A Comparison of NAEP Reading and NAEP Writing Assessments With Current-Generation State Assessments in English Language Arts: Expert Judgment Study

Sheila W. Valencia

Karen K. Wixson

Sami Kitmitto

Nancy Doorey

November 2020
Commissioned by the NAEP Validity Studies (NVS) Panel

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Keena Arbuthnot
Louisiana State University

Peter Behuniak
University of Connecticut

Jack Buckley
American Institutes for Research

James R. Chromy
Research Triangle Institute

Phil Daro
University of California, Berkeley

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Larry Hedges
Northwestern University

Gerunda Hughes
Howard University

Ina V.S. Mullis
Boston College

Scott Norton
Council of Chief State School Officers

James Pellegrino
University of Illinois at Chicago

Gary Phillips
American Institutes for Research

Lorrie Shepard
University of Colorado at Boulder

David Thissen
University of North Carolina, Chapel Hill

Gerald Tindal
University of Oregon

Sheila Valencia
University of Washington

Denny Way
College Board

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403
Email: fstancavage@air.org

CONTENTS

ENGLISH LANGUAGE ARTS STUDY OVERVIEW

Study Purpose

This study was designed to answer the following question:

To what extent are National Assessment of Educational Progress (NAEP) reading (NAEP-R) and NAEP writing (NAEP-W) tests similar to or different from reading and writing tests in current use by states?

To assist in unpacking this multifaceted question, the English language arts (ELA) study team considered the following subquestions:

1. *Content*: How does the balance of content assessed on the NAEP-R and NAEP-W assessments compare with that of state ELA assessments?
2. *Importance*: How well do NAEP-R and NAEP-W items compare with those on state ELA assessments in terms of focus on the most important aspects of the domain and/or the goals of college and career readiness standards that the assessment is designed to measure?
3. *Complexity*: How well do NAEP-R and NAEP-W items compare with those on state ELA assessments in terms of the content-specific complexity or depth of understanding/processing required by items on each assessment?
4. *Other test features*: How do NAEP-R and NAEP-W assessments compare with state assessments in terms of other item features (e.g., item format, scoring rubrics) and stimuli characteristics (e.g., reading passage and writing prompt engagingness, diversity of perspectives, difficulty)?
5. *Newer NAEP-R formats*: How do traditional NAEP-R reading blocks compare with the new scenario-based tasks (SBTs) with regard to item features and stimuli characteristics?

The intent behind the study is to generate information that can inform the consideration of whether, and to what extent, NAEP might need to update its frameworks and assessments to continue as a valid and useful monitor of student achievement in reading and writing, given changes in curriculum and assessments based on new college and career readiness standards.

This report is divided into three sections. The first section presents the methodology of the ELA Expert Judgement Study, including the definitions and rating rubrics used in the analyses. In the second part, the findings for each of the research questions are presented, including a quantitative analysis of the ratings and panelists' qualitative comments. Finally, we provide conclusions and implications for the NAEP-R and NAEP-W frameworks[1,2] and assessments.

Methodology Overview

The analyses for the ELA Expert Judgement Study were conducted using Grade 4 and Grade 8 items from NAEP-R and NAEP-W and four state assessment programs, identified in this report as state assessment (SA) 1 through SA 4. The four programs included two multistate

[1] National Assessment Governing Board. (2017). Reading Framework for the 2017 National Assessment of Educational Progress. Washington, DC: Author.

[2] National Assessment Governing Board. (2017). Writing Framework for the 2017 National Assessment of Educational Progress. Washington, DC: Author.

consortia that developed Common-Core State Standards-aligned assessments and two states that use their own assessments. Secure 2017 operational items from three state assessments and 2017 practice items from one state assessment were analyzed. These items were compared with operational items from the 2017 NAEP-R and NAEP-W assessments and the NAEP-R SBTs piloted in 2017 for use in the 2019 operational assessment.[3] The two sets of NAEP-R items (traditional operational items from 2017 and piloted SBTs), were pooled in all analyses; together, they represented the same proportions of traditional and SBT items intended for, and subsequently used in, the 2019 NAEP-R assessment. NAEP-R and NAEP-W cross-grade blocks of items (administered at both Grades 4 and 8 in 2017) were coded and analyzed separately for each grade. An overview of the assessments, including the number of items examined and each assessment's reporting categories, is summarized in Table 1.

Table 1. Overview of Assessments Reviewed

Grade 4 Assessments (# of items)	Grade 8 Assessments (# of items)	Reporting Categories
		NAEP
Reading (124)	Reading (148)	• Reading scale score and achievement level – Literary – Informational
Writing (9)	Writing (9)	• Writing scale score and achievement level
		State Assessments
ELA (52)	ELA (60)	• Reading score – Literary text – Informational text – Vocabulary • Writing score – Writing expression – Knowledge and use of language conventions • Overall ELA/Literacy scale score and performance level
ELA (50)	ELA (50) Writing (1)	• Reading/Writing process • Critical reading/writing • Vocabulary • Language • Research • Writing composite score • Overall ELA scale score and achievement level
ELA (37)	ELA (37)	• Reading • Listening/Speaking • Writing • Research/Inquiry • Overall ELA/Literacy scale score and achievement level
ELA (28)	ELA (29)	• Reading • Language • Composition: Topic development • Composition: Standard English conventions • Overall ELA scale score and achievement level

NOTE: For the state assessments, the equivalent of one form from each assessment was reviewed.

[3] The 2017 pilot SBTs were included in the analyses because these innovative item types would become part of the NAEP operational reading assessment beginning in 2019.

The study was conducted in two phases. During Phase 1, using a sample of items from each assessment, study leads worked with a small (five-person) Core Group of subject-matter experts to develop, pilot, and refine study materials, which included a Consolidated Content Framework as well as definitions, procedures, and rubrics for classifying and describing the characteristics of items across all the assessments. (See Appendix A for a list of the subject-matter experts who contributed to the study as members of the Core Group and/or Review Panel.) In addition, the project manager and study leads descriptively coded items and stimulus materials (reading stimuli and writing prompts) along several dimensions, including item format, passage length, and points contributed to total test score.

The refined study materials (see Appendices B through E) were then used in Phase 2, in which a larger group of panelists (24) evaluated the full samples of items. The panelists worked in groups of three or four to classify items into subdomains of the Consolidated Content Framework and score items for Importance and Complexity using the associated rubrics. After scoring a block of items, the panelists discussed their scores and made revisions as needed. At the end of the large-group workshop, panelists were given an opportunity to comment on individual items and sets of items. Small-group reflections were posted in writing and then reported to the whole group for further discussion. Those comments are integrated into the findings and conclusions reported below.

OVERVIEW OF ITEM AND STIMULUS MEASURES

Item Measures

In keeping with the study's subquestions, all items were classified or scored on the following dimensions:

- **Content:** defined as the "domains" and "subdomains" of ELA tested on the six assessments. These domains and subdomains were identified by an analysis of the specific standards and test blueprints associated with each assessment and then used to develop a Consolidated Content Framework for categorizing items. By comparing the relative emphasis across these categories, we obtained an indicator of balance (distribution of tested areas) for each assessment that could then be compared across assessments.
- **Importance:** an indicator of how well an item reflects the most important ideas, concepts, or understandings related to the content of the stimuli and/or how well an item assessed authentic, high-quality, grade-level expectations for the domain being assessed (e.g., for *Reading*, *Writing*).
- **Complexity:** refers to the complexity of ideas and/or depth of the thinking and reading or writing processes that students need to use to answer an item correctly or earn full credit.
- Other item features:
  - **Item format** in which the item elicits the student's response (e.g., single selected response, extended constructed response)
  - **Number of points** that are allocated to the item according to the assessment blueprint
  - **Reporting category** to which the item is assigned by the test developers and/or test administrators
- Other stimuli characteristics:
  - **All stimuli:** engagingness, diversity of perspectives, and grade-level appropriateness
  - **Reading passages:** difficulty, genre, and the number of other passages to which a passage is linked (i.e., associated with shared items)
  - Writing prompts: purpose

Content Measures. In this study, *Content* is defined as the domains and subdomains of ELA tested on the six assessments. Because the assessments are built upon different frameworks and have different item specifications, the study leads first conducted a careful analysis of test standards, blueprints, and related documents for each assessment in the study (including NAEP-R and NAEP-W) to construct a content framework that could be used for cross-assessment comparisons. A Consolidated Content Framework (a grid in which the standards for each assessment were sorted into a common set of domains and subdomains) was prepared, representing the knowledge, skills, and processes into which all test items could be categorized. That is, the Consolidated Framework captures the commonalities of content domains and subdomains across the assessments included in the study but is also broad enough to include content and processes that appear in only one or two of the assessments. Variations among test designs, such as whether two components (e.g., reading and writing) are assessed together or separately, also were accounted for in the Consolidated Framework.

Using items from each assessment, the ELA Expert Judgement Study leads worked with the Core Group of subject-matter experts in Phase 1 to pilot test the Consolidated Content Framework for classifying items from each of the six assessments. The framework (see Appendix B) was subsequently refined, resulting in three major domains: *Reading, Writing*, and *Conventions/Research Skills/Language*.

The *Reading* domain has four subdomains--*key ideas, craft, integrate/analyze,* and *vocabulary*. The *Writing* domain has two subdomains--*writing with sources* and *writing without sources*, and the *Conventions/Research Skills/Language* domain has three subdomains--*conventions, research skills,* and *language*. The seven subdomains were further defined to facilitate consistency in panelists' coding (see Appendix C, Expanded Definitions of Consolidated Content Framework Domains and Subdomains).

The content analysis process was carried out in two rounds. In the first, the study leads assigned each item to one of the three domains (*Reading, Writing,* or *Conventions/Research Skills/Language*) to facilitate panelists' use of the appropriate rating rubrics (which were customized by domain or, in the case of *Writing*, by subdomain). In the second round, as part of their review, panelists coded each *Reading* item into one of the four subdomains (*key ideas, craft/structure, integrate/analyze,* and *vocabulary/language*) by reading through both the item and its associated stimulus to determine which subdomain of reading was targeted for assessment. Within the *Writing* domain, items were coded by project leads into the two subdomains of *writing with sources* and *writing without sources*. In the domain of *Conventions/Research Skills/Language*, individual panelists indicated which of the three subdomains was best aligned with the item.

Importance Measures. The concept of Importance refers to the centrality of the assessed knowledge, skill, or process with respect to reading or writing performance. Separate rubrics for Importance were created for *Reading, writing with sources*, *writing without sources,* and *Conventions/Research Skills/Language;* all of the Importance rubrics had three levels. An example is given in Table 2; the full set of rubrics for Importance (and Complexity) can be found in Appendix D.

Table 2. Example Importance Rubric: *Reading* Importance

FOCUS & IMPORTANCE
Consider the focus & importance of the item with regard to students developing a deep understanding of the important ideas, concepts, and content of the text(s)

Level	Description
Level 1	• Item assesses understanding of minor or unimportant ideas, concepts, and/or information in the text(s)s
Level 2	• Item assesses understanding of ideas, concepts, and/or information of some importance that are related or helpful to understanding parts of the text(s)
Level 3	• Item assesses understanding of ideas, concepts, and/or information that are important to building a coherent and deep understanding of the text/s; for narratives, this is often related to plot, character development, theme; for informational text, this is often related to major concepts and key ideas

For both *Reading* and *Writing*, judgments about importance are made based on joint consideration of the item and any associated stimulus material(s) and/or scoring guide. In the domain of ***Reading*****,** Importance is defined as assessing important ideas, concepts, and content of the text(s) that are essential to building a coherent and deep understanding. *Reading* items rated "1" assess minor or unimportant understanding(s); that is, these items are judged to be "unimportant" for measuring students' understanding of major ideas in the text(s). Items rated "2" are considered on target and appropriate, and those rated "3" focus on the most important understandings.

In the domain of ***Writing***, Importance is defined by how well the prompt or item and the associated scoring rubric assess authentic, high-quality, grade-level expectations for writing, either with or without sources. Items rated "3" are judged to represent an important focus for writing instruction and assessment at the targeted grade level.

Panelists assigned ratings individually for Importance while working in groups of three or four, and then discussed their ratings as a group. After discussion, the groups reached 97 percent exact agreement for Importance ratings.

Complexity Measures. The concept of Complexity refers to the complexity of ideas and/or depth of processing involved in reading and writing that is required to correctly respond to, or receive full credit for, an item, taking into account the relation between the item and the stimulus material(s) to which it refers, as well as any related scoring guide. Drawing on the results of the Text-Task-Reader study,[4] each item was evaluated on specific variables associated with Complexity—level of inference, abstractness, amount of synthesis required, language, stimulus structure and features, and, for selected response items, appeal of distractors. This conception of a continuum of Complexity is consistent with the cognitive targets for NAEP reading items, as described in the NAEP Reading Framework (locate/recall, integrate/interpret, and critique/evaluate).

Separate rubrics were developed containing descriptors for each of four levels for Complexity within the context of *Reading, writing without sources, writing with sources,* and *Conventions/Research Skills/Language.* (For an example, see the Complexity rubric for *Reading* in Table 3.) Panelists were instructed to assign a level to each item by identifying the descriptors that best portrayed the item, and then selecting the level that contained all or most of the chosen descriptors. As with Importance*,* panelists assigned ratings individually for Complexity while working in groups of three to four, and then discussed their ratings as a group. After discussion, the groups reached 97 percent exact agreement for Complexity ratings.

[4] Valencia, S., Wixson, K., Ackerman, T., & Sanders, E. (2017). *Identifying text-task-reader interactions related to item and block difficulty in the National Assessment for Educational Progress reading assessment.* Retrieved from https://www.air.org/resource/identifying-text-task-reader-interactions-related-item-and-block-difficulty-national

Table 3. Example Complexity Rubric: *Reading* Complexity

DEPTH & COMPLEXITY
For each item, consider the complexity of the process students use to think across the item stem, text, and distractors (for SSR items) in order to select the correct answer or earn full-credit using the test scoring rubric.

