another. They don't tell us what the public might watch if they were offered half a dozen additional choices. A rating, at best, is an indication of how many people saw what you gave them. Unfortunately it does not reveal the depth of the penetration or the intensity of reaction, and it never reveals what the acceptance would have been if what you gave them had been better—if all the forces of art and creativity and daring and imagination had been unleashed. I believe in the people's good sense and good taste, and I am not convinced that the people's taste is as low as some of you assume. . . .

Certainly I hope you will agree that ratings should have little influence where children are concerned. The best estimates indicate that during the hours of 5 to 6 p.m., 60 percent of your audience is composed of children under twelve. And most young children today, believe it or not, spend as much time watching television as they do in the schoolroom. I repeat—let that sink in—most young children today spend as much time watching television as they do in the schoolroom. It used to be said that there were three great influences on a child: home, school and church. Today there is a fourth great influence, and you ladies and gentlemen control it.

If parents, teachers, and ministers conducted their responsibilities by following the ratings, children would have a steady diet of ice cream, school holidays, and no Sunday school. What about your responsibilities? Is there no room on television to teach, to inform, to uplift, to stretch, to enlarge the capacities of our children? Is there no room for programs deepening their understanding of children in other lands? Is there no room for a children's news show explaining something about the world to them at their level of understanding? Is there no room for reading the great literature of the past, teaching them the great traditions of freedom? There are some fine children's shows, but they are drowned out in the massive doses of cartoons, violence, and more violence. Must these be your trademarks? Search your consciences and see if you cannot offer more to your young beneficiaries whose future you guide so many hours each and every day.

What about adult programming and ratings? You know, newspaper publishers take popularity ratings too. The answers are pretty clear; it is almost always the comics, followed by the advice-to-the-lovelorn columns. But, ladies and gentlemen, the news is still on the front page of all newspapers, the editorials are not replaced by more comics, the newspapers have not become one long collection of advice to the lovelorn. Yet newspapers do not need a license from the government to be in business—they do not use public property. But in television—where your responsibilities as public trustees are so plain—the moment that the ratings indicate that Westerns are popular, there are new imitations of Westerns on the air faster than the old coaxial cable could take us from Hollywood to New York. . . .

Let me make clear that what I am talking about is balance. I believe that the public interest is made up of many interests. There are many people in this great country, and you must serve all of us. You will get no argument from me if you say that, given a choice between a Western and a symphony, more people will watch the Western. I like Westerns and private eyes too—but a

steady diet for the whole country is obviously not in the public interest. We all know that people would more often prefer to be entertained than stimulated or informed. But your **obligations** are not satisfied if you look only to popularity as a test of what to broadcast. You are not only in show business; you are free to communicate ideas as well as relaxation. You must provide a wider range of choices, more diversity, more alternatives. It is not enough to cater to the nation's whims—you must also serve the nation's needs. . . .

Let me address myself now to my role, not as a viewer but as chairman of the FCC. . .I want to make clear some of the fundamental principles which guide me.

First, the people own the air. They own it as much in prime evening time as they do at 6 o'clock Sunday morning. For every hour that the people give you, you owe them something. I intend to see that your debt is paid with service.

Second, I think it would be foolish and wasteful for us to continue any worn-out wrangle over the problems of payola, rigged quiz shows, and other mistakes of the past. . . .

Third, I believe in the free enterprise system. I want to see broadcasting improved and I want you to do the job. . . .

Fourth, I will do all I can to help educational television. There are still not enough educational stations, and major centers of the country still lack usable educational channels. . . .

Fifth, I am unalterably opposed to governmental censorship. There will be no suppression of programming which does not meet with bureaucratic tastes. Censorship strikes at the taproot of our free society.

Sixth, I did not come to Washington to idly observe the **squandering** of the public's airwaves. The squandering of our airwaves is no less important than the lavish waste of any precious natural resource . . . .

What you gentlemen broadcast through the people's air affects the people's taste, their knowledge, their opinions, their understanding of themselves and of their world. And their future. The power of instantaneous sight and sound is without **precedent** in mankind's history. This is an awesome power. It has limitless capabilities for good—and for evil. And it carries with it awesome responsibilities—responsibilities which you and I cannot escape....

Reprinted from the NAEP Web site.

Candidate Words for Item Development

debase

cajoling

susceptible

obligations

squandering

precedent

Grade 12 Sample Items

When Minow speaks about commercials as **cajoling**, he is saying that some commercials

(A) are as violent as television shows

* (B) gently persuade people to buy products

(C) exaggerate the quality of products

(D) seem longer than television shows

When Minow speaks about the **squandering** of the public's airwaves, he is saying that

(A) broadcasters should pay attention to public opinion

(B) some televisions shows are subject to censorship

(C) producing televisions shows is too expensive

* (D) most broadcast time is used irresponsibly

CHAPTER 9

SPECIAL STUDIES

Three special studies have been proposed as part of the development of the 2009 NAEP Reading Framework. Although very different in topic, they have the common goals of improving the quality of the NAEP assessment and gaining maximum information about student achievement in reading. One of the special studies—meaning vocabulary—if conducted prior to the administration of the 2009 NAEP Reading Assessment, can inform test development by providing information about new item types. Other studies propose the use of data gained from the 2009 NAEP Reading Assessment to examine English learners' reading achievement as well as factors that have an impact on the gender gap. The special studies are presented in priority order, from highest to lowest.

MEANING VOCABULARY ASSESSMENT

Purpose

The inclusion of meaning vocabulary items represents a significant change in the design of the 2009 NAEP Reading Assessment. This developmental study, to be conducted well in advance of the 2009 administration, will examine the validity of the proposed method of measuring meaning vocabulary and will inform the development and use of meaning vocabulary items on NAEP.

Background and Rationale

Our understanding of vocabulary as related to reading has expanded over the years; consequently, researchers and educators have moved away from the notion of vocabulary as discrete, isolated words and toward a consideration of vocabulary as a meaningful factor in readers' comprehension of texts. This movement is evidenced by research that examines the complexity of word knowledge (see Nagy and Scott [2000] for a discussion of incrementality, polysemy, multidimensionality, interrelatedness, and heterogeneity) and describes different levels of understanding of the same words (see Beck and McKeown's [1991] review of dimensions of knowledge). Nonetheless, much remains to be learned. As Baumann et al. (2000) aptly note, "We know too much to say we know too little, and we know too little to say that we know enough" (p. 752).

The 2009 NAEP Reading Assessment will include a measure of vocabulary with items that function both as a measure of comprehension of the passage in which the word is included and as a test of readers' specific knowledge of the word's meaning as intended by the passage author. As has been discussed, the inclusion of meaning vocabulary represents a change for NAEP. Although past NAEP assessments included a few vocabulary test items in the context of passages, the number of items was scant and there were no specific vocabulary criteria for selecting the items or distractors. Further, past reports from NAEP provided little information on how students performed on the vocabulary items and whether that performance was associated with comprehension achievement levels; thus, these reports did not provide a foundation for emphasizing the importance of vocabulary to reading comprehension. The importance of vocabulary in reading comprehension will be much more widely understood and disseminated with:

- NAEP's initiative specifying vocabulary as a major component of reading comprehension;
- NAEP reports providing quantitative data about the performance of grades 4, 8, and 12 students on meaning vocabulary questions and the developmental differences among grades; and
- NAEP reports describing the differences among advanced, proficient, basic, and below basic readers on vocabulary and the implications of these differences.

Guided by the following research questions, this developmental study will ensure that NAEP provides reliable and valid data for analysis and interpretation regarding meaning vocabulary, thus enabling a more comprehensive picture of students' reading achievement.

Research Questions

1. What is the correlation between reading comprehension and meaning vocabulary items, and how does the addition of meaning vocabulary items affect overall scores on the NAEP Reading Assessment?
2. How does the introduction of meaning vocabulary items affect the scores of ethnically, socioeconomically, and geographically varying groups and low-, average-, and high-performing readers?
3. What is the correlation between scores on the meaning vocabulary items and a vocabulary test such as the Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III)? Answers to this question will address the concurrent validity of NAEP's vocabulary measure.

Methodology

This study will evaluate the reliability and construct, content, and criterion validity of the proposed method of measuring meaning vocabulary, well in advance of the 2009 administration of the assessment. The contractor will develop the methodology to address the research questions. Methods may include developing and mapping meaning vocabulary items on the basis of released NAEP reading passages and using cognitive laboratory studies to determine student responses to these items.

References

Baumann, J.F., Kame'enui, E.J., & Ash, G.E. (2002). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J.R. Squire, & J.M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (pp. 752–785). Mahwah, NJ: Erlbaum.

Beck, I. & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 789–814). Mahwah, NJ: Erlbaum.

Nagy, W.E., & Scott, J.A. (2000). Vocabulary processes. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 269–284). Mahwah, NJ: Erlbaum.

ENGLISH LANGUAGE LEARNERS

Purpose

This special study will examine the patterns of achievement among English language learner (ELL) students and the link between NAEP scores and other indicators of students' ability and achievement, as well as the effects of the accommodations afforded students in these groups.

Background and Rationale

English language learners (ELLs)[27] represent the fastest growing school-age population in the United States. The National Clearinghouse for English Language Acquisition (NCELA) determined that whereas the K–12 population increased by 12% between the 1991–92 and 2001–02 school years to 48 million students, the number of youth classified as ELLs increased by 85% during the same time period, from 2.4 million students to 4.7 million students (NCELA, 2002). The ELL population is diverse, representing a great range in terms of socioeconomic status, native language, and place of birth (García, 2000). Spanish is the language spoken by nearly 80% of the population whose home language is not English; the remaining 20% speak more than 300 languages, including Vietnamese, Hmong, Chinese, and Korean.[28] The definition of ELL students is made more complex by the reclassified population, which includes students who have achieved a level of English proficiency necessary to progress in school, but may still require some level of ELL accommodations (Linquati, 2001).

The diversity of this student population has implications for reading instruction, as educators seek better ways to teach reading to ELL students; however, the research available to inform reading instruction for ELLs is somewhat limited. Garcia (2000) notes a general lack of research on bilingual reading; the focus has been mainly on oral language development and vocabulary acquisition, with some examination of students' use of metacognitive and cognitive reading strategies. The research has been somewhat limited in terms of the subgroups studied. A sampling of the literature related to ELLs indicates a strong focus on Spanish speakers' experiences in learning English (see, for example, García, 1991; Jiménez, García, & Pearson, 1996; Nagy, McClure, & Mir, 1997), and many studies focus on ELL students who are already good readers (Peregoy & Boyle, 2000).

The literature further notes implications for the ELL population in terms of assessment. Abedi (2004b) indicates that content-area tests administered in English may "inadvertently function" as English language proficiency tests for ELLs. Students may be unable to demonstrate their knowledge because of unfamiliarity with the structure of questions or vocabulary forms, or they might interpret an item literally. The literature is mixed on the appropriateness of providing accommodations when testing students identified as ELLs. Holmes and Anstrom (1997) define two continuums for assessing ELLs, based on the purpose of the assessment: If the assessment compares ELL students with their peers, no modifications are necessary, but to assess the skills and knowledge of individual ELL students, accommodations should be provided as necessary.

[27] The literature uses the following terms interchangeably to refer to students who are in the process of acquiring two languages: English Language Learners (ELL), English as a Second Language (ESL), Limited English Proficient (LEP), bilingual.

[28] National Clearinghouse for English Language Acquisition. (2001). *Language backgrounds of limited English proficient (LEP) students in the U.S. and outlying areas, 2000–2001*. Retrieved July 14, 2004, from the NCELA Web site: http://www.ncela.gwu.edu/stats/toplanguages/langsrank.pdf

The special study provides the opportunity for a timely, national-level examination of ELL students' reading achievement. Although some studies have examined the reading and test-taking strategies of ELLs, the applicability of those findings are limited because the bulk of the research has focused on students whose first language was Spanish (García, 1991; Jiménez, García, & Pearson, 1996; Nagy, McClure, & Mir, 1997). Although Spanish is the native language of over two-thirds of the ELL population in the public schools, it is imperative not to overlook the fact that the other one-third of the ELL population speak more than 300 languages.[29] Abedi (2004b) also points to the limited applicability of test scores from assessments that include ELLs; those scores are often not controlled for variables such as year of arrival—if the student was born abroad—and mode of English instruction. Ideally the special study would offer policymakers, educators, and the public more precise insights into how ELLs differ in the ways that they read and write as well as how the range of English language proficiency amongst ELLs could be expected to perform on national assessments such as NAEP. The literature on how ELLs labeled as reclassified fluent English proficient (RFEP)[30] perform in comparison to their native English speaking peers over a span of time is limited to reviews of data on state assessments and lacks any tests for statistical significance in the findings (Linquati, 2001).

Further, the special study could provide information on questions such as when and how to include ELLs in national assessments such as NAEP. There is also considerable debate over the types of accommodations that can be used to include ELLs in assessments, but the question of whether such accommodations would jeopardize the test validity and comparability of test scores across different categories of students has remained largely unanswered (Hafner, 1997; Kopriva, 2000; Olson & Goldstein, 1996; Abedi 2004a).

Although past NAEP reports have provided scores by ethnicity, they have not provided information about the link between language minority students and reading ability. This special study seeks to examine this link, informing the discussion of how to develop a dynamic assessment (adaptive testing) that more accurately maps the achievement of U.S. students.[31]

Research Questions

1. What miscues occur most frequently among different groups of English language learners, and are these miscues consistent with different groups of English learners' speech?

2. Are tests of English language proficiency predictive of NAEP comprehension and vocabulary scores?

3. What are the differential effects of English proficiency level on NAEP reading and vocabulary?

[29] NCELA (2001), *Op cit.*

[30] Students labeled RFEP are also known as Fluent English Proficient (FEP).

[31] The English language learner special study may be informed by the results of the National Literacy Panel's study on language minority children and youth. The NLP is conducting a comprehensive review of research on the development of literacy among language minority children and youth, to be completed in 2004.

4. How are reclassified fluent English proficient students (RFEP) achieving compared with other groups in reading comprehension and vocabulary, and how do they progress after one, two, or three years of reclassification?

5. At what minimum level of English proficiency is a student able to take the NAEP reading assessment in English?

6. Do accommodations provided to ELL students increase accessibility without changing the construct being measured by the assessment, or is the validity of the score compromised by these accommodations?

Methodology

This study will look specifically at the NAEP assessment design and at achievement data gathered from the 2009 administration of the assessment. The contractor will develop the methodology to address the research questions.

References

Abedi, J. (2004a). *Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues*. Los Angeles: University of California, Los Angeles, CRESST.

Abedi, J. (2004b). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4–14.

García, G.E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *International Reading Association, 25*(4), 371–392.

García, G.E. (2000). Bilingual children's reading. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *The handbook of reading research* (Vol. II, pp. 813–834). Mahwah, NJ: Erlbaum.

Hafner, A.L. (1997, October). *Assessment accommodations that provide valid inferences for LEP students: Valid inferences for LEP students.* Los Angeles: California State University, Los Angeles, Program Evaluation and Research Collaborative.

Holmes, D.H., & Anstrom, K. (1997, October). *High stakes assessment: A research agenda for English language learners—symposium summary*. Washington, DC: National Clearinghouse for Bilingual Education.

Jiménez, R.T., García, G.E., & Pearson, P.D. (1996). The reading strategies of Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, 31(1), 90–112.

Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners.* Washington, DC: Council of Chief State School Officers.

Linquati, R. (2001, September). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners.* Santa Barbara: The University of California Linguistic Minority Research Institute.

Nagy, W.E., McClure, E.F., & Mir, M. (1997). Linguistic transfer and the use of context by Spanish-English bilinguals. *Applied Psycholinguistics*,

Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficiency students in large-scale assessment: A summary of recent progress* (NCES Publication No. 97–482). Washington, DC: National Center for Education Statistics.

Peregoy, S.F., & Boyle, O.F. (2000, Autumn). English learners reading English: What we know, what we need to know. *Theory into Practice*, 39(4), 237–247.

GENDER DIFFERENCES

Purpose

This special study examines the differences in reading achievement between boys and girls, focusing on factors that are associated with the gender gap in reading.

Background and Rationale

The gender gap—a significant difference between the performance or achievement of boys versus girls—exists in a number of education-related settings and situations. Girls generally have higher high school graduation rates, college admission rates, and enrollment in Advanced Placement courses in the humanities, whereas boys have a higher incidence of diagnosed reading disorders. Although boys generally have higher mathematics and science achievement, the gender gap in the language arts favors girls. Results from the 2002 NAEP Reading Assessment indicate the following:

- The score gap between male and female grade 4 students in 2002 was smaller than in 2000, but it was not found to be significantly different from that in 1992.
- The score gap between boys and girls at grade 8 was smaller in 2002 than in all prior assessment years.
- The score gap between grade 12 boys and girls in 2002 is greater than it was in 1992.
- Girls outperformed boys at all three grades in 2002.

Other measures have revealed similar findings regarding reading performance among girls and boys. The PIRLS study, for instance, reported significantly higher average scale scores for girls in the United States in both 1991 and 2001. PISA (OECD, 2000) reported higher performance by females in reading literacy for all countries in the study, indicating a difference of 28 points between boys and girls in the United States. The enduring consistency of the gender gap in reading has generated interest in research related to specific aspects of reading among girls and boys—and more recently, the literature indicates a notable increase in research focused on boys' (rather than girls') school-related practices and experiences, including some aspects of reading (Weaver-Hightower, 2003).

In terms of reading preferences and practices, the literature indicates a long-standing recognition of distinct differences among boys and girls (for a review, see Greenlaw & Wielan, 1979). More recently, Coles and Hall (2002) report that boys tend to read science fiction and fantasy, but overall read less fiction than girls, preferring books about sports and war and humorous (comic or joke) books. Guzzetti, Young, Gretsawage, Fyfe, Hardenbrook (2002) similarly indicate that boys prefer visual media, including electronic text and periodicals, which focus on sports, games, and electronics. Additionally, Millard's (1997) research indicates that boys tend to describe school as the place where they read most, while many more girls reported reading as a leisure activity and sharing books with others as a social activity.

Clearly, children acquire literacy in a complex environment, including innumerable factors that make the causes of the gender gap in reading difficult to establish and manage. A significant literature examines these factors through various lenses, focusing on reading as a gendered practice and highlighting a variety of factors—existing in families, peer groups, schools, and other social institutions—that may influence boys' and girls' reading preferences and habits. Many researchers argue that literacy, as defined in the school setting, has been constructed in a way that conflicts with dominant conceptions of masculinity (see, for example, Alloway & Gilbert, 1997; Dutro, 2001; Martino, 1995). Gambell and Hunter (2000) provide a useful classification of five schools of thought—which they term division of family labor, character-personification, classroom interaction, assessment bias, and identification with genre—that the literature emphasizes as major factors in the perpetuation of the gender gap in reading.

Researchers have examined various aspects of gender-specific performance, often based on assessment data and often including other subjects as well as reading. Just a small sampling reveals the diversity of topics considered by such research: girls perform better on constructed-response items and boys do better with multiple-choice items (Pomplun & Sundbye, 2000); the range of scores tends to be broader among males than females, with more scores represented at the very high and very low ends of the spectrum (Cole, 1997); and gender differences in terms of achievement scores do not appear to vary much by race/ethnicity (Coley, 2001).

As educators continue to grapple with the gender gap's implications for instruction and assessment, this special study will facilitate a timely, comprehensive examination of the gender gap in reading at a national level. The special study will examine the relationships among many of the literature-based factors relevant to the gender gap in reading, focusing on variables in NAEP's assessment design and their relationship to the gender gap.

Research Questions

1. How are question response modes (e.g., multiple-choice, constructed-response) related to reading achievement?

2. How are the types of texts (e.g., fiction, informational, poetry) related to reading achievement?

3. How is the content of the selection (e.g., gender of main character, different themes, presence of moral) related to reading achievement?

Methodology

This study will look specifically at the NAEP assessment design and at achievement data gathered from the 2009 administration of the assessment. The contractor will develop the methodology to address the research questions.

References

Alloway, N., & Gilbert, P. (1997). Boys and literacy: Lessons from Australia. *Gender & Education, 9*(1), 49–61.

Cole, N. (1997). *The ETS gender study: How females and males perform in educational settings.* Princeton, NJ: Educational Testing Service.

Coles, M., & Hall, C. (2002). Gendered readings: Learning from children's reading choices. *Journal of Research in Reading, 25*(1), 96–108.

Coley, R. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work.* Princeton, NJ: Educational Testing Service.

Dutro, E. (2001). "But that's a girls' book!": Exploring gender boundaries in children's reading practices. *The Reading Teacher, 55*(4), 376–384.

Gambell, T., & Hunter, D. (2000). Surveying gender differences in Canadian school literacy. *Journal of Curriculum Studies, 32*(5), 689–719.

Greenlaw, M., & Wielan, O. (1979). Reading interests revisited. *Language Arts,* 56, 432–433.

Guzzetti, B., Young, J.P., Gritsavage, M., Fyfe, L., & Hardenbrook, M. (2002). *Reading, writing, and talking gender in literacy learning.* Newark, DE: International Reading Association.

Martino, W. (1995). Deconstructing masculinity in the English classroom: A site reconstituting gendered subjectivity. *Gender & Education, 7*(2), 205–221.

Millard, E. (1997). Differently literate: Gender identity and the construction of the developing reader. *Gender & Education, 9*(1), 31–49.

Organisation for Economic Cooperation and Development (OECD). (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: Author.

Pomplun, M., & Sundbye, N. (2000). Gender differences in constructed-response reading items. *Applied Measurement in Education, 12*(1), 95–109.

Weaver-Hightower, M. (2003). The "boy turn" in research on gender and education. *Review of Educational Research, 73*(4), 471–498.