Level	Description
Level 1	• Predominantly explicit information or simple inference within sentence, single paragraph* • Concrete content in text/s, item, rubric; little/no abstract reasoning • Very small amount of source text needed (e.g., 1 sentence or within 1 paragraph) • Understanding of literary or rhetorical features not required • 0-1 distractors are plausible or attractive based on prior knowledge or the text/s (may be N/A for CR items)
Level 2	• Several inferences from text segment or in multiple spots in text/s (may include summarizing) • Mix of concrete & some abstract reasoning in item, text/s, rubric • Amount of text needed – several paragraphs, often contiguous • Some understanding of single literary or rhetorical feature • 1-2 distractors may be plausible/attractive based on prior knowledge or the text/s (may be N/A for CR items);
Level 3	• Inferences require drawing conclusions, generalizations, synthesis, or analysis (e.g. theme), or a simple level of evaluation • Predominantly abstract content or reasoning in stem, text/s, rubric • Amount of text needed – more than 3 contiguous paragraphs or several non-contiguous paragraphs/places within the text/s; may be simple cross-text question • Understanding of several literary or rhetorical features across paragraphs • One or more distractors may be plausible/attractive based on prior knowledge or the text (may be N/A for CR items; distractors may contribute to complexity)
Level 4	• Analysis with critical evaluation, more complex reasoning, more pieces of evidence and/or alternative perspectives in single text OR Level 3 inferences across 2 or more texts • Abstractness - Level 3 across 2 or more texts OR application of concepts to new idea/content • Amount of text - Level 3 across 2 or more texts OR cohesive, integrated understanding of entire selection • Deep understanding of literary or rhetorical features across multiple aspects of the text OR across multiple texts • Multiple distractors may be plausible and/or attractive based on prior knowledge or the text (may be N/A for CR items; distractors may contribute to complexity)

* References to single paragraph or multiple paragraphs refer to "typical" paragraph (in contrast to dialogue). When there is dialogue, consider "typical" amount of text in a paragraph for this age/grade level.

In addition to these three major areas of analysis, several other measures were recorded for each item and each stimulus (i.e., reading passages and writing prompts).

Additional Item Measures. The following item measures were coded by the study leads and project manager.

- **Item format:** Based on the range of item formats used by the assessments reviewed in the study, each item was coded as a single selected response, multiple selected response (more than one answer to be selected), two-part selected response (questions with a Part A and a related Part B), short constructed response, or extended constructed response.

- **Number of points assigned to the item:** This variable facilitated analysis of the actual contribution of items to the total test score. For example, in some cases, item developers allocated more than 10 points to a single item while, in other cases, a single item was allocated only 1 point.
- **Reporting category to which the item was assigned:** Each assessment used a different framework for reporting scores. To assist with the analysis of test profiles, we coded the reporting category to which each item was assigned by the test developer or test administrator.

Stimuli Measures

Stimuli were defined as the reading passages and/or writing prompts that students engaged with before responding to test items. Because both reading and writing assessments require students to process text(s) and/or writing prompt(s) in addition to specific items, all the stimuli (reading selections and writing prompts) were rated on dimensions that have been shown to impact task difficulty and student performance. This is particularly important because college and career readiness standards place an emphasis on reading and responding to "complex" texts.

There were two main areas of analysis related to stimuli: (1) quantitative analyses measuring the word length, Lexile (difficulty), and linking (being part of a set of stimuli, all of which a student needs to process in order to respond to the associated items); and (2) qualitative analyses related to features of the text or writing prompt. The quantitative data were collected and entered by the project manager. The qualitative data were generated by the panelists, who used a 3-point scale to rate each reading and writing stimulus on three qualitative characteristics that have been shown to be related to reading comprehension and writing:

- Grade-level appropriateness: Is the stimulus at an appropriate level of challenge for students at the targeted grade?
- Diversity of perspectives: Does the stimulus reflect people, experiences, or perspectives from diverse ethnic, cultural, and social situations?
- Engagingness: Is the stimulus material likely to engage and interest students from the targeted grade level?

Because these qualitative variables are often interpreted differently, depending on educators' experiences with students, and because we deliberately selected panelists who represented a wide range of experiences, we did not aim for interrater agreement on these variables. During training, we discussed these variables as a whole group and then encouraged the small groups (three to four panelists) that would work together to further discuss these variables as they practiced rating. From that point on, we did not monitor for consensus in this area. To determine final ratings for the qualitative variables, we first looked for agreement among a majority of panelists who rated a specific stimulus. When a majority of panelists did not agree, the study leads reviewed the stimulus and the panelists" ratings, and then assigned a moderated rating. Approximately 12 percent of ratings for grade appropriateness required moderation; approximately 20 percent of ratings for diversity of perspectives and engagingness required moderation.

Other Stimuli Measures. Reading and writing stimuli also were descriptively coded by the project manager and the study leads on specific characteristics that are related to test design and, potentially, student performance:

- Reading genre: narrative, informational, poetry
- Writing purpose: convey/reflect, explain/inform, persuade/argue, text-based narrative, text-based analysis

FINDINGS

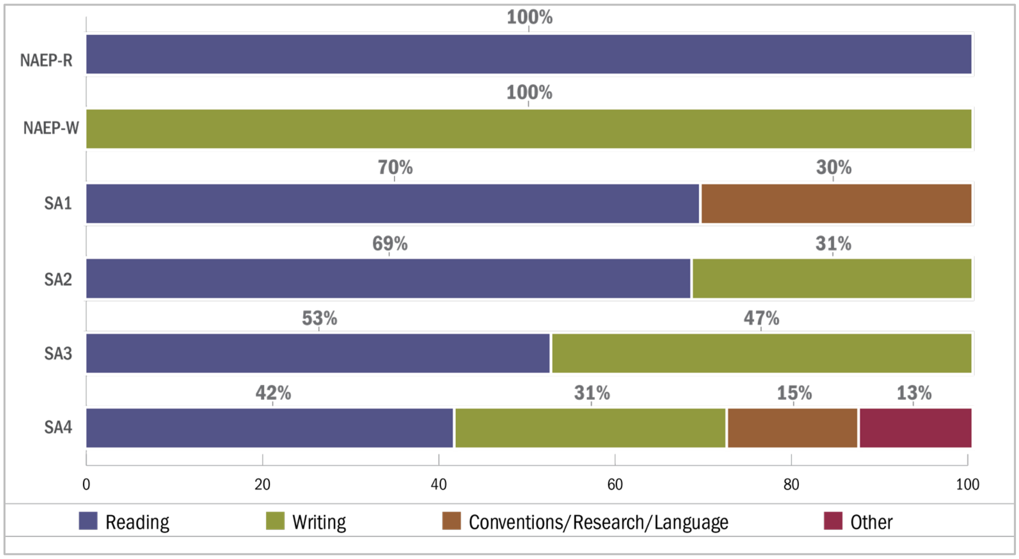
In this section we present the study findings, organized according to the research subquestions. It is important to note that item-level findings are reported in terms of the percentage of total score points allocated to items in a given category rather than the unweighted percentage of items. In some cases, "total score points" refers to total ELA score points; in others, it refers to the total within a particular domain, such as total *Reading* or total *Writing* score points, in order to provide more meaningful comparisons.

Content Analysis

Content Across ELA Domains

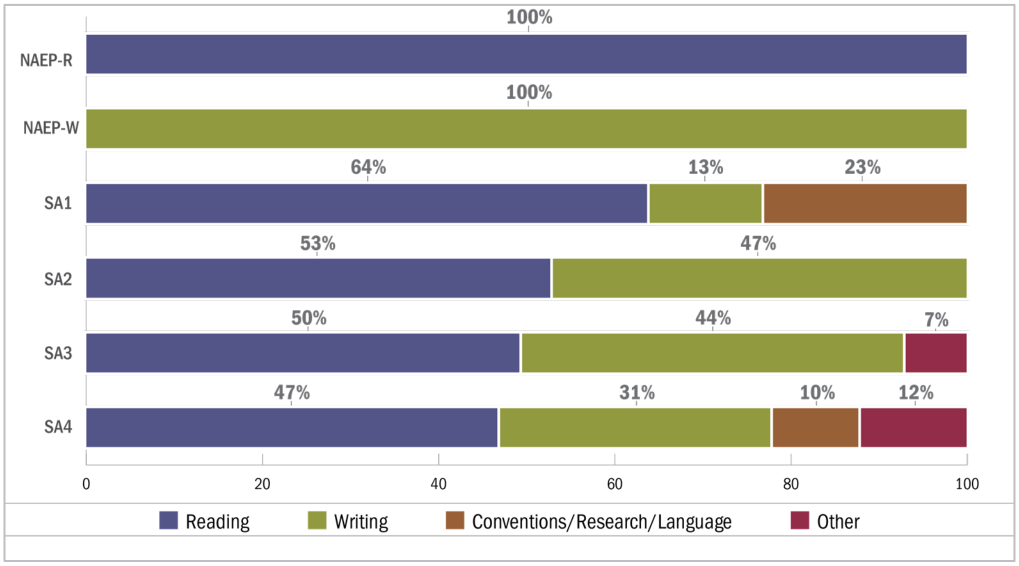
Figure 1 (Grade 4) and Figure 2 (Grade 8) present the distribution of total ELA score points across the three domains identified in the Consolidated Content Framework. Items coded as "Other" were judged to assess listening comprehension.

Figure 1. Distribution of ELA Score Points Across Domains: Grade 4



NOTE: SA=state assessment

Figure 2. Distribution of ELA Score Points Across Domains: Grade 8



NOTE: SA=state assessment

The analysis by content highlights four important differences between NAEP and the other assessments:

1. Most obviously, the total scores from each of the state assessments comprise items representing multiple domains of ELA, including, but not limited to, *Reading* and (in most cases) *Writing*.
2. By contrast, and consistent with the NAEP Reading and Writing Frameworks, NAEP assesses only *Reading* on NAEP-R and only *Writing* on NAEP-W.
3. Two of the state assessments include, as part of their total ELA score, stand-alone items that we classified under the domain of *Conventions/Research Skills/Language*. One assessment includes the results for these items in its subscores for Language, Research, or Vocabulary, depending on the item; the other includes the results in its Reading, Writing, or Research subscores.
4. Two of the state assessments include, as part of their total ELA score, items (classified as "Other") that we considered to be listening comprehension; that is, comprehension questions based on video (rather than written) stimuli. One reports the results in a subscore on Listening and Speaking, the other includes the results as part of one or another of the Reading subscores—Literary Text, Language, or Written Expression—depending on the item.

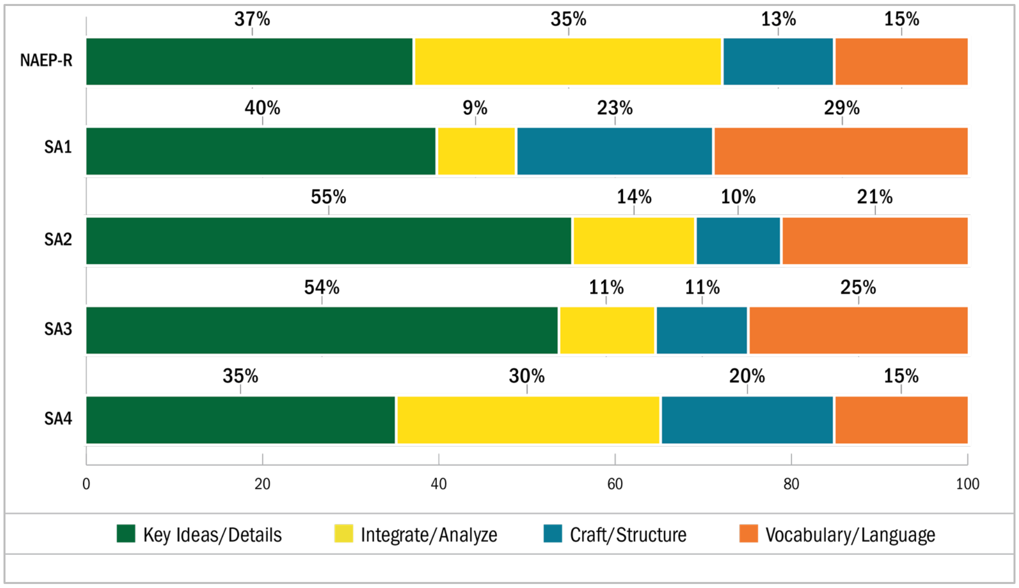
Because the state ELA assessments reviewed for this study include items related to domains (or subdomains) that are not represented in NAEP-R or NAEP-W, comparisons of NAEP total scores with state assessment total ELA scores are not valid. However, comparisons between NAEP-R scores and Reading subscores on the state assessments may be appropriate.

Content for ELA Subdomains

Reading Domain

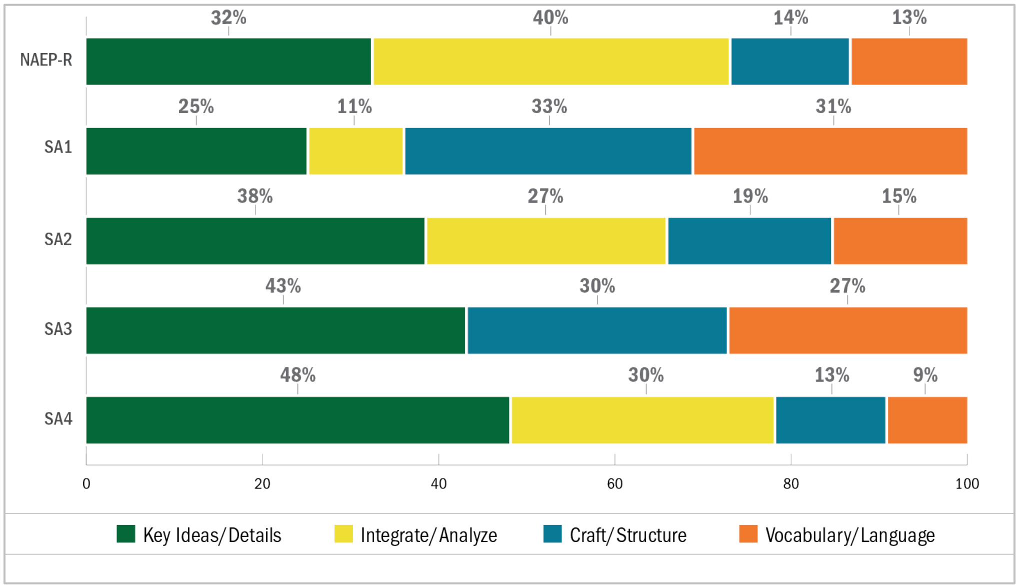
Within the domain of *Reading*, four subdomains (*key ideas, craft, integrate/analyze, vocabulary*) aligned with college and career readiness standards are identified in the Consolidated Content Framework. Figures 3 and 4 present the percentage of *Reading* score points allocated to each of the *Reading* subdomains for Grades 4 and 8, respectively.

Figure 3. Distribution of Reading Score Points Across Subdomains: Grade 4



NOTE: SA=state assessment

Figure 4. Distribution of Reading Score Points Across Subdomains: Grade 8



NOTE: SA=state assessment

Looking across the four subdomains within the *Reading* domain, panelists identified considerably different distributions for each of the assessments.

NAEP-R distributions are most similar to the distributions of SA 4 reading score points at both Grades 4 and 8. Among all the assessments, and at both grades, NAEP-R also has the largest percentage of score points allocated to items that assess *integrate/analyze*. This emphasis is consistent with the emphasis found in the NAEP Reading Framework and college and career readiness standards. The nature of the items in the various subdomains is explored further under the Importance and Complexity sections below.

Writing Domain

The *Writing* domain was parsed into two subdomains—*writing with source*s (*WS*) and *writing without sources* (*WO*)—depending on how writing was tested in each of the assessments. Tables 4 and 5 present the percentages of ELA score points allocated to *Writing* overall, and to each of the *Writing* subdomains for Grades 4 and 8, respectively.

Table 4. Percentage of ELA Score Points Allocated to the Overall *Writing* Domain and Each *Writing* Subdomain: Grade 4

	WS	*WO*	Overall
NAEP-W	0%	100%	**100%**
SA1	0%	0%	**0%**
SA2	31%	0%	**31%**
SA3	47%	0%	**47%**
SA4	27%	4%	**31%**

NOTE: SA=state assessment; *WS=writing with sources; WO=writing without sources*

Table 5. Percentage of ELA Score Points Allocated to the Overall *Writing* Domain and Each *Writing* Subdomain: Grade 8

	WS	*WO*	Overall
NAEP-W	0%	100%	**100%**
SA1	13%	0%	**13%**
SA2	47%	0%	**47%**
SA3	44%	0%	**44%**
SA4	24%	7%	**31%**

NOTE: SA=state assessment; WS=writing with sources; WO=writing without sources

All NAEP-W items are categorized as *WO*—students respond to a short prompt requiring them to write for a specific purpose (to convey experience, to explain/provide information, or to persuade/argue) without reference to extended accompanying reading material. In contrast, all the state assessments base their writing measure solely or primarily on *WS* items, with three of the four assessments devoting 24 to 47 percent of the total ELA score to this subdomain. (Only SA4 includes a few items testing *WO* in addition to its major emphasis on *WS*.) The *WS* items require students to read several related passages on a topic (most often informational passages) and then write an extended response to a prompt using information from the passages. Sometimes, a limited number of items that fall within the *Reading* domain also are associated with the same stimulus passages.

Points for *Writing* are reported differently across the assessments, with some reporting separate subscores for written expression, conventions, evidence provided, and so on. Some points associated with writing items also may be reported as part of a research subscore, when it is present.

The findings here indicate that NAEP-W items, which are all *WO*, may be assessing a different set of skills than the writing items on the other assessments reviewed here.

Conventions/Research Skills/Language Domain

This domain is divided into three subdomains in the Consolidated Content Framework. Items in the *conventions* and *language* subdomains assess areas such as grammar, spelling, or punctuation. Items in the *research skills* subdomain typically require students to identify, from a list, potential websites or resources that could be used to find relevant information for a specific purpose. Only two of the state assessments include items in the *Conventions/Research Skills/Language* domain (SA1 and SA4). SA1 includes the results for these items as part of its subscores for Language, Research, or Vocabulary, depending on the item. SA4 includes the results of *research skills* items in its Reading, Writing or Research subscores. Tables 6 and 7 present the percentage of total ELA score points allocated to this domain overall and to each of the three subdomains of *conventions, research skills, and language* at Grades 4 and 8, respectively

Table 6. Percentage of ELA Score Points Allocated to the Overall *Conventions/Research Skills/Language* Domain and Each of Its Subdomains: Grade 4

	Conventions	Research Skills	Language	Overall
NAEP-R	0%	0%	0%	**0%**
NAEP-W	0%	0%	0%	**0%**
SA1	10%	12%	8%	**30%**
SA2	0%	0%	0%	**0%**
SA3	0%	0%	0%	**0%**
SA4	4%	6%	4%	**14%**

NOTE: SA=state assessment. Only SA1 and SA4 had items in this domain. SA1 had 15 items and SA4 had 6 items.

Table 7. Percentage of ELA Score Points Allocated to the Overall *Conventions/Research Skills/Language* Domain and Each of Its Subdomains: Grade 8

	Conventions	Research Skills	Language	Overall
NAEP-R	0%	0%	0%	**0%**
NAEP-W	0%	0%	0%	**0%**
SA1	4%	13%	7%	**24%**
SA2	0%	0%	0%	**0%**
SA3	0%	0%	0%	**0%**
SA4	2%	8%	0%	**10%**

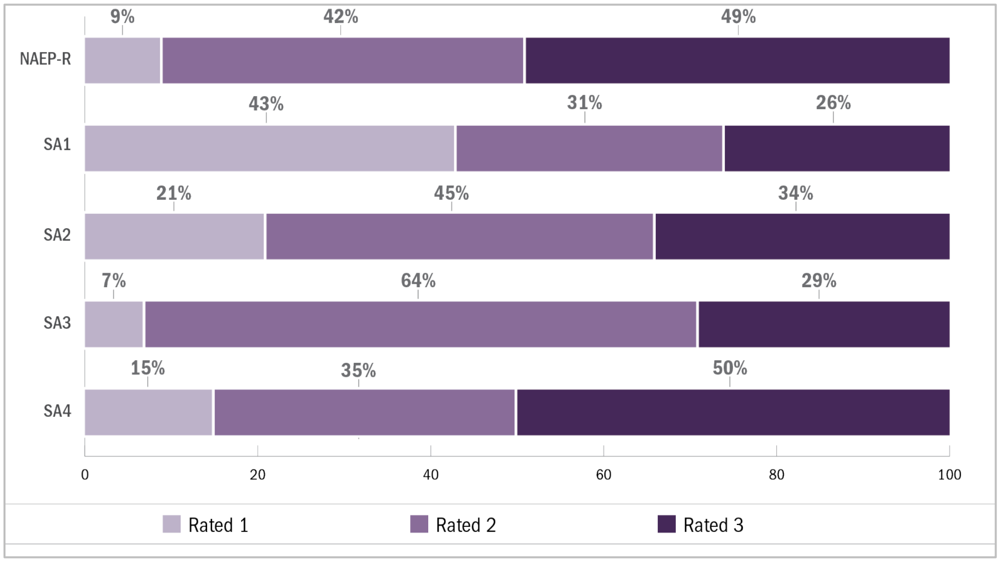
NOTE: SA=state assessment. Only SA1 and SA4 had items in this domain. SA1 had 13 items and SA4 had 4 items.

Importance and Complexity of Reading and Writing Items

Reading: Importance

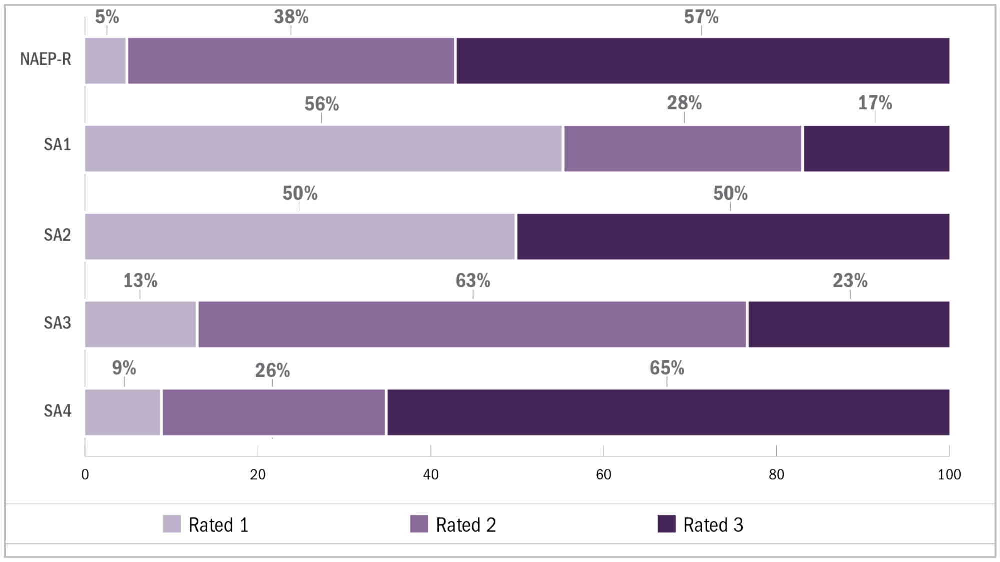
Figures 5 and 6 present, for Grades 4 and 8, respectively, the distribution of *Reading* score points across levels of Importance. Across both grades, the vast majority of NAEP-R score points derive from items that were rated at Level 2 or 3 on Importance, a pattern that is consistent with three of the four state assessments. Only a small percentage of NAEP-R score points come from items rated at Level 1 (unimportant), indicating that students are primarily being assessed on important aspects of text understanding.

Figure 5. Distribution of *Reading* Score Points Across Levels of Importance: Grade 4



NOTE: SA=state assessment

Figure 6. Distribution of *Reading* Score Points Across Levels of Importance*:* Grade 8



NOTE: SA=state assessment

A more in-depth analysis of the Level 1 Importance items may help inform NAEP-R item development. Tables 8 and 9 present the distribution of *Reading* score points, for Grades 4 and 8, respectively, by *Reading* subdomain. The data show that, for each grade level, the few NAEP-R score points at Level 1 are divided approximately equally between the *key ideas/details* and *vocabulary/language* subdomains. The words, phrases, and details assessed in this small percentage of items were judged as unimportant to understanding the central or major ideas in the text. Although this may seem problematic, it is important to note that, in the case of *vocabulary/language*, the NAEP Reading Framework specifies that vocabulary items should target familiar words and concepts rather than the central ideas of a passage, which may be less familiar. Except for Grade 4 SA1, where Level 1 reading score points are more evenly distributed between *craft/structure* and *vocabulary/language*, most of the state assessment items categorized as Level 1 also come from the *vocabulary/language* subdomain at both grade levels. Items in this category may require more scrutiny by test developers.

Table 8. Distribution of *Reading* Score Points Across Levels of Importance by Subdomain: Grade 4

	Percentage of Reading Score Points				
	NAEP-R	**SA1**	**SA2**	**SA3**	**SA4**
	(124 items)	**(35 items)**	**(26 items)**	**(49 items)**	**(18 items)**
Key Ideas/Details					
Importance = 1	5%	6%	3%	0%	5%
Importance = 2	16%	14%	34%	43%	15%
Importance = 3	17%	20%	17%	11%	15%
Total	**37%**	**40%**	**55%**	**54%**	**35%**
Craft/Structure					
Importance = 1	0%	20%	0%	4%	0%
Importance = 2	9%	0%	7%	4%	5%
Importance = 3	4%	3%	3%	4%	15%
Total	**13%**	**23%**	**10%**	**11%**	**20%**
Integrate/Analyze					
Importance = 1	0%	0%	0%	0%	0%
Importance = 2	9%	6%	0%	0%	10%
Importance = 3	27%	3%	14%	11%	20%
Total	**35%**	**9%**	**14%**	**11%**	**30%**
Vocabulary/Language					
Importance = 1	5%	17%	17%	4%	10%
Importance = 2	9%	11%	3%	18%	5%
Importance = 3	2%	0%	0%	4%	0%
Total	**15%**	**29%**	**21%**	**25%**	**15%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

Note: SA=state assessment

Table 9. Distribution of *Reading* Score Points Across Levels of *Importance*, by Subdomain: Grade 8

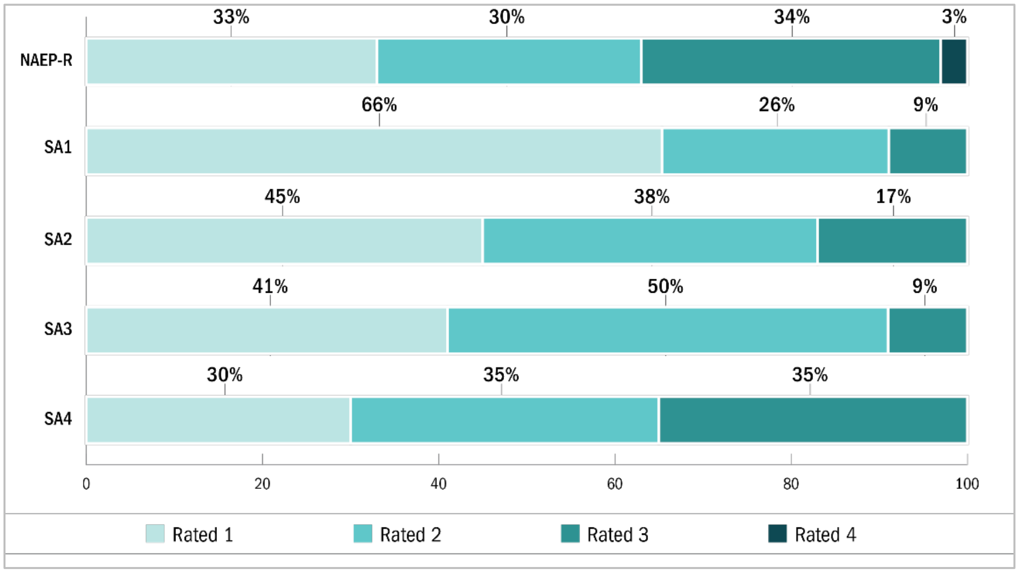
	Percentage of Reading Score Points				
	NAEP-R **(148 items)**	**SA1** **(36 items)**	**SA2** **(26 items)**	**SA3** **(53 items)**	**SA4** **(22 items)**
Key Ideas/Details					
Importance = 1	2%	3%	0%	7%	0%
Importance = 2	16%	14%	19%	13%	22%
Importance = 3	14%	8%	19%	23%	26%
Total	**32%**	**25%**	**38%**	**43%**	**48%**
Craft/Structure					
Importance = 1	0%	22%	0%	0%	0%
Importance = 2	8%	8%	8%	30%	4%
Importance = 3	7%	3%	12%	0%	9%
Total	**14%**	**33%**	**19%**	**30%**	**13%**
Integrate/Analyze					
Importance = 1	0%	3%	0%	0%	0%
Importance = 2	6%	3%	8%	0%	0%
Importance = 3	34%	6%	19%	0%	30%
Total	**40%**	**11%**	**27%**	**0%**	**30%**
Vocabulary/Language					
Importance = 1	3%	28%	0%	7%	9%
Importance = 2	9%	3%	15%	20%	0%
Importance = 3	2%	0%	0%	0%	0%
Total	**13%**	**31%**	**15%**	**27%**	**9%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Reading: Complexity