APPENDIX A

PARTICIPANTS IN THE DEVELOPMENT OF THE 2009 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN READING

STEERING COMMITTEE

Marilyn Adams
Chief Scientist
Soliloquy Learning Corporation
Needham, MA

Phyllis Aldrich
Gifted and Talented Coordinator
Saratoga-Warren BOCES
Saratoga Springs, NY

Francie Alexander
Vice President and Chief Academic Officer
Scholastic, Inc.
New York, NY

Patricia Alexander
Professor, College of Education
University of Maryland
College Park, MD

Lance Balla
Teacher
Snohomish High School
Snohomish, WA

Cynthia Teter Bowlin
Professor
Dallas County Community College
Dallas, TX

Wanda Brooks
Assistant Professor, Department of Education
University of Maryland, Baltimore County
Baltimore, MD

Leila Christenbury
Professor, School of Education
Virginia Commonwealth University
Richmond, VA

Mary Beth Curtis
Professor, School of Education
Director, Center for Special Education
Lesley University
Cambridge, MA

JoAnne Eresh
Senior Associate
Achieve, Inc.
Washington, DC

Alan Farstrup
Executive Director
International Reading Association
Newark, DE

Vincent Ferrandino
Executive Director
National Association of Elementary School Principals
Alexandria, VA

Mike Frye (Retired)
Section Chief, English Language Arts and Social Studies
North Carolina Department of Public Instruction
Raleigh, NC

Margo Gottlieb
Director, Assessment and Evaluation
Illinois Resource Center
Des Plaines, IL

Jane Hileman
Founder, 100 Book Challenge Company
King of Prussia, PA

Billie J. Orr (Retired)
President
Education Leaders Council
Washington, DC

Melvina Pritchett-Phillips
Resident Practitioner, Adolescent Literacy & Professional Development
National Association of Secondary School Principals
Reston, VA

Sandra Stotsky
Research Scholar
Northeastern University
Boston, MA

Julie Walker
Executive Director
American Association of School Librarians, a Division of the American Library Association
Chicago, IL

PLANNING COMMITTEE

Michael Kamil, Chair
Professor, School of Education
Stanford University
Stanford, CA

Peter Afflerbach
Professor, College of Education
University of Maryland
College Park, MD

Donna Alvermann
Professor, College of Education
University of Georgia
Athens, GA

Amy Benedicty
Teacher
Peninsula High School
San Bruno, CA

Robert Calfee
Dean, Graduate School of Education
University of California–Riverside
Riverside, CA

Mitchell Chester
Assistant Superintendent
Ohio Department of Education
Columbus, OH

Barbara Foorman
Director, Center for Academic and Reading Skills
University of Texas–Houston
Houston, TX

Irene Gaskins
Director
Benchmark School
Media, PA

Carol Jago
Teacher
Santa Monica High School
Santa Monica, CA

Jolene Jenkins
Teacher
Mahaffey Middle School
Fort Campbell, KY

Janet Jones
Reading Resource Teacher
Berry Elementary School
Waldorf, MD

Marilyn Joyce
Teacher
Brewer High School
Brewer, ME

Michael Kibby
Professor, Department of Learning and Instruction
SUNY Buffalo
Amherst, NY

Margaret McKeown
Senior Scientist
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA

Paula Moseley
Coordinator, Planning, Assessment and Research,
Student Testing Unit
Los Angeles Unified School District
Los Angeles, CA

Jean Osborn
Education Consultant
Champaign, IL

Charles Peters
Professor, School of Education
University of Michigan
Ann Arbor, MI

Carol Santa
Director of Education
Montana Academy
Kalispell, MT

Karen Wixson
Dean, School of Education
University of Michigan
Ann Arbor, MI

Junko Yokota
Professor, Reading and Language Arts
National-Louis University
Evanston, IL

Olivia Zarraluqui
Teacher
Our Lady of Lourdes Academy
Miami, FL

TECHNICAL ADVISORY PANEL

Patricia Gandara
Professor, School of Education
University of California at Davis
Davis, CA

Paul LaMarca
Director, Department of Assessment and Accountability
Nevada Department of Education
Carson City, NV

William Schafer
Affiliated Professor (Emeritus)
University of Maryland
College Park, MD

EXTERNAL REVIEW PANEL

To obtain an independent review of the draft 2009 NAEP Reading Framework, the National Assessment Governing Board (NAGB) commissioned a panel of prominent reading researchers and scholars to examine the draft document. After a three-month review period, the panel reported to NAGB on issues such as whether the Framework is supported by scientific research; whether the document reflects what students should know and be able to do in grades 4, 8, and 12; the appropriateness of proposed reading materials; and the clarity and organization of the draft Framework. Members of the Reading External Panel are listed below.

Dennis J. Kear, Panel Chair
Professor of Curriculum and Instruction, College of Education
Wichita State University
Wichita, KS

Ellin O. Keene
Deputy Director
Cornerstone National Literacy Initiative
University of Pennsylvania
Philadelphia, PA

Katherine A. Mitchell
Director, Alabama Reading Initiative
Alabama State Department of Education
Montgomery, AL

Keith E. Stanovich
Professor, Ontario Institute for Studies in Education
University of Toronto
Toronto, ON, Canada

Joanna P. Williams
Professor, Psychology and Education
Teachers College, Columbia University
New York, NY

APPENDIX B

GLOSSARY OF TERMS

GLOSSARY OF TERMS

This glossary provides brief definitions of terms used throughout the NAEP Reading Framework. The terms are defined according to their use in the Framework. Included in the list are terms that relate to types of reading materials, text structures and features, techniques of author's craft, and other key terms used in the Framework.

Allegory: A story in which the characters, settings, and events stand for abstract moral concepts.

Alliteration: The repetition of initial consonant sounds.

Allusion: A reference to a mythological, literary, or historical person, place, or thing.

Analogy: A comparison of two things to show their likenesses in certain respects.

Argumentation: Writing that seeks to influence through appeals that direct readers to specific goals or try to win them to specific beliefs.

Audience: A writer's targeted reader or readers.

Author's craft: The specific techniques that an author chooses to relay an intended message.

Autobiography: A written account of the author's own life.

Ballad: A song or songlike poem that tells a story.

Biography: An account of a person's life written by another person.

Causation: A text structure that presents causal or cause and effect relationships between the ideas presented in the text.

Cognitive target: A mental process or kind of thinking that underlies reading comprehension; cumulatively, the cognitive targets will guide the development of items for the assessment.

Coherence: The continuity of meaning that enables others to make sense of a text.

Comic relief: An event or character that serves as an antidote to the seriousness of dramatic events.

Comparison: A text structure in which ideas are related to one another on the basis of similarities and differences. The text presents ideas that are organized to compare, to contrast, or to provide an alternative perspective.

Conflict: A struggle or clash between opposing characters, forces, or emotions.

Connotation: The implicit rather than explicit meaning of a word. It consists of the suggestions, associations, and emotional overtones attached to a word.

Description: A text structure that presents a topic, along with the attributes, specifics, or setting information that describe that topic.

Denotation: The exact, literal definition of a word independent of any emotional association or secondary meaning.

Detail: A fact revealed by the author or speaker that supports the attitude or tone in a piece of poetry or prose. In informational text, details provide information to support the author's main point.

Diction: Word choice intended to convey a certain effect.

Elegy: A poem that mourns the death of a person or laments something lost.

Epic: A long narrative poem that relates the great deeds of a hero who embodies the values of a particular society.

Exaggeration or Hyperbole: A deliberate, extravagant, and often outrageous overstatement. It may be used for either serious or comic effect.

Exposition: One of the classifications of discourse whose function is to inform or to instruct or to present ideas and general truths objectively. Exposition presents information, provides explanations and definitions, and compares and contrasts.

Fable: A brief story that teaches a moral or practical lesson about life.

Fantasy: A story employing imaginary characters living in fictional settings where the rules of the real world are altered for effect.

Fiction: Imaginative literary works representing invented rather than actual persons, places, and events.

Figure of speech: A word or phrase that describes one thing in terms of something else, often involving an imaginative comparison between seemingly unlike things.

Flashback: A scene that interrupts the action of a work to show a previous event.

Fluency: The ability to read text quickly and accurately and comprehend what is read.

Foil: A character who sets off another character by strong contrast.

Folktale: A short story from the oral tradition that reflects the mores and beliefs of a particular culture.

Foreshadowing: The use of hints or clues in a narrative to suggest future action.

Free verse: Poetry that has no regular meter or rhyme scheme.

Genre: A category used to classify literary and other works by form, technique, or content.

Grammar: A coherent text structure on which readers rely as they seek to understand what they read; often referred to as "story grammar."

Graphic: A pictorial representation of data or ideas using columns, matrices, or other formats. Graphics can be simple or complex, present information in a straightforward way as in a list or pie graph, or embed or *nest* information within the document's structure. Graphics may be included in texts or be *stand-alone* documents (grade 12 only).

Historical fiction: A story that recreates a period or event in history and often uses historical figures as characters.

Iambic pentameter: A line of poetry made up of five metrical feet or units of measure, consisting of an unstressed syllable followed by a stressed syllable.

Imagery: Multiple words or a continuous phrase that a writer uses to represent persons, objects, actions, feelings, and ideas descriptively by appealing to the senses.

Inference: The act or process of deriving logical conclusions from premises known or assumed to be true; the conclusions drawn from this process.

Irony: The tension that arises from the discrepancy, either between what one says and what one means (verbal irony), between what a character believes and what a reader knows (dramatic irony) or between what occurs and what one expects to occur (situational irony).

Legend: An inscription or title on an object (e.g., a key to symbols used on a map). An unverified popular story handed down from earlier times.

Literary device: A literary technique used to achieve a particular effect.

Literary heritage: Works by authors whose writing influenced and continues to influence the thinking, history, and politics of the nation. These culturally and historically significant works comprise the literary and intellectual capital drawn on by later writers and to the present day.

Literary nonfiction: A text that conveys factual information but employs a narrative structure and characteristics, such as dialogue, and includes distinctly literary elements and devices. The text may be read to gain enjoyment and also to learn and to appreciate the specific craft behind authors' choices of words, phrases, and structural elements.

Lyrical poetry: Poems that focus on expressing emotions or thoughts.

Meaning Vocabulary: The application of one's understanding of word meanings to passage comprehension.

Memoir: A type of autobiography that usually focuses on a single time period or historical event.

Metaphor: A comparison of two unlike things without the use of *like* or *as*.

Mixed Text: Text that employs literary techniques usually associated with narrative or poetry while also presenting information or factual material with the dual purpose of informing and offering reading satisfaction; requires readers to discern bias from fact.

Monologue: A long, formal speech made by a character.

Mood: The atmosphere or predominant emotion in a literary work.

Motivation: A circumstance or set of circumstances that prompts a character to act a certain way or that determines the outcome of a situation or work.

Myth: A traditional story accepted as history which serves to explain the world view of a people.

Narration: The telling of a story in writing.

Narrative poetry: Poems that tell a story in verse, often focusing on a single incident.

Ode: A long lyric poem on a serious subject often for ceremonial or public occasions.

Onomatopoeia: The use of words that mimic the sounds they describe; imitative harmony.

Parody: The imitation of a work of literature, art, or music for amusement or instruction.

Parallel structure: The repetition of words, phrases, or sentences that have the same grammatical structure or that restate a similar idea.

Personification: A metaphor that gives inanimate objects or abstract ideas human characteristics.

Perspective: A position, stance, or viewpoint from which something is considered or evaluated.

Persuasion: A form of discourse whose function is to convince an audience or to prove or refute a point of view or an issue.

Plot: The sequence of events or actions in a short story, novel, or narrative poem.

Point of view: The perspective from which a narrative is told or the way in which the author reveals characters, actions, and ideas.

Problem/Solution: A text structure in which the main ideas are organized into two parts: a problem and a subsequent solution that responds to the problem or a question and an answer that responds to the question.

Procedural text: A text that conveys information in the form of directions for accomplishing a task. A distinguishing characteristic of such text is that it is composed of discrete steps to be performed in a strict sequence, with an implicit end product or goal.

Protagonist: The central character of a short story, novel, or narrative poem. The antagonist is the character who stands directly opposed to the protagonist.

Purpose: The specific reason or reasons for the writing. It conveys what the readers have to gain by reading the selection. Purpose is the objective or the goal that the writer wishes to establish.

Repetition: The deliberate use of any element of language more than once—sound, word, phrase, sentence, grammatical pattern, or rhythmical pattern.

Rhetoric: The art of using words to persuade in writing or speaking.

Rhetorical device: A technique used by writers to persuade an audience.