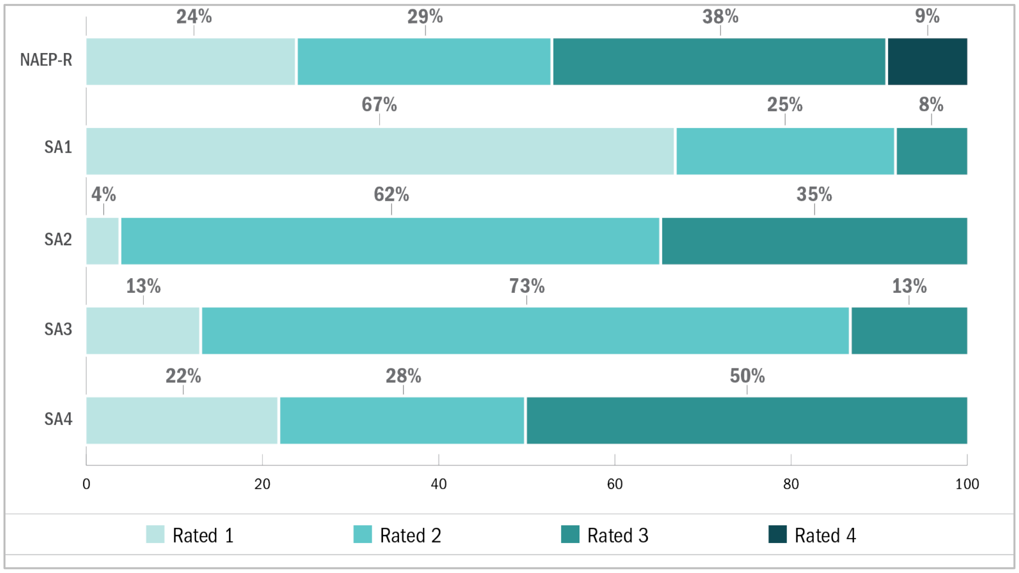
Complexity is an indication of the complexity of ideas and/or the depth of the reading processes and thinking in which a student must engage to receive full credit for a reading item. It is a function of three test components—the item, the scoring rubric, and the text. Figures 7 and 8 present the distribution of *Reading* score points across the levels of Complexity for Grades 4 and 8, respectively. For this study, items rated as Levels 3 and 4 are considered to be measuring complex reading, consistent with college and career readiness standards, while those at Levels 1 and 2 are considered to be measuring more explicit, concrete aspects of text. We would expect all reading assessments to have items that range from 1 to 4 on Complexity but, consistent with college and career readiness expectations, we also would expect that there would be greater emphasis in the 3–4 range.

Figure 7. Distribution of *Reading* Score Points Across Levels of Complexity: Grade 4



NOTE: SA=state assessment

Figure 8. Distribution of *Reading* Score Points Across Levels of Complexity: Grade 8



NOTE: SA=state assessment

Looking across the findings, we see that NAEP-R and SA4 have a fairly balanced representation of items across Levels 1–3 on the Complexity scale. At both grade levels, NAEP-R is the only assessment that has any *Reading* score points derived from items rated at Level 4 on Complexity. At Grade 4, only SA4 has a distribution of score points comparable to NAEP-R. Compared with the other state assessments, NAEP-R and SA4 have higher percentages of *Reading* score points (more than one third) coming from items rated at Level 3 or 4 and lower percentages of points (about one third) coming from items rated at Level 1. At Grade 8, the same pattern holds, but the percentages of *Reading* score points derived from items rated at Level 3 or 4 is closer to half, and the percentage derived from items rated at Level 1 is about one quarter.

Tables 10 and 11 present the distribution of *Reading* score points across levels of Complexity by *Reading* subdomain (*key ideas/details, craft/structure, integrate/analyze, vocabulary/language*) for Grades 4 and 8, respectively. At Grade 4, NAEP-R's Level-4 score points derive from items in the subdomains of *craft/structure* and *integrate/analyze* and, at Grade 8, they are all from items in the *integrate/analyze* subdomain. Overall, assessments were more likely to have items at Levels 3 and 4 for *craft/structure* and *integrate/analyze,* which naturally require deeper, more complex reading.

The pattern of Complexity ratings in the *key ideas/details* subdomain is fairly similar across all the assessments at Grade 4, falling mostly at Levels 1 and 2. At Grade 8, all the assessments except SA1 had some *key ideas /details* items that were rated at Level 3; there were no Level 4s in this subdomain for any assessment.

At both grades and across all the assessments, the majority of *vocabulary/language* items were rated at Level 1 for Complexity, with SA4 having 100 percent of *vocabulary/language* score points at this level. Only NAEP-R had *vocabulary/language* items that reached Level 3. (There were no *vocabulary/language* items rated at Level 4 on any of the assessments.) These findings may suggest that developers are using a relatively simple, definitional approach, to testing vocabulary.

Table 10. Distribution of *Reading* Score Points Across Levels of *Complexity* by Subdomain: Grade 4

	Percentage of Reading Score Points				
	NAEP-R **(124 items)**	**SA1** **(35 items)**	**SA2** **(26 items)**	**SA3** **(49 items)**	**SA4** **(18 items)**
Key Ideas/Details					
Complexity = 1	21%	23%	28%	20%	15%
Complexity = 2	13%	17%	28%	30%	20%
Complexity = 3	3%	0%	0%	4%	0%
Complexity = 4	0%	0%	0%	0%	0%
Total	**37%**	**40%**	**55%**	**54%**	**35%**
Craft/Structure					
Complexity = 1	1%	17%	3%	4%	0%
Complexity = 2	3%	3%	3%	7%	15%
Complexity = 3	7%	3%	3%	0%	5%
Complexity = 4	1%	0%	0%	0%	0%
Total	**13%**	**23%**	**10%**	**11%**	**20%**
Integrate/Analyze					
Complexity = 1	2%	0%	0%	0%	0%
Complexity = 2	10%	3%	0%	5%	0%
Complexity = 3	22%	6%	14%	5%	30%
Complexity = 4	2%	0%	0%	0%	0%
Total	**35%**	**9%**	**14%**	**11%**	**30%**
Vocabulary/Language					
Complexity = 1	9%	26%	14%	18%	15%
Complexity = 2	5%	3%	7%	7%	0%
Complexity = 3	1%	0%	0%	0%	0%
Complexity = 4	0%	0%	0%	0%	0%
Total	**15%**	**29%**	**21%**	**25%**	**15%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Table 11. Distribution of *Reading* Score Points Across Levels of *Complexity*, by Subdomain: Grade 8

	Percentage of Reading Score Points				
	NAEP-R **(148 items)**	**SA1** **(36 items)**	**SA2** **(26 items)**	**SA3** **(53 items)**	**SA4** **(22 items)**
Key Ideas/Details					
Complexity = 1	14%	14%	0%	5%	13%
Complexity = 2	16%	11%	27%	35%	20%
Complexity = 3	2%	0%	12%	3%	15%
Complexity = 4	0%	0%	0%	0%	0%
Total	**32%**	**25%**	**38%**	**43%**	**48%**
Craft/Structure					
Complexity = 1	0%	25%	0%	0%	0%
Complexity = 2	4%	6%	12%	20%	9%
Complexity = 3	11%	3%	8%	10%	4%
Complexity = 4	0%	0%	0%	0%	0%
Total	**14%**	**33%**	**19%**	**30%**	**13%**
Integrate/Analyze					
Complexity = 1	0%	0%	0%	0%	0%
Complexity = 2	6%	6%	12%	0%	0%
Complexity = 3	25%	6%	15%	0%	30%
Complexity = 4	9%	0%	0%	0%	0%
Total	**40%**	**11%**	**27%**	**0%**	**30%**
Vocabulary/Language					
Complexity = 1	9%	28%	4%	8%	9%
Complexity = 2	3%	3%	12%	18%	0%
Complexity = 3	1%	0%	0%	0%	0%
Complexity = 4	0%	0%	0%	0%	0%
Total	**13%**	**31%**	**15%**	**27%**	**9%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

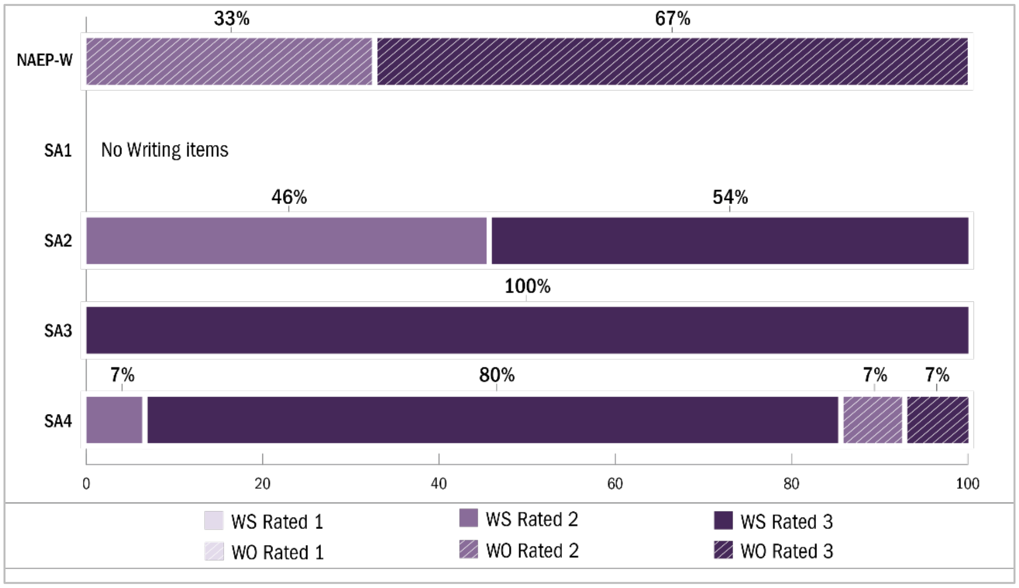
NOTE: SA=state assessment

Writing: Importance

As detailed above, *Writing* items were classified as falling into one of two subdomains—*WS* or *WO*. For both subdomains, Importance was evaluated by examining the extent to which each writing prompt or item addressed authentic, high-quality, grade-level expectations for writing; that is, how well each prompt or item represented an important focus for writing instruction and assessment at the targeted grade level.

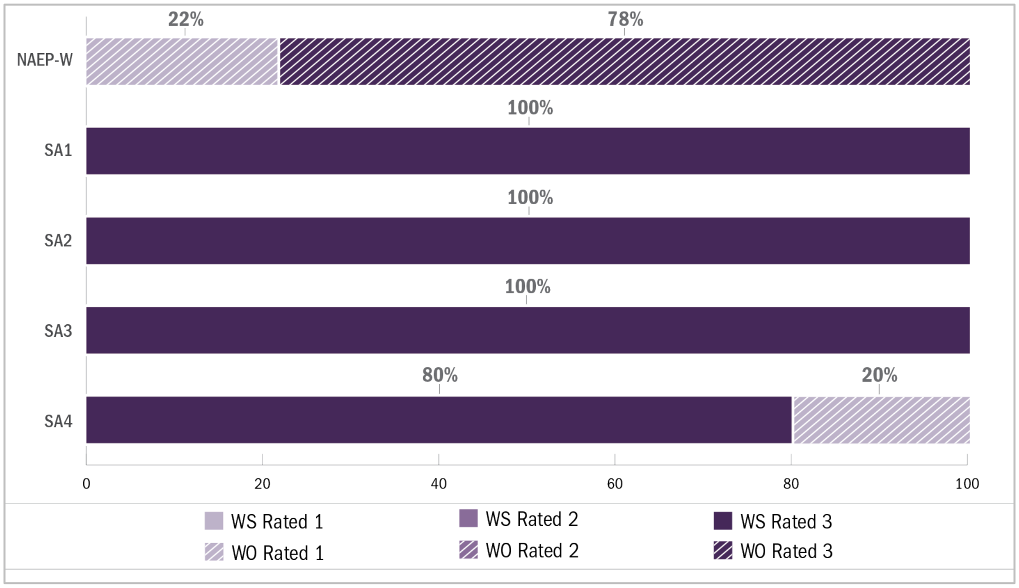
Figures 9 and 10 present the distribution of score points across levels of Importance for *Writing* items at Grades 4 and 8, respectively. At both grades, the majority of *Writing* items for the assessments that included writing received an Importance rating of 3 (SA1 does not assess writing at Grade 4.) Only NAEP-W and SA4 include *WO* items, but there is no meaningful way to compare *WO* items on these two assessments. NAEP-W items require students to write a complete composition using an extended constructed-response format, while SA4 *WO* items are all multiple choice. However, it is informative to compare Importance ratings for NAEP-W (which is entirely *WO*) with assessments that test *WS* using extended constructed response. (More than 75 percent of the *WS* items on the state assessments use extended constructed responses.)