Rhyme: The repetition of sounds in two or more words or phrases which appear close to each other in a poem. *End rhyme* occurs at the end of lines; *internal rhyme*, within a line. *Slant rhyme* is approximate rhyme. A *rhyme scheme* is the pattern of end rhymes.

Rhythm: The regular recurrence and speed of sound and stresses in a poem or work of prose.

Sarcasm: The use of verbal irony in which a person appears to be praising something but is actually insulting it.

Satire: A piece of prose in which witty language is used to convey insults or scorn.

Sequence: A text structure in which ideas are grouped on the basis of order or time.

Setting: The time and place in which events in a short story, novel or narrative poem take place.

Simile: A comparison of two different things or ideas through the use of the words *like* or *as*.

Soliloquy: A long speech in which a character who is usually alone onstage expresses his or her private thoughts or feelings.

Sonnet: A fourteen-line lyric poem, usually written in iambic pentameter.

Stanza: A division of a poem, composed of two or more lines.

Style: The writer's characteristic manner of employing language.

Symbol: An object, person, place, or action that has both a meaning in itself and that stands for something larger than itself, such as a quality, an attitude, a belief, or a value.

Syntax: The arrangement of words and the order of grammatical elements in a sentence.

Tall Tale: An improbable, incredible or fanciful story.

Theme: The central message of a literary work. A literary work can have more than one theme. Most themes are implied rather than being directly stated. A literary theme is not the same as a topic.

Tone: The writer's or speaker's attitude toward a subject, character, or audience conveyed through the author's choice of words and detail. Tone can be serious, humorous, sarcastic, indignant, objective, etc.

Trait: A distinguishing feature, as of a person's character.

Understatement: A kind of irony that deliberately represents something as being much less than it really is; the opposite of hyperbole or overstatement.

Voice: The distinctive style or manner of expression of an author or of a character.

APPENDIX C

PASSAGE MAPPING PROCEDURES

PASSAGE MAPPING PROCEDURES

WHAT IS PASSAGE MAPPING AND WHY IS IT IMPORTANT?

Mapping is the process of analyzing a text in ways that identify both important information and the relations among important ideas in a text. Mapping procedures provide a basis for text selection and analysis, and for item development that is consistent with current knowledge about the role of various text properties in text processing. Literary and Informational texts require different types of maps, because the structure of information is fundamentally different between these two general text types as well as among many of the various subgenre that comprise these larger categories of text.

There are a number of important reasons to map texts, most of which derive from how the text mapping process assists those who are selecting texts and writing items to reflect on the relations between text ideas and their organization within a text. More specifically, there are at least four ways in which the process of constructing maps and the maps themselves are useful. First, maps are used to determine the appropriateness of a particular text for use as part of an assessment of reading comprehension—e.g., to determine if a text is sufficiently rich and/or coherent. Second, mapping identifies the important information in a text and how the information is presented and organized—i.e., how the ideas are connected; how it is organized; how it makes use of specific text features (e.g., use of illustrations or subheadings), and how the author crafts the ideas through the use of selected literary and rhetorical devices (e.g., metaphor, imagery, embedded question). Third, mapping focuses item writers on key ideas, their relations, and how this information is organized within a text. Mapping also serves as a guide in creating scoring rubrics for constructed response items.

A fourth reason to use maps is that they capture information that is both explicit and implicit in texts. This means that a given text could be mapped in more than one way. For example, two people might have different events identified in a story or state them differently. As a result, it is important that the mapping process involve discussion of initial map/s before revising and/or finalizing them. This discussion should occur before item development. In addition, the reading item development committee should review newly developed items and suggested rubrics for constructed response items in relation to the information on the map.

DESCRIPTIONS OF MAPS

The development of mapping procedures requires operational decisions about distinctions among types of texts and how they are parsed. These decisions have been driven by the need for clear and helpful guidelines for those who are selecting texts and developing items and scoring rubrics, rather than by the nuances of literary or linguistic theory. For the purposes of NAEP reading tests, Literary and Informational texts should be analyzed using three types of maps: narrative, non-narrative, and poetry. Narrative maps are used for any Literary texts that have a plot—i.e., some combination of problem, conflict, and resolution, and include a variety of texts such as tales, mysteries, and realistic and historical fiction. Non-narrative maps are used for

Literary and Informational texts that do not have a plot, most likely speeches, exposition, documents, and persuasive essays. In addition to the narrative and non-narrative maps, there are separate guidelines for analyzing poetry.

Narrative Map

The narrative map (see Figure 1 below) captures the structure of fiction—themes, plot structure, setting characters, along with elements of author's craft. Since themes operate on multiple levels, the narrative map identifies themes at two levels—in terms of both the specific events of the narrative (story level) and the more general concepts that run through the narrative (abstract level). This allows for theme questions along a continuum of explicitness within the narrative. The narrative maps are also designed to emphasize the interrelatedness among these elements. For example, rather than treating setting as merely a locational feature, it is related to the other elements of the plot and the themes. Similarly, the category of character provides information about how the characters are portrayed in the major events of the story, what their various roles are and how they function in relation to the themes. Sample narrative maps are provided at then end of this Passage Mapping section.

Figure 1--NARRATIVE MAP
(To be used with texts with a plot structure.)

TITLE:

GENRE:

STORY LEVEL THEME:

ABSTRACT LEVEL THEME:

PLOT:

Problem:

Conflict:

Resolution:

SETTING (and how it is connected to the themes and significant ideas in the text):

CHARACTER/S (traits that are connected to significant ideas in the text):

MAJOR EVENTS:

AUTHOR'S CRAFT:

VOCABULARY WORDS:

Non-narrative Map

The non-narrative map (see Figure 2) captures the organization of information in texts that do not have a plot structure, specifying the genre, dominant organizational patterns, and the central ideas. The non-narrative map also identifies major and supporting ideas, and the "organizing elements" that specify the relations between these ideas and the larger text. This structure allows for the identification of both multiple patterns of organization that exist within any given text and the dominant pattern/s of organization relative to a particular central idea. The analysis of major and supporting ideas in relation to the organization of the text, assists those who are selecting texts in determining the coherence of a passage and helps item writers generate questions about these relations. For example, in an argumentative selection, this mapping process helps the item writers develop questions that focus on the quality of the argument and the support provided by the author to substantive claims, rather than simply focusing on questions about the understanding of key arguments. In sum, the non-narrative map reflects the relations among organizational patterns, three levels of ideas (central, major, and supporting), the text features (e.g., subheadings, charts, and illustrations), and author's craft (e.g., figurative language and rhetorical devices) for a variety of texts including exposition, argumentation, and documents.

Whenever possible, it is important to adhere closely to the structure presented by the text, to ensure that the mapping is not based on a reconceptualization of the information by the individuals constructing the map. However, in cases where information is not presented in a logical order, it may be desirable for the map to reorder the information. For example, it is not uncommon for some magazine articles to present information in a nonlinear manner in an effort to make the information more interesting and appealing to the reader as in biographical sketches where events in an author's life may not be presented in chronological order. When this occurs, the map may not follow the linear sequence of the article but rather restructure the ideas so they flow in a more logical order. When completed the map should provide a logical flow of the ideas in the selection which are linked to one another through the organizing elements of the selection. Sample maps are provided at the end of this appendix for three types of Informational texts: exposition, argumentation, and document.

The non-narrative map will most likely be used for texts that fit the definition of Literary Non-fiction, because these texts rarely have a plot. These texts include biographical texts and personal essays. See the sample map for a biographical sketch at the end of this appendix. It is possible, however, particularly at grade 4, for a biographical text to have enough of a plot structure to require the use of the narrative map.

Figure 2--NON-NARRATIVE MAP
(To be used for any Literary and Informational texts that do not have a plot.)

TITLE:

GENRE:

CENTRAL IDEA:

DOMINANT ORGANIZATIONAL PATTERN/S:

MAJOR/SUPPORTING IDEAS AND ROLE IN TEXT ORGANIZATION:

Org. Element--_____:

Major Idea:
Supporting Idea/s:

Org. Element--_____:

Major Idea:
Supporting Idea/s:

Org Element--_____:

Major Idea:
Supporting Idea/s:

Org. Element—_____:

Major Idea:
Supporting Idea/s:

Org. Element—_____:

Major Idea:
Supporting Idea/s:

TEXT FEATURES:

AUTHOR'S CRAFT:

VOCABULARY WORDS:

Poetry Map
The poetry analysis (see Figure 3) captures the structure of a poem—theme, key events or ideas, text structures and features, author's craft, and word choice. When mapping poetry, it is important to pay attention to the author's style—the author's word choice, sentence structure, figurative language, and sentence arrangement, because all work together to establish mood, images, and the meaning of the poem. See the sample poetry map at the end of this appendix.

Figure 3—POETRY ANALYSIS

TITLE:

GENRE:

CONCRETE LEVEL THEME:

ABSTRACT LEVEL THEME:

SUMMARY:

TEXT STRUCTURE AND FEATURES:

AUTHOR'S CRAFT:

WORDS/PHRASES:

MAPPING PROCEDURES

Mapping begins by reading through the text carefully to determine if it has a plot structure. If the text has a plot, i.e., problem, conflict, and resolution, then use the narrative mapping procedures. For all other texts, except poetry, use the non-narrative mapping procedures.

Developing a Narrative Map

Once it has been determined that the Narrative map is most appropriate, proceed with the steps in the following stages (see the examples at the end of this appendix for further guidance).

Stage 1 of Narrative Mapping Procedure

Step 1: Begin by noting the title of the selection and the genre (e.g., fable, historical fiction, fantasy, etc.). Next, identify the problem, the conflict, and the resolution. These elements provide the basis for generating important themes at both the explicit level of the story and at a more abstract level.

- Problem: major initiating event around which the narrative is focused; this is most often introduced at the beginning of the narrative
- Conflict: forces, often two in number, that are in conflict within the narrative; these may be physical, psychological, social, ethical, etc., in nature
- Resolution/solution: how the conflict is resolved

Step 2: Use the elements of the plot--problem, conflict, and resolution—to identify important themes. Themes should be written at two levels—i.e., the story level and abstract level. A story level theme is based on the explicit events in the story, and an abstract theme captures generalizations that can be drawn from a story, but do not reference specific events or characters from the story. For example, in the fable about the tortoise and the hare, a story level theme might be that a rabbit loses the race to the tortoise because the rabbit is overconfident. An abstract theme might be that arrogance and judgments based on superficial qualities often lead to poor decisions.

Stage 2 of Narrative Mapping Procedure

Step 3: Describe the setting (location and importance to theme) and the characters (traits and functions in the narrative).

- List the physical setting/s how they are connected to the theme/s and important ideas of the narrative
- List important physical and psychological traits of each of the major characters
- Identify the function of each of the major characters in relation to the plot and theme of the narrative. When traits are listed, they should be significant characteristics that capture the complexity of well-rounded characters. They can include physical, psychological, and emotional dimensions of a character. In addition, accompanying the trait should be a sentence or two explanation of how the trait is connected to the plot of the narrative.

Step 4: List the events of the story essential to the themes, conflicts, or resolution of the conflict. List the major actions that take place.

Step 5: Determine appropriateness of the passage for use on NAEP—i.e., is it rich enough to support the development of 10-12 items. If so, move on to Stage 3

Stage 3 of Narrative Mapping Procedure

Step 6: During the initial reading of the narrative selection, identify elements of the author's craft that appear in the story. Once the other elements of the map are in place, decide whether the literary and rhetorical devices you noted are important enough to be included in the map.

Step 7: Use the vocabulary mapping procedures to identify words for vocabulary items.