Figure 9. Distribution of *Writing* Score Points Across Levels of Importance: Overall and by Subdomain, Grade 4



NOTE: SA=state assessment; WS=writing with sources; WO=writing without sources

Figure 10. Distribution of *Writing* Score Points Across Levels of Importance: Overall and by Subdomain, Grade 8



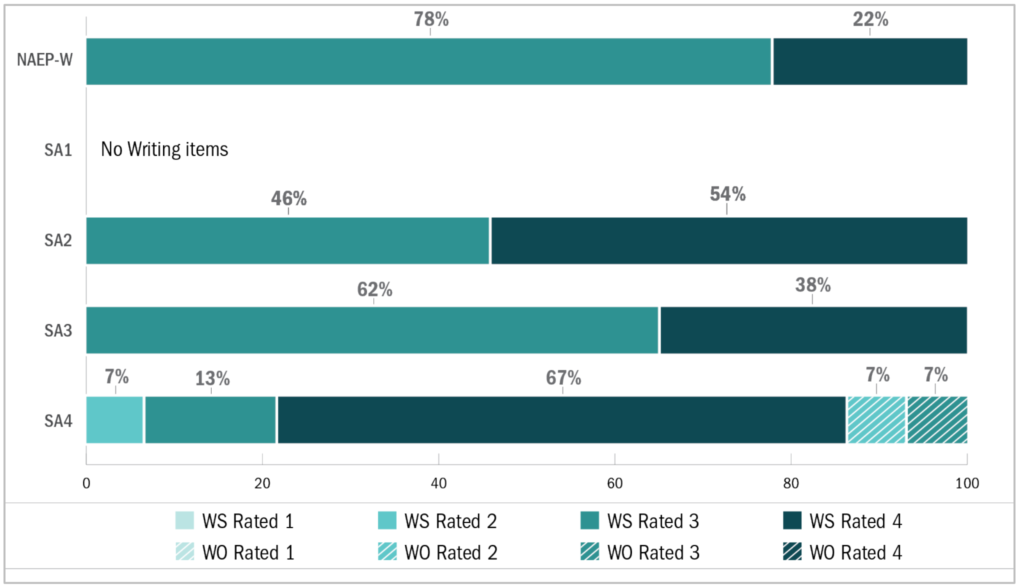
NOTE: SA=state assessment; WS=writing with sources; WO=writing without sources

In general, NAEP-W has a larger percentage of points derived from items with lower Importance ratings (e.g., Level 2) compared with assessments that use *WS* items, although SA2 is an exception at Grade 4. This suggests that it may be more difficult to attain higher levels of Importance when students are responding to prompts that are not accompanied by reading passages.

Writing: Complexity

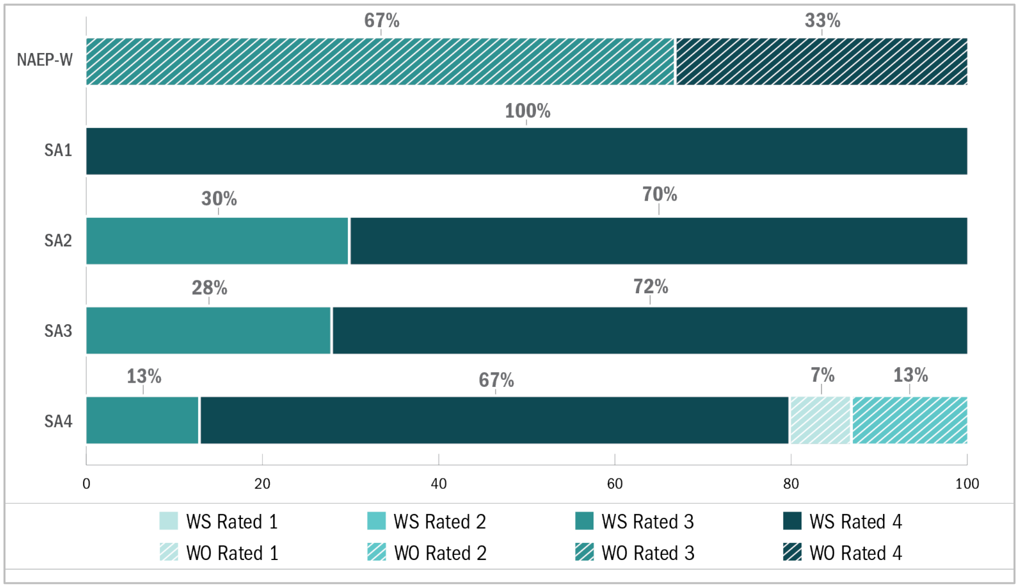
As noted previously, all NAEP-W items are classified *as WO*. Although SA4 also has *WO* items, they only account for a small proportion of *Writing* score points on this assessment (4 percent and 7 percent, respectively, for Grades 4 and 8). Also, the nature of the *WO* items on NAEP-W is quite different from the *WO* items on SA4. This difference is reflected in the comparison of Complexity ratings for *WO* items on these two assessments: NAEP-W items are rated at Levels 3 and 4, while SA4 *WO* items are rated much lower at both grade levels. Figures 11 and 12 present the distribution of *Writing* score points across levels of Complexity*,* overall and by subdomain at Grades 4 and 8, respectively.

Figure 11. Distribution of *Writing* Score Points Across Levels of Complexity: Overall and by Subdomain, Grade 4



NOTE: SA=state assessment; WS=writing with sources; WO=writing without sources

Figure 12. Distribution of *Writing* Score Points Across Levels of Complexity: Overall and by Subdomain, Grade 8



NOTE: SA=state assessment; WS=writing with sources; WO=writing without sources

It also is informative to look at the Complexity ratings for items classified as *WS* in comparison to NAEP-W items. Although NAEP-W has no *WS* items, three of the four state assessments have these items at Grade 4 (one state does not assess writing at Grade 4), and all have them at Grade 8. What is notable here is that a significant percentage of the points coming from *WS* items are rated at Level 4. Paralleling the findings for Importance, this suggests that *WS* items provide opportunities for students to engage in more complex reasoning than *WO* items, the current model for NAEP writing assessments.

Other Features: Item Format

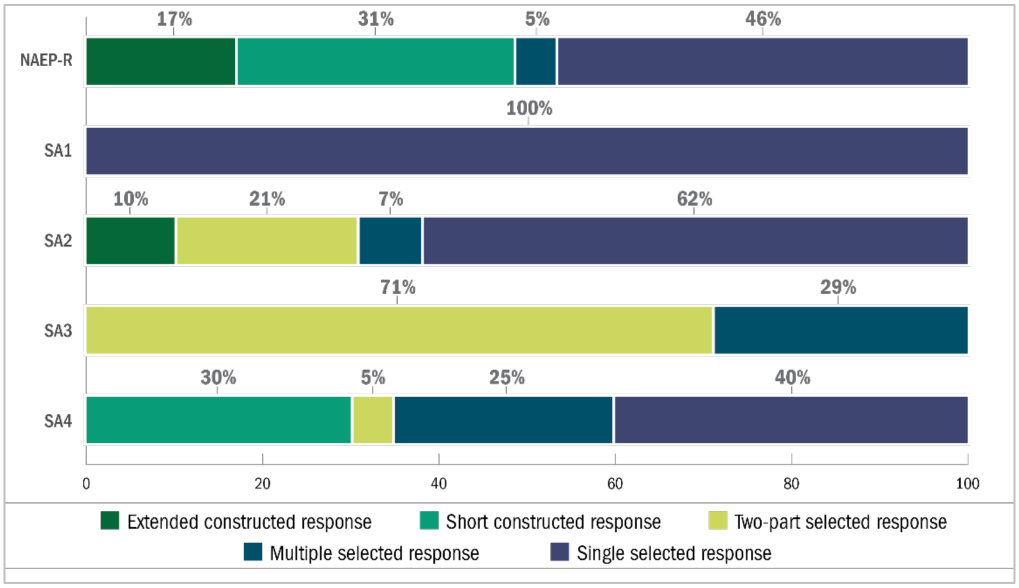
Reading

Each item was reviewed and coded as

- Single selected response,
- Multiple selected response, where more than one answer can be selected,
- Two-part selected response, used for questions with a Part A and a related Part B,
- Short constructed response, or
- Extended constructed response

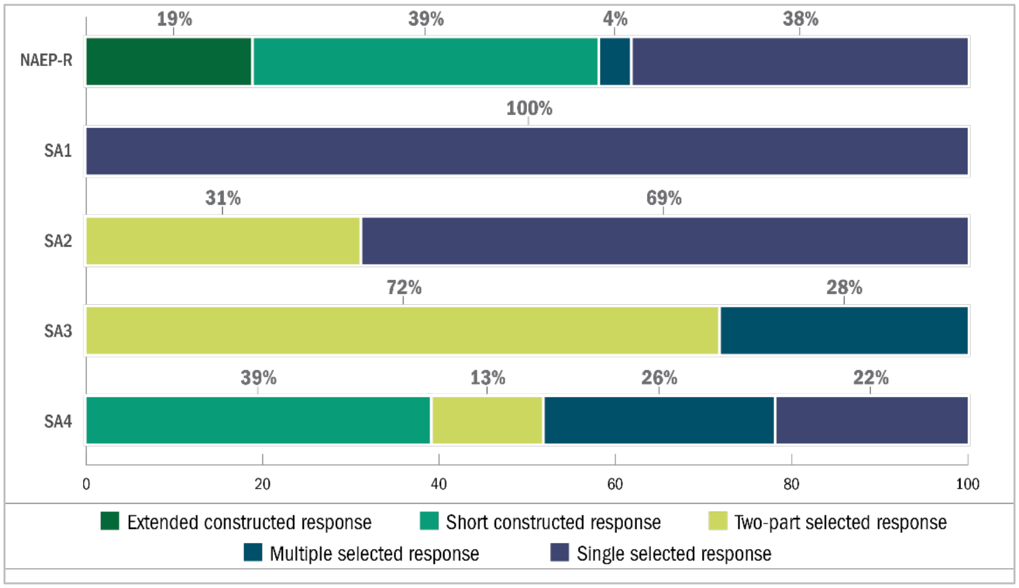
Figures 13 and 14 present the distribution of *Reading* score points across item formats for Grades 4 and 8, respectively.

Figure 13. Distribution of *Reading* Score Points Across Item Formats: Grade 4



NOTE: SA=state assessment

Figure 14. Distribution of *Reading* Score Points Across Item Formats: Grade 8



NOTE: SA=state assessment

It is interesting to note the differences in *Reading* item formats across assessments. NAEP-R employed all formats except the two-part selected response; it also was the only assessment to employ the extended constructed-response format at Grade 8 and one of only two assessments to do so at Grade 4. Furthermore, NAEP-R has the greatest number of *Reading* score points coming from constructed-response items (short or extended) at both Grades 4 and 8. Use of the single selected response format varied considerably across assessments; it was the most predominant (or only) item format at both grade levels in two of the assessments (SA1 and SA2) and was completely absent from another (SA3).

Tables 12 and 13 provide information about Importance ratings by item format for Grades 4 and 8, respectively. In general, there was a higher incidence of Level 1 Importance ratings associated with single selected response items compared with other item formats. This pattern occurred at both grade levels.

Table 12. Distribution of *Reading* Score Points Across Levels of Importance, by Item Format: Grade 4

	Percentage of Reading Score Points				
	NAEP-R (124 items)	SA1 (35 items)	SA2 (26 items)	SA3 (49 items)	SA4 (18 items)
Extended Constructed Response					
Importance = 1	0%	0%	0%	0%	0%
Importance = 2	2%	0%	0%	0%	0%
Importance = 3	15%	0%	10%	0%	0%
Total	**17%**	**0%**	**10%**	**0%**	**0%**
Short Constructed Response					
Importance = 1	1%	0%	0%	0%	0%
Importance = 2	19%	0%	0%	0%	10%
Importance = 3	12%	0%	0%	0%	20%
Total	**31%**	**0%**	**0%**	**0%**	**30%**
Two-Part Selected Response					
Importance = 1	0%	0%	7%	4%	5%
Importance = 2	0%	0%	7%	45%	0%
Importance = 3	0%	0%	7%	23%	0%
Total	**0%**	**0%**	**21%**	**71%**	**5%**
Multiple Selected Response					
Importance = 1	0%	0%	0%	4%	0%
Importance = 2	0%	0%	7%	20%	15%
Importance = 3	5%	0%	0%	5%	10%
Total	**5%**	**0%**	**7%**	**29%**	**25%**
Single Selected Response					
Importance = 1	9%	43%	14%	0%	10%
Importance = 2	21%	31%	31%	0%	10%
Importance = 3	17%	26%	17%	0%	20%
Total	**46%**	**100%**	**62%**	**0%**	**40%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Table 13. Distribution of *Reading* Score Points Across Levels of Importance, by Item Format: Grade 8

	Percentage of Reading Score Points				
	NAEP-R (148 items)	SA1 (36 items)	SA2 (26 items)	SA3 (53 items)	SA4 (22 items)
Extended Constructed Response					
Importance = 1	0%	0%	0%	0%	0%
Importance = 2	3%	0%	0%	0%	0%
Importance = 3	16%	0%	0%	0%	0%
Total	**19%**	**0%**	**0%**	**0%**	**0%**
Short Constructed Response					
Importance = 1	0%	0%	0%	0%	0%
Importance = 2	13%	0%	0%	0%	13%
Importance = 3	26%	0%	0%	0%	26%
Total	**39%**	**0%**	**0%**	**0%**	**39%**
Two-Part Selected Response					
Importance = 1	0%	0%	0%	10%	0%
Importance = 2	0%	0%	12%	52%	4%
Importance = 3	0%	0%	19%	10%	9%
Total	**0%**	**0%**	**31%**	**72%**	**13%**
Multiple Selected Response					
Importance = 1	0%	0%	0%	3%	4%
Importance = 2	2%	0%	0%	12%	0%
Importance = 3	2%	0%	0%	13%	22%
Total	**4%**	**0%**	**0%**	**28%**	**26%**
Single Selected Response					
Importance = 1	5%	56%	0%	0%	4%
Importance = 2	20%	28%	38%	0%	9%
Importance = 3	13%	17%	31%	0%	9%
Total	**38%**	**100%**	**69%**	**0%**	**22%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Tables 14 and 15 provide information about Complexity ratings by item format for each grade. NAEP-R is the only assessment with items rated at Level 4 for Complexity at both grades, and all of these points derive from short or extended constructed response items. Level 1 Complexity ratings are most common for single selected response items across all assessments at both grades, with the exception of SA3, which did not have any single response items. The majority of Level 1 Complexity ratings for SA3 at both grades were associated with two-part selected response items.

Table 14. Distribution of *Reading* Score Points Across Levels of Complexity, by Item Format: Grade 4

	Percentage of Reading Score Points				
	NAEP-R (124 items)	SA1 (35 items)	SA2 (26 items)	SA3 (49 items)	SA4 (18 items)
Extended Constructed Response					
Complexity = 1	0%	0%	0%	0%	0%
Complexity = 2	2%	0%	0%	0%	0%
Complexity = 3	14%	0%	10%	0%	0%
Complexity = 4	2%	0%	0%	0%	0%
Total	**17%**	**0%**	**10%**	**0%**	**0%**
Short Constructed Response					
Complexity = 1	5%	0%	0%	0%	0%
Complexity = 2	11%	0%	0%	0%	0%
Complexity = 3	14%	0%	0%	0%	30%
Complexity = 4	1%	0%	0%	0%	0%
Total	**31%**	**0%**	**0%**	**0%**	**30%**
Two-Part Selected Response					
Complexity = 1	0%	0%	10%	34%	5%
Complexity = 2	0%	0%	10%	30%	0%
Complexity = 3	0%	0%	0%	7%	0%
Complexity = 4	0%	0%	0%	0%	0%
Total	**0%**	**0%**	**21%**	**71%**	**5%**
Multiple Selected Response					
Complexity = 1	1%	0%	0%	7%	5%
Complexity = 2	3%	0%	7%	20%	20%
Complexity = 3	1%	0%	0%	2%	0%
Complexity = 4	0%	0%	0%	0%	0%
Total	**5%**	**0%**	**7%**	**29%**	**25%**
Single Selected Response					
Complexity = 1	27%	66%	34%	0%	20%
Complexity = 2	14%	26%	21%	0%	15%
Complexity = 3	5%	9%	7%	0%	5%
Complexity = 4	0%	0%	0%	0%	0%
Total	**46%**	**100%**	**62%**	**0%**	**40%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Table 15. Distribution of *Reading* Score Points Across Levels of Complexity, by Item Format: Grade 8

	Percentage of Reading Score Points				
	NAEP-R (148 items)	SA1 (36 items)	SA2 (26 items)	SA3 (53 items)	SA4 (22 items)
Extended Constructed Response					
Complexity = 1	0%	0%	0%	0%	0%
Complexity = 2	1%	0%	0%	0%	0%
Complexity = 3	11%	0%	0%	0%	0%
Complexity = 4	7%	0%	0%	0%	0%
Total	**19%**	**0%**	**0%**	**0%**	**0%**
Short Constructed Response					
Complexity = 1	4%	0%	0%	0%	0%
Complexity = 2	10%	0%	0%	0%	9%
Complexity = 3	23%	0%	0%	0%	30%
Complexity = 4	2%	0%	0%	0%	0%
Total	**39%**	**0%**	**0%**	**0%**	**39%**
Two-Part Selected Response					
Complexity = 1	0%	0%	0%	10%	4%
Complexity = 2	0%	0%	23%	52%	7%
Complexity = 3	0%	0%	8%	10%	2%
Complexity = 4	0%	0%	0%	0%	0%
Total	**0%**	**0%**	**31%**	**72%**	**13%**
Multiple Selected Response					
Complexity = 1	0%	0%	0%	3%	4%
Complexity = 2	4%	0%	0%	22%	9%
Complexity = 3	0%	0%	0%	3%	13%
Complexity = 4	0%	0%	0%	0%	0%
Total	**4%**	**0%**	**0%**	**28%**	**26%**
Single Selected Response					
Complexity = 1	19%	67%	4%	0%	13%
Complexity = 2	13%	25%	38%	0%	4%
Complexity = 3	5%	8%	27%	0%	4%
Complexity = 4	0%	0%	0%	0%	0%
Total	**38%**	**100%**	**69%**	**0%**	**22%**
Total Reading	**100%**	**100%**	**100%**	**100%**	**100%**

NOTE: SA=state assessment

Writing

At Grade 4, three of four of the state assessments included items that assessed *writing from sources* (one state does not test writing at Grade 4). The format for these items varied across assessments, with the majority of items calling for extended constructed response.

At Grade 8, all the state assessments included items that assessed *writing from sources*. The format for these was either extended constructed response or short constructed response, and item format was related to the number of sources that students read before responding, with larger numbers of sources associated with the extended constructed-response items.

Conventions, Research Skills, and Language

Only SA1 and SA4 have items that fall into this domain. For SA1, at both Grades 4 and 8, 100 percent of the *Conventions/Research Skills/Language* score points derive from single selected response items. This compares with SA4, where the percentage of *Conventions/Research Skills/Language* score points coming from single selected response items is 29 percent at Grade 4 and zero at Grade 8. At both grades, SA4 also had items in this domain that were classified as multiple selected response or short constructed response. At Grade 4, 43 percent of the *Conventions/Research Skills/Language* score points come from multiple selected response items and 29 percent come from short constructed-response items, compared with 60 percent multiple selected response and 40 percent short constructed response at Grade 8.

Other Features: Stimuli Characteristics

Reading

In *Reading*, stimuli were defined as the reading selections that students were expected to read prior to responding to the reading items. A good deal of attention has been focused on reading assessment stimuli because college and career readiness standards emphasize the reading of complex material; this has led to concerns about the difficulty levels of reading materials that are assigned in K–12 schools.

Grade-Appropriate Difficulty, Length, and Linkage of Reading Stimuli. Tables 16 and 17 provide information on the Lexile scores, average text length, and percentage of "linked" texts (i.e., texts that are presented as a shared set, which, in turn, serves as the stimulus for a set of reading items). In general, and at both grade levels, NAEP-R is comparable to the state assessments on these three dimensions. However, it is notable that, at Grade 8, NAEP-R has no texts that appear in linked sets of three or more. In comparison, three of the four state assessments have between 20 percent and 67 percent of texts linked together in sets of three or more. Such stimulus sets do occur in NAEP-R at Grade 4, specifically in the newly developed SBTs.

Table 16. Difficulty, Length, and Linking for *Reading* Stimuli: Grade 4

		Lexile*	Length			Linking	
Assessment	***n***	**Mean (SD)**	**Mean (SD)**	**Min**	**Max**	**% Linked to One Additional Stimulus**	**% Linked to Two+ Additional Stimuli**
NAEP-R	18	895 (135)	604 (296)	47	925	33%	17%
SA1	14	705 (141)	419 (147)	220	694	43%	0%
SA2	4	750 (82)	572 (430)	153	1,172	50%	0%
SA3	8	927 (261)	425 (226)	156	862	25%	38%
SA4	5	870 (130)	477 (176)	185	681	0%	60%

*Excludes poetry and video.

NOTE: SA=state assessment

Table 17. Difficulty, Length, and Linking for *Reading* Stimuli: Grade 8

		Lexile*	Length			Linking	
Assessment	***N***	**Mean (SD)**	**Mean (SD)**	**Min**	**Max**	**% Linked to One Additional Stimulus**	**% Linked to Two+ Additional Stimuli**
NAEP-R	19	1,026 (174)	711 (324)	211	1,429	42%	0%
SA1	13	1,058 (155)	580 (254)	190	941	27%	20%
SA2	5	750 (71)	785 (518)	234	1,596	80%	0%
SA3	9	997 (163)	626 (193)	448	1,004	45%	33%
SA4	6	1,117 (175)	537 (329)	155	967	0%	67%

*Excludes poetry and video.

NOTE: SA=state assessment

Based on the new Lexile recommendations in the Common Core State Standards (see Table 18) all assessments at both grade levels are within the suggested difficulty range except for SA2, which has a lower than recommended Lexile range at Grade 8.

Table 18. Lexile Levels Recommended by the Common Core State Standards

Text Complexity Grade Band in the Standards	**Old Lexile Ranges**	**Lexile Ranges Aligned to CCR Expectations**
K–1	N/A	N/A
2–3	450–725	450–790
4–5	645–845	770–980
6–8	860–1010	955–1155
9–10	960–1115	1080–1305
11–CCR	1070–1220	1215–1355

NOTE: CCR=college and career readiness

Tables 19 and 20 provide information on the percentages of Grade 4 and Grade 8 *Reading* stimuli that were rated by the panelists as below, on, or above grade level. For all the assessments, the overwhelming majority of reading stimuli were rated as "on grade level." The exception is SA1, in which approximately one third of the stimuli were rated below grade level at both grades. NAEP-R is most comparable to SA3 in that both had some stimuli that were rated as too difficult at Grade 4 and some that were rated as too easy at Grade 8. In the case of NAEP-R, these ratings are most likely attributable to the cross-grade blocks that are administered at both Grades 4 and 8.

Table 19. Distribution of *Reading* Stimuli Across Levels of Grade-Appropriate Difficulty Ratings: Grade 4

	# of stimuli	Percentage of Stimuli		
		Below grade level	On grade level	Above grade level
NAEP-R	18	0%	83%	17%
SA1	14	36%	64%	0%
SA2	4	0%	100%	0%
SA3	8	0%	75%	25%
SA4	5	0%	100%	0%

NOTE: SA=state assessment

Table 20. Distribution of *Reading* Stimuli Across Levels of Grade-Appropriate Difficulty Ratings: Grade 8

	# of stimuli	Percentage of Stimuli		
		Below grade level	On grade level	Above grade level
NAEP-R	19	5%	95%	0%
SA1	13	33%	67%	0%
SA2	5	0%	100%	0%
SA3	9	11%	89%	0%
SA4	6	17%	83%	0%

NOTE: SA=state assessment

Engagingness and Diversity of Perspectives. For the attributes of engagingness and diversity of perspectives, *Reading* stimuli were rated on a 3-point scale of "none," "some," or "a lot." ("Diversity of perspectives" captures the extent to which the stimuli reflect diverse experiences, people, and perspectives.) Tables 21 and 22 show the percentages of *Reading* stimuli that were rated as exhibiting some or a lot of these characteristics, for Grades 4 and 8, respectively.

Table 21. Percentage of *Reading* Stimuli Rated as Exhibiting "Some" or "a Lot" of Engagingness and Diversity of Perspectives: Grade 4

		Percentage of Stimuli	
	# of stimuli	Engagingness	Diversity of Perspectives
NAEP R	18	94%	61%
SA1	14	64%	50%
SA2	4	100%	25%
SA3	8	63%	75%
SA4	5	100%	100%

NOTE: SA=state assessment

Table 22. Percentage of *Reading* Stimuli Rated as Exhibiting "Some" or "a Lot" of Engagingness and Diversity of Perspectives, Grade 8

		Percentage of Stimuli	
	# of stimuli	Engagingness	Diversity of Perspectives
NAEP R	19	90%	74%
SA1	13	53%	20%
SA2	5	80%	100%
SA3	9	0%	0%
SA4	6	67%	17%

NOTE: SA=state assessment

At Grade 4, the percentage of NAEP-R stimuli rated as having some or a lot of engagingness was comparable to SA2 and SA4, but higher than the other two state assessments. At Grade 8, notably, the percentage of engaging stimuli was quite a bit higher for NAEP-R than any of the state assessments, although SA2 came closest, with 80 percent judged to be engaging.

The percentages of *Reading* stimuli rated as reflecting some or a lot of diversity of perspectives ranged from 0 to 100 across assessments. At Grade 4, NAEP-R had a higher percentage than two of the state assessments (SA1 and SA3) and a lower percentage than the other two. At Grade 8, SA2 had the highest percentage of passages with diversity of perspectives, followed by NAEP-R. The other three state assessments had percentages that were considerably lower.

Genre. Tables 23 and 24 present information at each grade level for the percentage of *Reading* stimuli reflecting three genres—informational, literary, and poetry.

Similar to the state assessments (except for SA2, Grade 4), NAEP-R included a higher percentage of informational reading stimuli than stimuli from either of the other genres, a finding consistent with recommendations in college- and career-ready standards and the NAEP Reading Framework. Across assessments, the highest percentage of informational texts was found in SA4 at both grade levels. For poetry, the percentages were more varied. NAEP-R was the only assessment that included poetry at both grade levels. Only one of the state assessments (SA3) included poetry at Grade 4, and only two (SA1 and SA2) included it at Grade 8.

Table 23. Distribution of *Reading* Stimuli Across *Genre*: Grade 4

		Percentage of Stimuli		
	# of stimuli	Informational	Literary	Poetry
NAEP-R	18	56%	28%	17%
SA1	14	50%	50%	0%
SA2	4	40%	60%	0%
SA3	8	50%	38%	15%
SA4	5	70%	10%	0%

NOTE: SA=state assessment

Table 24. Distribution of *Reading* Stimuli Across *Genre*: Grade 8

		Percentage of Stimuli		
	# of stimuli	Informational	Literary	Poetry
NAEP-R	19	58%	32%	11%
SA1	13	53%	40%	7%
SA2	5	50%	25%	25%
SA3	9	56%	44%	0%
SA4	6	70%	10%	0%

NOTE: SA=state assessment

Reading Stimuli Summary. Overall, NAEP-R stimuli aligned with other state assessments and college and career readiness expectations in terms of difficulty and emphasis on informational text (genre). The presence of above-grade-level stimuli at Grade 4 and below-grade-level stimuli at Grade 8 are most likely associated with the cross-grade blocks (Grades 4 and 8) included in the NAEP-R design. Panelists considered NAEP-R stimuli at both Grades 4 and 8 to be highly engaging compared with several of the state assessments, especially at Grade 8. The percentage of NAEP-R stimuli rated high for diversity of perspectives was in the midrange.

Writing

Writing stimuli were defined as the prompts used to elicit writing from students.

Grade-Appropriate Difficulty. Tables 25 and 26 provide information on the percentages of Grade 4 and Grade 8 *Writing* stimuli that were rated by the panelists as below, on, or above grade level. For all the assessments, the overwhelming majority of writing stimuli were rated as on grade level, although SA3 and SA4 have some writing stimuli (40 percent and 20 percent, respectively) rated above grade level at Grade 4, and NAEP-W and SA4 have some rated below grade level at one grade or another (12 percent at Grade 4 for NAEP-W and 17 percent at Grade 8 for SA4).