Stage 4 of Narrative Mapping Procedure

After items are written, chart their distribution on the Item Classification Matrix

Determine if coverage is adequate and appropriate for the grade level. Check the items against the map to ascertain that the items reflect important ideas in the text and to serve as a final check that the map has covered all important ideas.

Write new items if needed, referring to information from analysis of the passage.

Developing a Non-Narrative Map

Once it has been determined that the Non-narrative map is the most appropriate, proceed with the steps in the following four stages. Examples of Non-narrative maps are provided at the end of this appendix.

Note that when mapping procedural texts, a less extensive version of the Non-Narrative may be used. Since some procedural texts will have subheadings, these can be used in identifying the organizational elements. In addition, the supporting details are generally not specified because many of them may refer to specific steps or detailed directions that do not need to be captured in the map. It is also unlikely that author's craft will be a consideration with procedural texts and documents. See the sample at the end of this appendix.

Stage 1 of Non-Narrative Mapping Procedure

Step 1: Read the entire piece carefully to get a general understanding of its content and purpose. Identify the type of text using categories from the Framework, for example:

- Informational: Expository (news article)
- Informational: Procedural (directions)
- Informational: Argumentation/persuasive (simple persuasive essay
- Literary: Literary Non-fiction (biographical sketch).

Step 2: Write a central purpose statement that synthesizes the key ideas and supporting ideas in the selection into a sentence or two. Use words that reflect the Dominant Organizational pattern whenever possible (e.g., to describe, to argue, to provide chronology).

Step 3: Identify the Dominant Organizational pattern or patterns (e.g., persuasion, argumentation, description, problem, solution, theory, and evidence). This is the organizational pattern that is dominant for the central purpose that has been identified.

Stage 2 of Non-Narrative Mapping Procedure

Step 4: Read the first paragraph or continuous section of text carefully.

- Identify major ideas and write a brief description (one or two sentences) next to the major ideas subheading.
- Identify the role that this major idea plays in terms of the overall organization of the text—e.g., description, chronology, reason/result, etc.—and write a key word descriptor next to the Organizing Element heading. For example, in a persuasive selection the organizing elements might be argument, evidence, and counter argument. In a biographical sketch they might be chronology, overview, and description. The organizing elements are directly connected to the major ideas associated with the logical flow of ideas in the selection.

Step 5: Identify any relevant ideas that support this major idea and write them next to the supporting idea/s subheading. There should be a clear connection between the major ideas and the supporting ideas. Subtitles may or may not be useful depending on how well they describe the content in the section of text they refer to and/or how important the section is to the overall purpose of the text. Supporting ideas are likely to be several sentences in length.

Step 6: Review the map to ascertain whether the original determination of purpose was correct; adjust if needed.

- Review the map carefully to be sure that the organizational elements are accurate and consistent with the Dominant Organizational pattern and there is a clear logical flow to the ideas. Make adjustments as necessary.
- Eliminate the text from consideration if the purpose is unclear or if the text lacks coherent organization structures

Stage 3 of Non-Narrative Mapping Procedure

Step 7: Determine the extent to which text features such as charts or bold print have been used to support readers' comprehension of the text. If they appear to have been used for this purpose, note them on the map along with the purpose they serve.

Step 8: Identify any elements of author's craft that seem particularly important. State how interpretation of these elements can help readers understand the important ideas or concepts presented in the text.

Step 9--Use the vocabulary mapping procedures to identify words for vocabulary items.

Stage 4 of Non-Narrative Mapping Procedure

After the items are written, chart their distribution on the Item Classification Matrix.

Determine if coverage is adequate and appropriate for the grade level. Check the items against the map to ascertain that the items reflect important ideas in the text and to serve as a final check that the map has covered all important ideas. Write new items if needed.

Analyzing Poetry

Stage 1 of the Poetry Analysis

Step 1: Note the title of the poem and the genre, e.g., free verse, ode, or lyric poem. Begin the analysis with a close reading of the poem to identify the theme/s. Themes are written at two levels: the concrete level of the poem (see earlier section on concrete vs. story level) and a more abstract level that captures important generalizations. Also, pay attention to the text structure and features, author's craft, and word choices that seem important.

Step 2: Use the themes of the poem to develop a summary to capture the story, situation, or vivid images presented by the speaker of the poem. The summary can include descriptions of people, places, a mood, an emotion, an attitude, and images. Use line references when noting specific aspects of the poem.

Step 3: Determine the appropriateness of the poem for use on NAEP—i.e., Is it rich enough to support the development of 10 to 12 items? If the poem is short and is intended to be paired with another text, will the poem support development of 4-5 items? If so, move to Stage 2.

Stage 2 of the Poetry Analysis

Step 4: Determine the extent to which structural features in the poem, e.g., stanza, verse, rhyme scheme and rhythm patterns, are used to present or reinforce the events in the poem. If they appear to have been used for this purpose, note them as part of the analysis along with the purpose they serve.

Step 5: Decide whether the elements of author's craft that were noted previously are important enough to be included.

- When analyzing poems, pay particular attention to the images, similes, and metaphors that are used by the author.

- Note the use of word choice or phrasing in the poem. Pay particular attention to connotation and denotation of words. Decide whether they are important enough to be included in the analysis.

Stage 3 of the Poetry Analysis

After the items are written, chart their distribution on the Item Classification Matrix. Determine if coverage is adequate and appropriate for the grade level. Check the items against the map to ascertain that the items reflect important ideas in the poem and to serve as a final check that the map has covered all important ideas. Write new items if needed, referring to information from analysis of the poem.

Note: These passage procedures will be used to develop passage maps beginning with the 2011 NAEP Reading assessment.

Sample Passage Maps

For the reading passages, refer to released NAEP passages at nces.ed.gov/nationsreportcard, under "Sample Questions." See the Chapter 8 of the Specifications for additional released NAEP reading passages.

NARRATIVE MAP
(To be used for texts with a plot structure)

GRADE 4

TITLE: How The Brazilian Beetles Got Their Coats

GENRE: Tale

STORY LEVEL THEME: A rat lost what he wanted because he made faulty assumptions about a beetle based on superficial information

ABSTRACT LEVEL THEME: Arrogance and judgments based on superficial qualities often lead to poor decisions

PLOT:

Problem: An arrogant rat makes a superficial judgment about his opponent in a race

Conflict: The outcome of the race between the arrogant, boastful rat and the polite, restrained beetle

Resolution: Not one to boast, the beetle is confident in her abilities and wins a race that results in her getting a new colorful coat

SETTING (and how it is connected to the themes and significant ideas in the text): The story takes place in the jungles of Brazil, which provides both the geographic setting for a race between a beetle and a rat, and the context for colors of the beetle's coat.

CHARACTER/S (traits that are connected to significant ideas in the text):

Rat

Arrogant—thinks he is better than a beetle because he is much faster than her

Over confident—because he can run faster than the beetle he believes he will win a race against her

Superficial—judges people on their superficial qualities

Surprised—learns that the beetle wins the race because she has wings and can fly

Beetle

Polite—she is taught by her mother to always be polite and does not respond to the rat's critical comments about her nor does she make critical remarks about him

Confident—she is does not boast about her wings when the rat criticizes her for being slow. In fact, she tells the parrot what colors she would like her new coat to be

Clever—when the parrot is explaining the rules of the race, she does not say anything about her wings

Respectful—she accepts the rat for who he is and does not criticize him when she wins the race

MAJOR EVENTS:

1. A big gray rat criticized a little brown beetle for how slowly she was walking
2. The rat demonstrates his arrogance by demonstrating just how fast he is.
3. Despite his criticism of her, the beetle politely acknowledges that the rat is a fast runner.
4. The beetle is content with who she is and does not boast about her own talents.
5. A brightly colored parrot observes the conversation and proposes a race to the top of a palm tree on a cliff with a prize of a brightly colored coat to the winner. The only rule in the contest is getting to the top of the cliff first.
6. Both the rat and the beetle accept the challenge and begin the race.
7. The rat is overly confident that he will win the race and criticizes the beetle for thinking she can win at all.
8. The rat starts off very fast and thinks he should slow down because there is not way the beetle can catch him; however, his heart tells him to keep running as fast as he can.
9. When the rat arrives at the top of the cliff, he is surprised that the beetle is already there and asks her how she got there first.
10. The beetle shows the rat her wings
11. The rat said he didn't realize she could fly, and the parrot tells the rat that in the future he should not judge others by their looks alone
12. The beetle wins the prize and chooses a coat of green and gold to match the colors of the parrot, the palm trees, the mangoes, and the golden sunshine on the green hills
13. Beetles in Brazil continue to wear a coat of green and gold, while rats wear a dull gray coat

AUTHOR'S CRAFT:

Foreshadow—the beetle who is not boastful believes she can win the race so this is a clue to the reader that there is more to her than the rat has observed.

Metaphor—concealed wings is a metaphor for you need to look beyond the surface to fully understand the capabilities of a person.

Imagery and Symbolism—dull gray, the color of the rat today, is symbolic of the type of person the rat turned out to be vs. the colorful coat of the beetle, which represents all that is wonderful about the Brazilian jungle.

Tone—there is a shift in tone near the end of the tale—the rat remarks, "`I did not know you could fly,' said the big gray rat in a subdued little voice."

Rhetorical Question—"What is the use of hurrying." is used to reinforce the overconfidence of the rat.

Rhetorical Device—purpose of the first paragraph is to suggest the beetle does get its colorful coat because it already has one and then transitions into, "Once upon a time…

Word Choice—dull has a double meaning: the color of the rat's coat and the lack of insight he possesses.

VOCABULARY WORDS

scornfully
boasts
accomplishments
gaunt
subdued
concealed

NARRATIVE MAP
(To be used for texts with a plot structure)

GRADE 8

TITLE: Thank You M'am

GENRE: Literary--Realistic Fiction

STORY LEVEL THEME: A woman's tough, but sympathetic, response to a teenage boy who tries to steal her purse causes the boy to change his behavior/attitude

ABSTRACT THEME: Kindness, trust, and generosity are used to teach a young boy a lesson about right and wrong

PLOT:

Problem: Roger attempts to steal Mrs. Jones' purse in hopes of getting money to buy a pair of shoes he cannot afford to purchase

Conflict: Will Roger run or will he let Mrs. Jones help him

Resolution: Roger reciprocates the trust and caring demonstrated by Mrs. Jones, and is given a chance to change his life

SETTING (and how it is connected to the themes and significant ideas in the text): Urban area and small apartment where everything is in view provide a woman with an opportunity to help a young boy to see the wrongness of his actions

CHARACTER/S (traits that are connected to significant ideas in the text):

Mrs. Luella Bates Washington Jones /Woman

Trusting—she leaves her purse where the boy could take it if he wanted to; provides him with a choice about going to the store with her money to buy food or eating what she has on hand

Honest—she is straightforward with the boy and never tries to deceive him

Caring—she does not turn him over to the police, gives him food and money

Stern—establishes that she is in control of the situation and directs the boy when she has him in her physical and psychological grasp

Person of conviction—wants to teach the boy right from wrong and does not give up easily

Virtuous—she shows moral goodness, decency, and possesses a sense of justice toward the boy throughout the story

Insightful—she is able to judge Roger's nature and uses it as away to build trust with him

Roger/Boy

Thief
Frail and unkempt

Observant—he quickly recognizes who is in control and knows he dare not leave the apartment; he checks out the apartment and knows that she is not alone

Cautious—once he is in Mrs. Jones' apartment he is very aware of his circumstance

Respectful—once he realizes his fate is in Mrs. Jones's hands, he is polite and cooperative

MAJOR EVENTS:

1. Roger attempts to steal a purse of an older woman but is thwarted in his attempt by a woman who is not easily taken advantage of..
2. The woman quickly establishes her physical and emotional control over the boy.
3. She is able to judge the character of the boy and use her insights and experience to build trust between them.
4. Mrs. Jones probes Roger to find out about Roger's background and his motive for attempting to steal her purse but it is done in a caring and sensitive manner.
5. Mrs. Jones takes him home. When she makes him wash his face he considers running, but does as he's told instead.
6. Mrs. Jones speculates that Roger tried to steal because he was hungry, but he tells her that he wanted a pair of blue, suede shoes.
7. Through her experience and wisdom, she does not set herself up as a morally superior person and tells Roger everyone does things he or she is not proud of.
8. Roger can be respectful—he refers to her as M'am and he does not want her to think he is going to continue to steal from her.
9. She demonstrates her control over the boy on several different levels, i.e., physical, social, and moral.
10. She does not give up on Roger and acts in a way that demonstrates she cares about his future, e.g., she gives him money, feeds him, and sends him away with food.
11. She builds trust between herself and the boy throughout the story, e.g., trusting him by leaving her purse where he could easily take it and run out the door.
12. She uses the gift of money to hopefully teach him a lesson about controlling his desires to steal from others.

AUTHOR'S CRAFT:

Tone: one of authority in the beginning changing to one of concern

Rhetorical Devices

Use of italics
Significance of the title and use of M'am throughout
Use of slang diction
Use of "run" image throughout

VOCABULARY WORDS:
permit
release
frail
presentable
barren

NON-NARRATIVE MAP
(To be used for any Literary and Informational texts that do not have a plot)

GRADE 4

TITLE: Watch Out for Wombats

GENRE: Informational--Exposition

CENTRAL IDEA: To describe the physical appearance, eating habits, reproductive characteristics of the Australian wombat and compare it to other mammals and marsupials

DOMINANT ORGANIZATIONAL PATTERN/S: Description, comparison

CENTRAL PURPOSE: To describe the physical appearance, eating habits, reproductive characteristics of the Australian wombat and compare it to other mammals and marsupials.

MAJOR/SUPPORTING IDEAS AND ROLE IN TEXT ORGANIZATION:

Org. Element--Introduction

Major Idea: Road signs in Australia warning motorists to watch out for wombats

make those who are not familiar with these animals curious about them

Supporting Idea/s: Looking for them midday is likely to be unsuccessful, because they are nocturnal

Org. Element--Description

Major Idea: Two types of wombats in Australia: hairy-nosed and coarse-haired

Supporting Idea/s: they live in different parts of Australia
both have soft brown fur, short ears and thick bodies
resemble North American badgers
hairy-nosed is smaller and has pointier ears
they are shy and gentle

Org. Element--Comparison

Major Idea: There are similarities and differences between wombats and koalas

Supporting Idea/s: both have strong forelimbs and powerful claws
koala uses claws to cling to high tree branches, but wombat uses them to dig large underground burrows
burrows are usually 9-15 ft across, but can be as big as 90 ft
wombat builds nest of bark in burrow for sleeping

Org. Element—Characteristic

Major Idea: Wombats are vegetarians

Supporting Idea/s: uses its claws to tear up grasses and roots for food
teeth, which grow throughout its life, are sharp and good for cutting

Org. Element—Comparison

Major Idea: Wombats reproduction is similar to other mammals and marsupials

Supporting Idea/s: like other marsupials they have only one baby at a time, usually in winter (May-July)
a baby wombat is called a joey
baby wombats are born at an early stage of development (only 1 inch at birth) and use their forelimbs to crawl along their mother's underside to get into her pouch
once inside the pouch, the joey finds a nipple and, like other mammals, is nourished by its mother's milk
the joey stays in the pouch for next four months, grows rapidly,

and emerges when it is able to survive on its own

Org. Element—Comparison

Major Idea: Most marsupials pouches open upward, but wombat and koala pouches open downward

Supporting Idea/s: a strong muscle keeps the pouch tightly closed so young wombat or koala doesn't fall out
downward facing pouch means less dirt is less likely to get inside when wombat is burrowing

Org. Element— Problem/Solution

Major Idea: Wombats are endangered and in some places becoming extinct; however animal reserves are being set up

Supporting Idea/s: more and more people are moving into territory where they live destroying their homes and foods but with reserves being set up they prosper again

TEXT FEATURES: Wombat road signs

AUTHOR'S CRAFT:

Use of Wombat crossing signs to engage reader's interest
Literal interpretation of "Watch Out for Wombats" as a vehicle to provide information that wombats are nocturnal
Using human behavior to make the point that wombats carry their young with them

VOCABULARY WORDS:

loomed
nourished
advantage
plentiful
prosper

NON-NARRATIVE MAP
(To be used for any Literary and Informational texts that do not have a plot)

GRADE 4

TITLE: Dr. Shannon Lucid: Space Pioneer

GENRE: Literary—Literary Nonfiction—Biographical Sketch

CENTRAL PURPOSE: To describe significant events in Shannon Lucid's life that led her to achieve her life's goal to become a space explorer and inspire others to achieve their dreams by taking every opportunity that comes their way.

DOMINANT ORGANIZATIONAL PATTERN/S: Chronology, description

MAJOR/SUPPORTING IDEAS AND ROLE IN TEXT ORGANIZATION:

Org. Element—Overview

> **Major Idea:** While Shannon Lucid dreamed of becoming a space explorer, it was not always easy
>
> **Supporting Idea/s**: She thought it was too late to be a pioneer; however, she learns about Robert Goddard and realizes she can explore space. She eventually becomes the second woman in space and holds the American record for the longest time in space.

Org. Element—Chronology 1

> **Major Idea:** Lucid learns the spirit of adventure from her parents and pursues her interest in exploring space
>
> **Supporting Idea/s**: She loves science and earns degrees in chemistry and biochemistry; applies to space program in 1959 but they do not accept women at that time; she persists and is finally admitted in 1978

Org. Element—Chronology 2

> **Major Idea:** In 1996, Lucid became second American astronaut to live aboard Russian space station Mir
>
> **Supporting Idea/s**: She was part of a program to study how long-term travel in space affects the human body

Org. Element—Problem

Major Idea: On Mir, she had to cope with weightlessness, a small living space, and a diet of dehydrated foods

Supporting Idea/s: Her predecessor had lost weight in part because he missed familiar foods, and had been lonely

Org. Element—Solution

Major Idea: Lucid made provisions to have food she liked and to stay in touch with family. She also maintained her weight and loss only a little strength in her bones and muscles sufficiently that she was able to walk off space shuttle when she returned rather than being carried

Supporting Idea/s: Lucid took along favorite foods and supply ships delivered more; she stayed in touch with her family through daily e-mail and "visits" on radio and TV; to stay fit, she spent many hours exercising on a treadmill and stationary bicycle

Org. Element—Description

Major Idea: Lucid did research while on MIR

Supporting Idea/s: Studied the effects of weightlessness on plants and animals

Org. Element—Description

Major Idea: Lucid's companions on Mir were two Russian cosmonauts, both named Yuri

Supporting Idea/s: Even though she and the cosmonauts grew up when U.S. and Russia were enemies, they became friends

Org. Element—Description

Major Idea: Lucid enjoyed her time in space

Supporting Idea/s: She liked observing earth and watching seasons change; missed bookstores, desserts, spending time with daughters, etc.

TEXT FEATURES:

Photographs of the space station and Shannon Lucid
Subheadings

AUTHOR'S CRAFT:

Embedded questions for rhetorical purposes

VOCABULARY WORDS:

adaptation
isolated
mortally
inspire
opportunity

NON-NARRATIVE MAP

(To be used for any Literary and Informational texts that do not have a plot)

GRADE 8

TITLE: Ellis Island

GENRE: Informational—Exposition—Informational Trade Book

CENTRAL IDEA: To provide a historical account of immigrants told in the words of immigrants who can to the US through Ellis Island between 1892 and 1954

DOMINANT ORGANIZATIONAL PATTERN/S: Description, Cause/Effect

MAJOR IDEAS:

Org. Element—Description/Introduction

Major Idea: Between 1892 and 1954, Ellis Island was the "doorway to America" for 17 million people

Supporting Idea/s: not everyone was welcome
"land of the free" was not so free to everyone

Org. Element—Cause

Major Idea: Immigrants came from Europe to escape oppression/poverty and/or seek a better life

Supporting Idea/s: first hand accounts from a woman escaping Turkish oppression in Armenia, and a man from the Ukraine seeking opportunities

offered by U.S.

Org Element—Effect/Problem

Major Idea: Those who wanted to immigrate had to endure great hardship to travel to U.S.

Supporting Idea/s: they had to contend with border guards, thieves, dishonest immigration agents, and bad conditions on the ships they crossed on.
once they saw NY and Statue of Liberty—they felt it was worth it

Org. Element—Problem

Major Idea: Once they arrived, many immigrants experienced problems entering the U.S.

Supporting Idea/s: Sent to Ellis Island where they were screened

Org. Element—Cause

Major Idea: As the number of immigrants increased, U.S. worried that immigrants would take away jobs and wages from American workers

Supporting Idea/s: Peak immigration was 1907 with 1,285,349 immigrants

Org. Element—Effect

Major Idea: Medical examinations and literacy tests became more difficult. However, medical examinations caused the most concerns.

Supporting Idea/s: deported those who could not pass literacy and/or medical exams
only 3% of those who arrived at Ellis Island between 1892 and 1954, when it closed, were deported

Org. Element—Effect

Major Idea: The Immigration Act of 1924 set a quota of 164,000 immigrants per year

Supporting Idea/s: those who immigrated to the U.S. during this period found it to be the land of opportunity:

TEXT FEATURES:
Subheadings, illustrations, use of italics to set off quotations from past immigrants
Illustration of "cattle-pen-like" method of processing

AUTHOR'S CRAFT:

Use of first hand accounts to illustrate the points about the immigrant experience in general and on Ellis Island
Use of a doorway to America/doorway metaphor

VOCABULARY WORDS:

opportunity
oppressive
flee
turmoil
threat
quota

NON-NARRATIVE MAP

(To be used for any Literary and Informational texts that do not have a plot)

GRADE 12

TITLE: Address to the Broadcasting Industry

GENRE: Informational—Argumentative—Speech

CENTRAL IDEA: In his address delivered in May 1961 before the National Association of Broadcasters Newton Minow, attempted to persuade the organization that television had become a "vast wasteland," because programming decisions were heavily influenced by ratings. He suggested that because of its responsibility to the public, the television industry must act in a responsible manner to resolve the problem by providing the public with a wider range of choices.

DOMINANT ORGANIZATIONAL PATTERN/S: Argument, claims, evidence, enumeration

MAJOR/SUPPORTING IDEAS AND ROLE IN TEXT ORGANIZATION:

Org Element—Overview 1

Major Idea: A brief biographical overview provided a context for reading the speech delivered by Newton Minow

Supporting Ideas: President Kennedy appointed Minow chairman of the Federal Communications Commission
in the speech Minow delivered while chairman of the FCC to the National Association of Broadcasters, he indicted the television industry for its poor selection of programs.