Table 25. Distribution of *Writing* Stimuli Across Levels of Grade-Appropriate Difficulty Ratings: Grade 4

		Percentage of Stimuli		
	# of stimuli	Below grade level	On grade level	Above grade level
NAEP-W	8	12%	88%	0%
SA1	0	0%	0%	0%
SA2	2	0%	100%	0%
SA3	5	0%	60%	40%
SA4	5	0%	80%	20%

NOTE: SA=state assessment

Table 26. Distribution of *Writing* Stimuli Across Levels of Grade-Appropriate Difficulty Ratings: Grade 8

		Percentage of Stimuli		
	# of stimuli	Below grade level	On grade level	Above grade level
NAEP-W	9	0%	100%	0%
SA1	2	0%	100%	0%
SA2	5	0%	100%	0%
SA3	5	0%	100%	0%
SA4	6	17%	83%	0%

NOTE: SA=state assessment

Engagingness and Diversity of Perspectives. Tables 27 and 28 present the percentages of *Writing* stimuli that were rated as having a lot or some engagingness and diversity of perspectives at each grade level. NAEP-W was the only assessment in which all the prompts were rated at these levels for engagingness at both Grades 4 and 8. In terms of diversity of perspectives, NAEP *Writing* prompts fell in the midrange at both grade levels compared with the state assessments.

Table 27. Percentage of *Writing* Stimuli Rated as Exhibiting “Some” or “a Lot” of Engagingness and Diversity of Perspectives: Grade 4

		Percentage of Stimuli	
	# of stimuli	Engagingness	Diversity of perspectives
NAEP-R	8	100%	75%
SA1	0	0%	0%
SA2	2	100%	50%
SA3	5	60%	80%
SA4	5	80%	100%

NOTE: SA=state assessment

Table 28. Percentage of *Writing* Stimuli Rated as Exhibiting "Some" or "a Lot" of Engagingness and Diversity of Perspectives: Grade 8

		Percentage of Stimuli	
	# of stimuli	Engagingness	Diversity of perspectives
NAEP-R	9	100%	89%
SA1	2	100%	100%
SA2	5	80%	100%
SA3	5	0%	0%
SA4	6	67%	17%

NOTE: SA=state assessment

Comparison of NAEP-R SBTs and Traditional Reading Blocks

The NAEP-R items and stimuli rated in this study included traditional blocks from the 2017 operational NAEP-R as well as four SBTs, one literary and one informational at each grade, which were piloted in 2017 for use in the 2019 NAEP-R operational assessment. In this section, we examine similarities and differences between the traditional blocks and SBTs.

The SBTs differ from the traditional NAEP-R blocks in several ways. The SBTs begin by presenting students with a specific purpose and task (e.g., building knowledge to create a website or a poster for a science fair). Students are then guided through the task using features that provide scaffolding, including avatars (in some cases), continuous directions, and feedback designed to affirm or reset student thinking. The traditional blocks, by contrast, simply present the stimuli and the items, leaving students to proceed through the task as they see fit rather than in the controlled manner in which they are guided through the SBT.

Two of the four state assessments in this study included performance or *WS* tasks with some of the same features as the NAEP-R SBTs. Common features include a stated purpose, multiple linked stimuli that sometimes include multimedia, and items that require synthesis of ideas and information across stimuli. The NAEP-R SBTs were included in this study because they will be part of NAEP-R assessments starting in 2019, and they share similarities with the performance and/or *WS* tasks in the state assessments. With these points in mind, a comparison of NAEP-R SBTs and traditional blocks was conducted to assist NAEP-R in thinking about future development of SBTs and other types of performance tasks that have become prominent in state assessments.

Comparison of Importance and Complexity for SBT and Traditional Blocks of Items

The results of the comparison between SBTs and traditional blocks of items in NAEP-R indicate that SBT items were rated higher on the measures of Importance and Complexity at both Grades 4 and 8. The specific differences are as follows.

Importance. SBTs at both grade levels had no *Reading* score points from items rated unimportant (Level 1); this compares with 12 percent and 6 percent of Level-1 score points at Grades 4 and 8, respectively, for traditional blocks. Conversely, the percentage of *Reading* score points coming from items rated at Level 3 was greater for SBTs than traditional blocks at both grades (Grade 4: 59 percent versus 46 percent; Grade 8: 82 percent versus 53 percent).

Importance by Reading Subdomain. At both grades, Level-1 *Reading* score points from traditional items were located in the *key ideas/details* and *vocabulary/language* subdomains; as noted, SBT items generated no Level-1 score points in any subdomains. Conversely, the percentages of Level-3 score points in the *craft/structure* and i*ntegrate/analyze* subdomains were greater for SBTs at both grades (*craft/structure*: Grade 4, 50 percent versus 17 percent; Grade 8, 100 percent versus 43 percent; *integrate/analyze*: Grade 4, 91 percent versus 73 percent; Grade 8, 100 percent versus 82 percent).

Complexity. The percentage of Level-1 *Reading* score points was lower for SBTs than traditional blocks at both grades (Grade 4: 20 percent versus 37 percent; Grade 8: 0 percent versus 28 percent). Only a small percentage of score points derived items that were rated at Level 4 for Complexity in either SBTs or traditional reading blocks. However, the percentage of score points from items rated at Level 2 or 3 was greater for SBTs than traditional blocks at both grades (Grade 4: 76 percent versus 61 percent; Grade 8: 90 percent versus 54 percent).

Complexity by Reading Subdomain. The percentage of *Reading* score points coming from *key ideas/details* items rated at Level 1 was greater for traditional blocks than SBTs at both grades (Grade 4: 65 percent versus 35 percent; Grade 8: 55 percent versus 0 percent). As mentioned above, only a small percentage of score points derived from items rated at Level 4 were found in either the SBTs or traditional blocks. However, all of these Level-4 items fell into the *craft/structure* or *integrate/analyze* categories for both types of blocks and at both grades.

Comparison of Characteristics of SBT Stimuli and Traditional Stimuli

The results of the comparison between SBT and traditional stimuli in NAEP-R indicate that SBT stimuli were rated higher than stimuli in traditional blocks on the measures of grade appropriateness, diversity of perspectives, and engagingness at both Grades 4 and 8. The specific differences are as follows.

Grade Appropriateness. A greater percentage of stimuli were rated as grade appropriate for SBTs than traditional blocks at both grades. The SBT stimuli were rated as 100 percent grade appropriate at both grade levels, while 77 percent of the traditional Grade 4 stimuli and 93 percent of the traditional Grade 8 stimuli earned this rating. It also is interesting to note that 23 percent of the traditional stimuli at Grade 4 were rated as above grade level and 7 percent of the traditional stimuli at Grade 8 were rated as below grade level, while no SBT stimuli were rated as either above or below grade level. This is likely due to the presence of cross-Grade 4/8 blocks in the traditional item pool and the absence of cross-grade SBTs.

Engagingness. A larger percentage of SBT than traditional stimuli were rated at higher levels of engagingness at both grades. At Grade 4, the distribution of ratings for SBT stimuli were 0 percent, 20 percent, and 80 percent, respectively, for none, some, and a lot of engagingness. This compares with the distribution of ratings for traditional stimuli, which were 8 percent, 46 percent, and 46 percent, respectively. At Grade 8, the corresponding percentages were 0 percent, 25 percent, and 75 percent, respectively, for SBT stimuli and 13 percent, 87 percent, and 0 percent, respectively, for traditional stimuli.

Diversity of Perspectives. In general, a larger percentage of stimuli were rated high for reflecting diversity of perspectives for SBTs than for traditional blocks at both grades. At Grade 4, 46 percent of the traditional stimuli were rated as not reflecting diversity of perspectives compared with 20 percent of the SBT stimuli. However, it also should be noted that 23 percent of the traditional stimuli were rated as strongly reflecting diversity of perspectives compared with 0 percent of the SBT stimuli. At Grade 8, the ratings for traditional stimuli distributed as 33 percent, 47 percent, and 20 percent for little, some, and strong reflections of diversity of perspectives, respectively, compared with 0 percent, 50 percent, and 50 percent, respectively, for SBT stimuli.

Comparison of SBT and Traditional NAEP-R Summary. Overall, the new NAEP-R SBT items were rated more favorably than the traditional NAEP-R items at the corresponding grade level in the areas of Importance and Complexit*y*. SBT stimuli also were rated more favorably than their traditional counterparts in the areas of grade-level appropriateness, engagingness, and diversity of perspectives. Panelists also commented favorably about the SBT approach to assessing reading.

OVERALL STUDY CONCLUSIONS AND THEIR IMPLICATIONS FOR NAEP FRAMEWORKS AND THE DESIGN OF NAEP READING AND WRITING ASSESSMENTS

The purpose of this study was to answer the following question:

> *To what extent are NAEP reading (NAEP-R) and NAEP writing (NAEP-W) assessments similar to/different from reading and writing assessments in current use by states?*

To assist in unpacking this multifaceted question, the ELA Expert Judgement Study team considered a number of subquestions that spanned areas specific to the reading and writing assessments. Below, we address these questions by summarizing findings and implications for possible modifications to NAEP frameworks and/or assessments. Several of these conclusions reach across subquestions to highlight their interrelatedness. in conceptions of the domains and as used in measuring reading and writing achievement.

Content

How does the balance of content assessed on the NAEP-R and NAEP-W assessments compare with that of state ELA assessments?

- **Conclusion 1 (Reading and Writing)**: NAEP administers separate assessments and produces separate scores for reading and writing. In contrast, each of the state assessments analyzed for this study produces a total test score for ELA that is based on items assessing reading, writing, and, often, other ELA skills. Therefore, student performance on NAEP-R or NAEP-W cannot be compared with the total ELA scores reported by any of the state assessments analyzed for this study.

  Implications: The lack of content alignment between NAEP (both NAEP-R and NAEP-W) and the items that are included in the total ELA scores produced by state assessments analyzed for this study makes comparisons between total scores on state assessments and NAEP scores inappropriate. Valid comparisons may be possible using subsets of reading items or reading subtest scores from state assessments with NAEP-R scores. The content analysis of reading items conducted for this study suggests that such a targeted comparison of reading performance could be valid, but it would need to be tested empirically. Comparisons between NAEP-W and subsets of writing items or writing subtest scores from state assessments are not recommended due to differences in the types of writing tasks assessed (see Conclusion 2).

- **Conclusion 2 (Writing)**: Major differences were identified between NAEP-W items (*writing without sources [WO]*) and the vast majority of writing items on the state assessments (*writing with sources [WS]*). *WS* is aligned with the college and career readiness emphasis on integrating reading and writing and conducting research and inquiry. On the state assessments reviewed for this study, *WS* was almost always associated with informational text. In addition, at both Grades 4 and 8, items that called for *WS* were generally rated as more Important and Complex than NAEP-W *WO*. Consequently, state assessments measure some writing processes and skills that are currently missing from NAEP-W.

Implications: This analysis, as well as panelists' narrative feedback, strongly suggest that NAEP should pursue strategies for assessing *WS* on NAEP-R and/or NAEP-W—these types of tasks are consistent with current curricular emphases and are likely to be rated as more Important and Complex than *WO*. Because *WS* is often associated with longer "performance" tasks, the current NAEP block design (30 minutes) would likely need to be restructured to enable the inclusion of such tasks. At the same time, current trends in the field of writing and panelists' feedback suggest that NAEP should continue to assess some forms of *WO*. NAEP will need to address these different types, purposes, audiences, and processes for writing in revisions to NAEP frameworks and assessments for both *Reading* and *Writing*.

- **Conclusion 3 (Reading and Writing):** NAEP-R and NAEP-W do not currently include items similar to those classified as *Conventions/Research Skills/Language* on the state assessments reviewed for this study. These were stand-alone items that were not linked to either reading stimuli or writing prompts.

  Implications: It is our opinion that NAEP-R and NAEP-W should not include items that assess the subdomains of *conventions, research skills,* or *language* in isolation, even though these item types are present in some of the state assessments reviewed for this study. However, the panelists, supported by evidence from researchers in the field, urge NAEP to investigate new avenues for assessing students' research strategies and critical literacy skills, which include their proficiency in searching for, evaluating, and using online digital sources. This multimodal, online critical literacy is essential to college and career readiness and is already part of curricula in many schools. Panelists agreed that, if NAEP is to continue to keep pace and lead, it should address these skills in the revised frameworks and assessments. The SBTs may provide a basis from which to extend to digital critical literacy; the current block design of NAEP may need to be reexamined to allow for these longer, multilayered tasks.

  Conclusion 4 (Reading): Distributions across the four *Reading* subdomains (*key ideas/details, craft/structure, integrate/analyze,* and *vocabulary/language*) varied across the five assessments. Compared with the state assessments, NAEP had the greatest proportion of *Reading* score points coming from the subdomain of *integrate/analyze*. This emphasis is consistent with the emphasis found in the NAEP Reading Framework and college and career readiness standards.

  Implications: Although there are no agreed-upon criteria in the reading field for the proportion of items that should be devoted to different subdomains, NAEP's emphasis on the *Reading* subdomain of *integrate/analyze* should be maintained—it is consistent with college- and career-ready goals for deeper comprehension (higher ratings on Complexity*)* and with two of the three cognitive targets (integrate/interpret, critique/evaluate) in the NAEP Reading Framework. Furthermore, panelists recommended *additional* emphasis on items that assess critical evaluation, including evaluation of sources linked to research. Although none of the assessments examined here has moved into these new areas, as noted under Conclusion 3, panelists felt strongly that NAEP should address them in the frameworks and in item development.

Importance

How well do NAEP-R and NAEP-W items compare with those of state ELA assessments in terms of focus on the most important aspects of the domains and/or the goals of college and career readiness standards that the test is designed to measure?

- **Conclusion 5 (Reading):** Panelists' ratings of Importance revealed that the overwhelming majority of NAEP-R items targeted understanding of important concepts and information in texts. This is a strength that also is found in three of the four state assessments at both Grades 4 and 8. For all tests, high percentages of Level 1 (unimportant) *Reading* score points were categorized in the subdomain of *vocabulary/language*. These items tended to focus on words or phrases that were less important to understanding central ideas in the associated text, both on NAEP-R and on the state assessments. The focus on noncentral vocabulary in NAEP-R is, however, consistent with the current NAEP Reading Framework.