Org Element—Claim 1

Major Idea: Television is a powerful voice in American and it is the duty of the industry to set high standards in order to promote high quality programming

Supporting Idea/s: the television industry has grown and it needs to assume a leadership role much like newspapers and magazines did years ago

we now live in the television age and the overall impact television will have on society is yet to be determined

Org Element—Claim 2

Major Idea: Television programming has its good qualities

Supporting Idea/s: Minow supports his claim by offering examples of specific programs at the time

such programs represent television at its best and they enrich the lives of Americans

Org Element—Claim 3

Major Idea: Television has its bad qualities and because of them, he considers it a "vast wasteland"

Supporting Idea/s: He supports his claim by offering evidence of programs like formula comedies, game shows, and endless commercials and so much of it is bad because of the competition for higher ratings

Org Element—Problem/Solution

Major Idea: Because it has a responsibility to the public, the television industry, which is driven too much by the desire of high ratings and sponsors, needs to rethink its perspective on ratings

Supporting Idea/s: Parents would not approach raising children by giving them what they preferred. Newspapers do not treat their readers this way either. This is in contrast to how the television industry approaches the programs it offers its viewers—they offer them what they think they want. A problem with interpreting the ratings the way the television industry does is that it that the ratings do not tell what the public might watch. The people's taste is not as low as the television industry assumes. Of particular importance is that ratings should not be used to make decisions about children's programming, because children are spending as much time watching television as in the schoolroom.

Org Element—Perspective:

Major Idea: Minow clarifies his position by calling for a balance in the types of programming that is offered, because the public is made up of many interests

Supporting Idea/s: The public interest is made up of many different interests and television must serve them all. Programs must stimulate and inform. Therefore, a wider range of choices must be provided.

Org Element—Description

Major Idea: Minow delineates the principles that he believes should be used to guide decisions about television programming

Supporting Idea/s: Minow provides examples of six principles he uses as chairman of the FCC to guide his decisions. Some of these are: (1) people own the air, (2) don't dwell on mistakes of the past; (3) he believes in the free enterprise system, and (4) opposed to government censorship.

TEXT FEATURES

Bolded introduction
Use of ellipsis
Sequence words

AUTHOR'S CRAFT

Tone—serious, professional
Sarcasm
Rhetorical Devices, e.g., questions and phrases (Repeat—let that sink in), analogies
Uses numbered principles to bring message home
Metaphor of wasteland
Use of contrasting words: destroy/rebuild; debase/enrich

VOCABULARY WORDS:

debase
cajoling
susceptible
obligations
squandering
precedent

NON-NARRATIVE MAP
(To be used for any Literary and Informational texts that do not have a plot)

GRADE 12

TITLE: MetroGuide

GENRE: Informational—Procedural—Directions

CENTRAL PURPOSE: To summarize metro policies and provide all the information one would need to ride the metro

DOMINANT ORGANIZATIONAL ELEMENT/S: Description, explanation

MAJOR IDEAS:

Org. Element—Introduction

Major Idea: Metro is easy to use, will save money, and enhance travel in DC

Supporting Idea/s:

Org. Element—Description

Major Idea: Basic bus fares

Supporting Idea/s: footnotes for special situations

Org Element—Description

Major Idea: Basic policies and fares for using Metrorail

Supporting Idea/s:

Org. Element—Description

Major Idea: Basic information on transfers

Supporting Idea/s:

Org. Element—Description

Major Idea: Basic information on where and how to purchase passes

Supporting Idea/s: Alternative methods of purchase

Org. Element—Description

Major Idea: Replacement policy

Supporting Idea/s: contact information

Org. Element—Summary

Major Idea: Table detailing type of pass, cost, period of use, and description for metrobus and metrorail passes

Supporting Idea/s:

TEXT FEATURES:
Headings, subheadings, illustrations, shaded boxes, table

AUTHOR'S CRAFT:

VOCABULARY WORDS:
responsible
replacement
retain
disruption
eligible

POETRY ANALYSIS

GRADE 8

TITLE: The Fish

GENRE: Free Verse

CONCRETE LEVEL THEME: The speaker's respect and admiration for an old fish he has caught grows as he observes it more closely and sees evidence of its struggles to survive and decides to let it go

ABSTRACT LEVEL THEME: The physical and psychological struggles in life may leave scars but they also contribute to the wisdom that comes from them.

SUMMARY OF KEY EVENTS:

1. The speaker describes a "tremendous" old fish that hold along side the boat and is surprised that it did not fight (1-5).
2. The speaker begins describing the physical appearance of the fish both externally and internally (e.g., skin, gills, barnacles that cling to the fish, entrails, big bones, little bones, 9-34)
3. The speaker attempts to learn more about the fish by looking into eyes which he notes are bigger than his and old and scratch and effect what is seen "old scratched isinglass that don't return his stare (34-42).
4. The speaker can tell from the other fish hooks that are still in its mouth that fish has been around a long time and has fought hard to survive (50-64)
5. The speaker gains respect for fish and see him in a new light (rainbow) and decides to let him go (65-75)

AUTHOR'S CRAFT:

1. Simile—skin is like old wallpaper; like brown roses stained with age; lip is weapon like (50); like medals with their ribbons (61) recognizes his prowess as survivor.
2. Imagery—the images of the fish become deeper and more insight as the poem develops. At first there is the image of the aging fish described from a physical sense. Then you get the image of a fish who is a survivor (the fishhooks) and then it the poem ends with an imagine of a rainbow.
3. Symbolism—rainbow that symbolizes his freedom, the pot of gold at the end of the rainbow. The rainbow comes at the end of the story. The fish also symbolizes the impact of one who struggles through life and the physical and psychological scares it leaves.

WORDS/PHRASES:

1. eyes that don't look back—sense of hopelessness (42)
2. Sullen face (45)
3. Five-haired beard of wisdom (63)
4. He stares and stares and victory filled the little boat (65-70)
5. Repeat rainbow 3 times before deciding to release the fish (75)

APPENDIX D

NAEP ITEM DEVELOPMENT AND REVIEW POLICY STATEMENT

Adopted: May 18, 2002

NATIONAL ASSESSMENT GOVERNING BOARD

NAEP ITEM DEVELOPMENT AND REVIEW

POLICY STATEMENT

It is the policy of the National Assessment Governing Board to require the highest standards of fairness, accuracy, and technical quality in the design, construction, and final approval of all test questions and assessments developed and administered under the National Assessment of Educational Progress (NAEP). All NAEP test questions or items must be designed and constructed to reflect carefully the assessment objectives approved by the National Assessment Governing Board. The final assessments shall adhere to the requirements outlined in the following Guiding Principles, Policies and Procedures for NAEP Item Development and Review.

The Governing Board's Assessment Development Committee, with assistance from other Board members as needed, shall be responsible for reviewing and approving NAEP test questions at several stages during the development cycle. In so doing, the Guiding Principles, Policies and Procedures must be adhered to rigorously.

Introduction

The *No Child Left Behind Act of 2001* (P.L. 107-110) contains a number of important provisions regarding item development and review for the National Assessment of Educational Progress (NAEP). The legislation requires that:

- "the purpose [of NAEP] is to provide…a fair and accurate measurement of student academic achievement;"

- "[NAEP shall]…use widely accepted professional testing standards, objectively measure academic achievement, knowledge, and skills, and ensure that any academic assessment authorized….be tests that do not evaluate or assess personal or family beliefs and attitudes or publicly disclose personally identifiable information;"

- "[NAEP shall]…only collect information that is directly related to the appraisal of academic achievement, and to the fair and accurate presentation of such information;"

- "the Board shall develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards;"

- "the Board shall have final authority on the appropriateness of all assessment items;"

- "the Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological;" and

- "the Board shall develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public."

Given the importance of these mandates, it is incumbent upon the Board to ensure that the highest standards of test fairness and technical quality are employed in the design, construction, and final approval of all test questions for the National Assessment. The validity of educational inferences made using NAEP data could be seriously impaired without high standards and rigorous procedures for test item development, review, and selection.

Test questions used in the National Assessment must yield assessment data that are both valid and reliable in order to be appropriate. Consequently, technical acceptability is a necessary, but not sufficient, condition for judging the appropriateness of items. In addition, the process for item development must be thorough and accurate, with sufficient reviews and checkpoints to ensure that accuracy. The Guiding Principles, Policies, and Procedures governing item development, if fully implemented throughout the development cycle, will result in items that are fair and of the highest technical quality, and which will yield valid and reliable assessment data.

Each of the following Guiding Principles is accompanied by Policies and Procedures. Full implementation of this policy will require supporting documentation from the National Center for Education Statistics (NCES) regarding all aspects of the Policies and Procedures for which they are responsible.

This policy complies with the documents listed below which express acceptable technical and professional standards for item development and use. These standards reflect the current agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.

Standards for educational and psychological testing. (1999). Washington, DC: American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME).

Code of fair testing practices in education. (1988). Washington, DC: Joint Committee on Testing Practices.

National Center for Education Statistics (NCES) Statistical Standards, DRAFT, February 2002.

GUIDING PRINCIPLES – ITEM DEVELOPMENT AND REVIEW POLICY

Principle 1

NAEP test questions selected for a given content area shall be representative of the content domain to which inferences will be made and shall match the NAEP assessment framework and specifications for a particular assessment.

Principle 2

The achievement level descriptions for basic, proficient, and advanced performance shall be an important consideration in all phases of NAEP development and review.

Principle 3

The Governing Board shall have final authority over all NAEP test questions. This authority includes, but is not limited to, the development of items, establishing the criteria for reviewing items, and the process for review.

Principle 4

The Governing Board shall review all NAEP test questions that are to be administered in conjunction with a pilot test, field test, operational assessment, or special study administered as part of NAEP.

Principle 5

NAEP test questions will be accurate in their presentation and free from error. Scoring criteria will be accurate, clear, and explicit.

Principle 6

All NAEP test questions will be free from racial, cultural, gender, or regional bias, and must be secular, neutral, and non-ideological. NAEP will not evaluate or assess personal or family beliefs, feelings, and attitudes, or publicly disclose personally identifiable information.

POLICIES AND PROCEDURES FOR GUIDING PRINCIPLES

Principle 1

NAEP test questions selected for a given content area shall be representative of the content domain to which inferences will be made and shall match the NAEP assessment framework and specifications for a particular assessment.

Policies and Procedures

1. Under the direction of the Board, the framework for each assessment will be developed in a manner that defines the content to be assessed, consistent with NAEP's purpose and the context of a large-scale assessment. The framework development process shall result in a rationale for each NAEP assessment, which delineates the scope of the assessment relative to the content domain. The framework will consist of a statement of purpose, assessment objectives, format requirements, and other guidelines for developing the assessment and items.

2. In addition to the framework, the Board shall develop assessment and item specifications to define the: a) content and process dimensions for the assessment; b) distribution of items across content and process dimensions at each grade level; c) stimulus and response attributes (or what the test question provides to students and the format for answering the item); d) types of scoring procedures; e) test administration conditions; and f) other specifications pertaining to the particular subject area assessment.

3. The Board will forward the framework and specifications to NCES, in accordance with an appropriate timeline, so that NCES may carry out its responsibilities for assessment development and administration.

4. In order to ensure that valid inferences can be made from the assessment, it is critical that the pool of test questions measures the construct as defined in the framework. Demonstrating that the items selected for the assessment are representative of the subject matter to which inferences will be made is a major type of validity evidence needed to establish the appropriateness of items.

5. A second type of validity evidence is needed to ensure that NAEP test items match the specific objectives of a given assessment. The items must reflect the objectives, and the item pool must match the percentage distribution for the content and cognitive dimensions at each grade level, as stated in the framework. Minor deviations, if any, from the content domain as defined by the framework will be explained in supporting materials.

6. Supporting material submitted with the NAEP items will provide a description of procedures followed by item writers during development of NAEP test questions. This description will include the expertise, training, and demographic characteristics of the groups. This supporting material must show that all item writing and review groups have

the required expertise and training in the subject matter, bias, fairness, and assessment development.

7. In submitting items for review by the Board, NCES will provide information on the relationship of the specifications and the content/process elements of the pool of NAEP items. This will include procedures used in classifying each item.

8. The item types used in an assessment must match the content requirements as stated in the framework and specifications, to the extent possible. The match between an objective and the item format must be informed by specifications pertaining to the content, knowledge or skill to be measured, cognitive complexity, overall appropriateness, and efficiency of the item type. NAEP assessments shall use a variety of item types as best fit the requirements stated in the framework and specifications.

9. In order to ensure consistency between the framework and specifications documents and the item pools, NCES will ensure that the development contractor engages a minimum of 20% of the membership of the framework project committees in each subject area to serve on the item writing and review groups as the NAEP test questions are being developed. This overlap between the framework development committees and the item developers will provide stability throughout the NAEP development process, and ensure that the framework and specifications approved by the Board have been faithfully executed in developing NAEP test questions.

Principle 2

The achievement level descriptions for basic, proficient, and advanced performance shall be an important consideration in all phases of NAEP development and review.

Policies and Procedures

1. During the framework development process, the project committees shall draft preliminary descriptions of the achievement levels for each grade to be assessed. These preliminary descriptions will define what students should know and be able to do at each grade, in terms of the content and process dimensions of the framework at the basic, proficient, and advanced levels. Subsequent to Board adoption, the final achievement level descriptions shall be an important consideration in all future test item development for a given subject area framework.

2. The achievement level descriptions will be used to ensure a match between the descriptions and the resulting NAEP items. The achievement level descriptions will be examined, and appropriate instruction provided to item writers to ensure that the items represent the stated descriptions, while adhering to the content and process requirements of the framework and specifications. The descriptions will be used to evaluate the test questions to make certain that the pool of questions encompasses the range of content and process demands specified in the achievement level descriptions, including items within each achievement level interval, and items that scale below basic.

3. As the NAEP item pool is being constructed, additional questions may need to be written for certain content/skill areas if there appear to be any gaps in the pool, relative to the achievement level descriptions.

4. Supporting materials will show the relationship between the achievement levels descriptions and the pool of NAEP test questions.

Principle 3

The Governing Board shall have final authority over all NAEP test questions. This authority includes, but is not limited to, the development of items, establishing the criteria for reviewing items, and the process for review.

Policies and Procedures

1. Under the *No Child Left Behind Act*, a primary duty of the Governing Board pertains to "All Cognitive and Noncognitive Assessment Items." Specifically, the statute states that, "The Board shall have final authority on the appropriateness of all assessment items." Under the law, the Board is therefore responsible for all NAEP test questions as well as all NAEP background questions administered as part of the assessment.

2. To meet this statutory requirement, the Board's Policy on NAEP Item Development and Review shall be adhered to during all phases of NAEP item writing, reviewing, editing, and assessment construction. The National Center for Education Statistic (NCES), which oversees the operational aspects of NAEP, shall ensure that all internal and external groups involved in NAEP item development activities follow the Guiding Principles, Policies and Procedures as set forth in this Board policy.

3. Final review of all NAEP test questions for bias and appropriateness shall be performed by the Board, after all other review procedures have been completed, and prior to administration of the items to students.

Principle 4

The Governing Board shall review all NAEP test questions that are to be administered in conjunction with a pilot test, field test, operational assessment, or special study administered as part of NAEP.

Policies and Procedures

1. To fulfill its statutory responsibility for NAEP item review, the Board shall receive, in a timely manner and with appropriate documentation, all test questions that will be administered to students under the auspices of a NAEP assessment. These items include those slated for pilot testing, field testing, and operational administration.

2. The Board shall review all test items developed for special studies, where the purpose of the special study is to investigate alternate item formats or new technologies for possible

future inclusion as part of main NAEP, or as part of a special study to augment main NAEP data collection.

3. The Board shall not review items being administered as part of test development activities, such as small-scale, informal try-outs with limited groups of students designed to refine items prior to large-scale pilot, field, or operational assessment.

4. NCES shall submit NAEP items to the Board for review in accordance with a mutually agreeable timeline. Items will be accompanied by appropriate documentation as required in this policy. Such information shall consist of procedures and personnel involved in item development and review, the match between the item pool and the framework content and process dimensions, and other related information.

5. For its first review, the Board will examine all items prior to the pilot test or field test stage. In the case of the NAEP reading assessment, all reading passages will be reviewed by the Board prior to item development. For each reading passage, NCES will provide the source, author, publication date, passage length, rationale for minor editing to the passage (if any), and notation of such editing applied to the original passage. NCES will provide information and explanatory material on passages deleted in its fairness review procedures.

6. For its second review, the Board will examine items following pilot or field testing. The items will be accompanied by statistics obtained during the pilot test or field test stage. These statistics shall be provided in a clear format, with definitions for each item analysis statistic collected. Such statistics shall include, but shall not be limited to: p-values for multiple-choice items, number and percentage of students selecting each option for a multiple-choice item, number and percentage not reaching or omitting the item (for multiple-choice and open-ended), number and percentage of students receiving various score points for open-ended questions, mean score point value for open-ended items, appropriate biserial statistics, and other relevant data.

7. At a third stage, for some assessments, the Board will receive a report from the calibration field test stage, which occurs prior to the operational administration. This "exceptions report" will contain information pertaining to any items that were dropped due to differential item functioning (DIF) analysis for bias, other items to be deleted from the operational assessment and the rationale for this decision, and the final match between the framework distribution and the item pool. If the technology becomes available to perform statistically sound item-level substitutions at this point in the cycle (from the initial field test pool), the Board shall be informed of this process as well.

8. All NAEP test items will be reviewed by the Board in a secure manner via in-person meetings, teleconference or videoconference settings, or on-line via a password-protected Internet site. The Board's Assessment Development Committee shall have primary responsibility for item review and approval. However, the Assessment Development Committee, in consultation with the Board Chair, may involve other NAGB members in the item review process on an *ad hoc* basis. The Board may also submit items to external experts, identified by the Board for their subject area expertise, to assist in various duties related to item review. Such experts will follow strict procedures to maintain item security, including signing a Nondisclosure Agreement.

9. Items that are edited between assessments by NCES and/or its item review committees, for potential use in a subsequent assessment, shall be re-examined by the Board prior to a second round of pilot or field testing.

10. Documentation of the Board's final written decision on editing and deleting NAEP items shall be provided to NCES within 10 business days following completion of Board review at each stage in the process.

Principle 5

NAEP test questions will be accurate in their presentation, and free from error. Scoring criteria will be accurate, clear, and explicit.

Policies and Procedures

1. NCES, through its subject area content experts, trained item writers, and item review panels, will examine each item carefully to ensure its accuracy. All materials taken from published sources must be carefully documented by the item writer. Graphics that accompany test items must be clear, correctly labeled, and include the data source where appropriate. Items will be clear, grammatically correct, succinct, and unambiguous, using language appropriate to the grade level being assessed. Item writers will adhere to the specifications document regarding appropriate and inappropriate stimulus materials, terminology, answer choices or distractors, and other requirements for a given subject area. Items will not contain extraneous or irrelevant information that may differentially distract or disadvantage various subgroups of students from the main task of the item.

2. Scoring criteria will accompany each constructed-response item. Such criteria will be clear, accurate, and explicit. Carefully constructed scoring criteria will ensure valid and reliable use of those criteria to evaluate student responses to maximize the accuracy and efficiency of scoring.

3. Constructed-response scoring criteria will be developed initially by the item writers, refined during item review, and finalized during pilot or field test scoring. During pilot or field test scoring, the scoring guides will be expanded to include examples of actual student responses to illustrate each score point. Actual student responses will be used as well, to inform scorers of unacceptable answers.

4. Procedures used to train scorers and to conduct scoring of constructed-response items must be provided to the Board, along with information regarding the reliability and validity of such scoring. If the technology becomes available to score student responses electronically, the Board must be informed of the reliability and validity of such scoring protocol, as compared to human scoring.

Principle 6

All NAEP test questions will be free from racial, cultural, gender, or regional bias, and must be secular, neutral, and non-ideological. NAEP will not evaluate or assess personal or family beliefs, feelings, and attitudes, or publicly disclose personally identifiable information.

Policies and Procedures

1. An item is considered biased if it unfairly disadvantages a particular subgroup of students by requiring knowledge of obscure information unrelated to the construct being assessed. A test question or passage is biased if it contains material derisive or derogatory toward a particular group. For example, a geometry item requiring prior knowledge of the specific dimensions of a basketball court would result in lower scores for students unfamiliar with that sport, even if those students know the geometric concept being measured. Use of a regional term for a soft drink in an item context may provide an unfair advantage to students from that area of the country. Also, an item that refers to a low-achieving student as "slow" would be unacceptable.

2. In conducting bias reviews, steps should be taken to rid the item pool of questions that, because of their content or format, either appear biased on their face, or yield biased estimates of performance for certain subpopulations based on gender, race, ethnicity, or regional culture. A statistical finding of differential item functioning (DIF) will result in a review aimed at identifying possible explanations for the finding. However, such an item will not automatically be deleted if it is deemed valid for measuring what was intended, based on the NAEP assessment framework. Items in which clear bias is found will be eliminated. This policy acknowledges that there may be real and substantial differences in performance among subgroups of students. Learning about such differences, so that performance may be improved, is part of the value of the National Assessment.

3. Items shall be secular, neutral, and non-ideological. Neither NAEP nor its questions shall advocate a particular religious belief or political stance. Where appropriate, NAEP questions may deal with religious and political issues in a fair and objective way.

The following definitions shall apply to the review of all NAEP test questions, reading passages, and supplementary materials used in the assessment of various subject areas:

- Secular—NAEP questions will not contain language that advocates or opposes any particular religious views or beliefs, nor will items compare one religion unfavorably to another. However, items may contain references to religions, religious symbolism, or members of religious groups where appropriate.

  Examples: The following phrases would be acceptable: "shaped like a Christmas tree", "religious tolerance is one of the key aspects of a free society," "Dr. Martin Luther King, Jr. was a Baptist minister," or "Hinduism is the predominant religion in India."

- Neutral and Non-ideological—Items will not advocate for a particular political party or partisan issue, for any specific legislative or electoral result, or for a single perspective on a controversial issue. An item may ask students to explain both sides of a debate, or it may ask them to analyze an issue, or to explain the arguments of proponents or opponents, without requiring students to endorse personally the position they are describing. Item writers should have the flexibility to develop questions that measure important knowledge and skills without requiring both pro and con responses to every item.

Examples: Students may be asked to compare and contrast positions on states rights, based on excerpts from speeches by X and Y; to analyze the themes of Franklin D. Roosevelt's first and second inaugural addresses; to identify the purpose of the Monroe Doctrine; or to select a position on the issue of suburban growth and cite evidence to support this position. Or, students may be asked to provide arguments either for or against Woodrow Wilson's decision to enter World War I. A NAEP question could ask students to summarize the dissenting opinion in a landmark Supreme Court case.

The criteria of neutral and non-ideological also pertain to decisions about the pool of test questions in a subject area, taken as a whole. The Board shall review the entire item pool for a subject area to ensure that it is balanced in terms of the perspectives and issues presented.

4. The Board shall review both stimulus materials and test items to ensure adherence to the NAEP statute and the polices in this statement. Stimulus materials include reading passages, articles, documents, graphs, maps, photographs, quotations, and all other information provided to students in a NAEP test question.

5. NAEP questions will not ask a student to reveal personal or family beliefs, feelings, or attitudes, or publicly disclose personally identifiable information.