  Implications: NAEP should review the definitions, descriptions, and item specifications for assessing vocabulary specified in the Reading Framework to determine if they are aligned with current thinking, research, and other assessments. If changes are warranted, they should be made in the framework and in the assessment.

- **Conclusion (Writing):** See Conclusion 2 above under Content.

Complexity

How well do NAEP-R and NAEP-W items compare with those on state ELA assessments in terms of content-specific Complexity or depth of understanding/processing required by items on each assessment?

- **Conclusion 6 (Reading):** The Complexity and depth of processing required by NAEP-R items compares favorably with the Complexity of reading items on the state assessments. Specifically, NAEP-R has better coverage of the full range of Complexity ratings. More than 30 percent of NAEP-R score points came from items rated at the higher end of the Complexity scale (Levels 3 and 4), a balance that is supported by college and career readiness standards. Furthermore, NAEP-R is the only assessment in the study to have items rated at the highest level – Level 4—for Complexity and depth of understanding. The panelists felt it was likely that even more of the NAEP-R items would have been rated at Level 4 if the scoring guides for the constructed-response items were more rigorous. Items rated as low in Complexity were categorized predominantly as *vocabulary/language* or *key idea/details*, both in NAEP-R and in all the state assessments examined.

  Implications: NAEP should conduct a review of scoring rubrics and anchor papers associated with extended and short constructed-response reading items to be sure that Complexity and depth of processing expectations are consistent with the intent of individual items. In addition, reading items in the subdomains of *vocabulary/language* and *key ideas/details* (across all item formats) should be reviewed to determine if low Complexity ratings (Levels 1 and 2) are consistent with the intent of these items. NAEP should be sure that these particular items are not inadvertently testing a different level of Complexity than originally intended. NAEP should review its item development specifications to be sure that items represent a range of complexity (1–4).

- **Conclusion (Writing):** See Conclusion 2 above under Content.

Other Test Features

How do NAEP-R and NAEP-W assessments compare with state assessments with regard to other item features (e.g., item format, scoring rubrics) and stimuli characteristics (e.g., reading passage and writing prompt engagingness, diversity of perspectives, and difficulty)?

- **Conclusion 7 (Reading):** The analysis produced a number of conclusions and recommendations related to NAEP-R reading selections: (a) NAEP passages were judged to be highly engaging for students and representative of a moderate range of diverse perspectives; they also compared favorably with the other assessments. (b) The majority of NAEP-R passages were judged to be aligned with grade-level expectations for difficulty. However, the presence of cross-grade blocks introduced more challenging texts at Grade 4 and easier texts at Grade 8. (c) Similar to the state assessments in this study, NAEP-R uses a variety of genre/subgenres and is the only assessment to include poetry at both Grades 4 and 8.

  Implications: NAEP should continue to make text selection a priority, with an emphasis on including material from across the narrative, informational, and poetry genres as well as texts that represent diversity of perspectives. Panelists also encouraged NAEP to increase the presence of digital and multimodal stimuli (also see Conclusion 3, above, under Content). Findings from this study, together with evidence from prior studies about floor effects at Grade 4, suggest that NAEP should reexamine the use of cross-grade blocks.

- **Conclusion 8 (Reading)**: Compared with the majority of the state assessments, NAEP-R has fewer items that require students to read and respond across three or more texts. This multitext structure can place more demands on reading comprehension and is highlighted in college and career readiness standards, especially with respect to building knowledge from text and conducting research.

  Implications: NAEP should expand the current practice of pairing passages to include more and different cross-text tasks (similar to the new SBTs). These linked passages also could be used as text for items in the subdomains of *writing with sources* and *research,* and they are associated with more complex reading and writing.

- **Conclusion 9 (Reading/Writing)**: NAEP uses a variety of formats to assess reading. It does not use the two-part selected-response item format that was dominant on two of the state assessments, but compared with the other assessments, NAEP-R includes a greater proportion of extended and short constructed-response items. This is consistent with the NAEP Reading Framework. These items, which are scored and reported as part of NAEP-R (not NAEP-W), were most often associated with higher levels of reading Complexity and Importance*,* making them particularly useful in assessing the higher levels of comprehension targeted by college and career readiness standards.

  Implications: NAEP should maintain its emphasis on constructed-response items in NAEP-R. It should investigate possible strategies for restructuring these items to serve as vehicles for assessing the integration of reading and writing, conducting *research*, and/or *writing with sources*. NAEP should not emulate the two-part selected response item format (Part A, Part B) that is dominant on two of the state assessments reviewed for this study; panelists found these items confusing and they were often rated low on Complexity.

Newer NAEP Formats

How do traditional NAEP-R reading blocks compare with the new scenario-based tasks (SBTs) with regard to item features and stimuli characteristics?

- **Conclusion 10:** The comparison between the SBTs and traditional blocks on NAEP-R indicated that SBTs provide increased opportunities for items to assess Important content and more Complex understanding. Panelists also noted that SBT tasks are more similar to performance/*WS* tasks on the state assessments (e.g., purpose driven, multiple stimuli, cross-stimuli items) than are tasks in traditional NAEP-R or NAEP-W blocks. SBTs also may provide NAEP with opportunities to integrate the assessment of reading and writing.

  Implications: NAEP should continue to develop and implement reading blocks that use new formats similar to SBTs or other alternatives that prioritize purpose-driven, performance-oriented, multisource tasks. These new task formats may provide opportunities to address *WS* (see Conclusion 2 above) and critical digital literacy skills (see Conclusion 3, above). However, as noted previously, test block designs may need to be reexamined to accommodate these longer tasks. Item development specifications should be sure to address a range of complexity in SBTs to monitor difficulty and challenge.

In sum, the purpose of this study was to examine the similarities and differences between the NAEP reading and writing assessments and a sample of state ELA assessments currently in use by states. The intent behind this study was to generate information that can inform considerations of whether, and to what extent, NAEP might need to update its frameworks and assessments in order to continue as a valid and useful monitor of student achievement in reading and writing given changes in curriculum and assessments that draw on new college and career readiness standards. The results and implications of this study are somewhat different for NAEP-R and NAEP-W. For *Reading*, the results indicate that, although NAEP-R items are representative of the types of reading assessment items found on this sample of tests, state assessments are no longer assessing reading in isolation from other areas of ELA. For *Writing*, the results indicate that NAEP-W is *not* representative of the writing measures on these assessments, as states are primarily assessing *WS* rather than *WO*. The results for both NAEP-R and NAEP-W need to be taken into consideration in the future development of NAEP frameworks and assessments.

APPENDIX A. ENGLISH LANGUAGE ARTS (ELA) EXPERT PANELISTS

Core Group

Sally Hampton	National Writing Project (retired)
Christy Howard	Assistant Professor, Eastern Carolina University
Don Leu	Professor, University of Connecticut
Annemarie Palincsar	Professor, University of Michigan

Review Panel

Melissa Adams-Budde	Assistant Professor, West Chester University of Pennsylvania
Eurydice Bauer	Professor, University of South Carolina
Gina Biancarosa	Associate Professor, University of Oregon
Jensa Bushey	Literary Coach, Shelburne Community School, Vermont
Jill Castek	Associate Professor, University of Arizona
Brady Donaldson	Language Arts Coach, Salt Lake City Schools
Elizabeth Dutro	Associate Professor, University of Colorado
Georgia (Joey) Garcia	Professor, University of Illinois, Urbana-Champaign
Virginia (Ginny) Goatley	Professor, University at Albany, SUNY
Christy Howard	Assistant Professor, Eastern Carolina University
Julie Learned	Assistant Professor, University at Albany, SUNY
Tamera K. Lipsey	Reading Coordinator, Nashville Public Schools
Jeannette Mancilla-Martinez	Associate Professor, Vanderbilt University
Nicole Martin	Assistant Professor, Ball State University
Sharon O'Neal	Associate Professor, Texas State University
Nate Phillips	Assistant Professor, University of Illinois, Chicago
Laura Roop	Director, Western Pennsylvania Writing Project, University of Pittsburgh
Nancy Roser	Professor Emeritus, University of Texas at Austin
Laura Schiller	Director, Oakland Writing Project (retired)
Amy Vetter	Associate Professor, University of North Carolina, Greensboro
Brandon Wallace	Director of Special Education, National Office, Urban Teachers
Rebecca Woodard	Assoc. Professor, University of Illinois, Chicago
Victoria Young	Texas Education Agency (retired)
Melody Zoch	Assistant Professor, University of North Carolina, Greensboro

APPENDIX B. CONSOLIDATED CONTENT FRAMEWORK FOR ENGLISH LANGUAGE ARTS (ELA)

Consolidated Content Framework for ELA—Grade 4

	READING					WRITING	
	Understanding Central Ideas and Details	Understanding/ Using Author's Craft and Text Structure	Integration and Analysis of Knowledge and Ideas	Vocabulary and Language Meaning in Text	Conventions, Research Skills, and Language	Writing Qualities and Production Without Sources	Reading and Writing With Sources: Research and Literary Response
	R-T1 R-T2 R-T8 R-T9 RS-4	R-T6 R-T13	R-T4 R-T5 R-T11 R-T12	R-T3 R-T7 R-T10 R-T14	W-T8 W-T9	W-T1a&b W-T2 W-T3a&b W-T4 W-T6a&b W-T7 W-T8 W-T9 W-T10	RS1 RS2 RS3 RS4 W-T10
	RL1 RL2 RL3 RI1 RI2 RI3	RL5 RL6 RI5 RI6	RL7 RL9 RI7 RI8 RI9	RL4 RI4 L4 L5 L6	L6	W1 W2 W3 W4 W5 W6	W7 W8 W9 W10
	RL1 RL2 RL3 RI1 RI2 RI3	RL5 RL6 RLMA.8.A RI5 RI6	RL7 RL9 RI7 RI8 RI9	RL4 RI4 L3 L4 L5 L6	L1 L2 L3 L6	W1 W2 W3 MA.3.A W4 W5 W6	W7 W8 W9
	4.2.R.1 4.2.R.3 4.2.R4	4.2.R.2 4.3.R.1 4.3.R.2 4.3.R.3 4.3.R.4 4.3.R.6	4.3.R.5 4.3.R.7 4.6.R.2 4.6.R.3	4.4.R.1 4.4.R.2 4.4.R.3 4.4.R.4 4.4.R.5	4.4.W.1 4.4.W.2 4.5.R.1 4.5.R.2 4.5.R.3 4.5.R.4 4.5.R.5 4.5.W.1 4.5.W.2 4.5.W.3 4.5.W.4 4.2.W.3 4.2.W.4	4.2.W.1 4.2.W.2 4.2.W.3 4.2.W.4 4.3.W.1 4.3.W.2 4.3.W.3	4.2.W.4 4.6.R.1 4.6.R.2 4.6.R.3 4.6.W.1 4.6.W.2 4.6.W.3
NAEP*							

Reading Standards
Writing Standards
Language Standards

Consolidated Content Framework for ELA—Grade 8

READING

WRITING

	Understanding Central Ideas and Details	Understanding/ Using Author's Craft and Text Structure	Integration and Analysis of Knowledge and Ideas	Vocabulary and Language Meaning in Text	Conventions, Research Skills, and Language	Writing Qualities and Production Without Sources	Reading and Writing With Sources: Research and Literary Response
	R-T1 R-T2 R-T8 R-T9	R-T6 R-T13	R-T4 R-T5 R-T11 R-T12	R-T3 R-T7 R-T10 R-T14	W-T8 W-T9	W-T1a&b W-T2 W-T3a&b W-T4 W-T6a&b W-T7 W-T8 W-T9 W-T10	RS1 RS2 RS3 RS4 W-T10
	RL1 RL2 RL3 RI1 RST1 RH1 RI2 RST2 RH2 RI3 RST3 RH3	RL5 RL6 RI5 RI6 RST5 RH5 RST6 RST6	RL7 RL9 RI7 RI8 RI9 RH7 RH8 RH RST7 RST8 RST9	RL4 RI4 RH4 RST4 L3 L4 L5 L6	L6	W1 W2 W3 W4 W5 W6 W10	W7 W8 W9 WH7 WST7 WH8 WST8 WH9 WST9 W10
	RL1 RL2 RL3 RI1 RST1 RH1 RI2 RST2 RH2 RI3 RST3 RH3	RL5 RL6 RLMA.8. A RI5 RI6 RST5 RH5 RST6 RST6	RL7 RL9 RI7 RI8 RI9 RH7 RH8 RH RST7 RST8 RST9	RL4 RI4 RH4 RST4 L3 L4 L5 L6	L1 L2 L3 L6	W1 W2 W3 MA.3.A W4 W5 W6	W7 W8 W9 WH7 WST7 WH8 WST8 WH9 WST9
	8.2.R.1 8.2.R.3	8.3.R.1 8.3.R.2 8.3.R.3 8.3.R.4 8.3.R.6 8.2.R.2	8.3.R.5 8.3.R.7 8.6.R.2 8.6.R.3	8.4.R.1 8.4.R.2 8.4.R.3 8.4.R.4 8.4.R.5	8.2.W.5 8.4.W.1 8.4.W.2 8.5.R.1 8.5.R.2 8.5.R.3 8.5.R.4 8.5.W.1 8.5.W.2 8.5.W.3 8.5.W.4 8.5.W.5	8.2.W.1 8.2.W.2 8.2.W.3 8.2.W.4 8.2.W.5 8.3.W.1 8.3.W.2 8.3.W.3 8.3.W.4	8.2.W.1 8.2.W.2 8.2.W.3 8.2.W.4 8.2.W.5 8.6.R.1 8.6.R.2 8.6.R.3 8.6.W.1 8.6.W.2 8.6.W.3 8.6.W.4
NAEP*							

Reading Standards
Writing Standards
Language Standards

* NAEP uses reading and writing frameworks as the basis for developing its reading and writing assessments. These frameworks do not include specific assessment standards that are comparable to the standards provided by the other assessments in this study. Rather, they provide definitions and general guidelines for the development of the assessments, as described in the next section, NAEP Definitions.

NAEP Definitions

- Reading is an active and complex process that involves:
  - Understanding written text.
  - Developing and interpreting meaning.
  - Using meaning as appropriate to type of text, purpose, and situation. (p. 2)
- Writing is a complex, multifaceted, and purposeful act of communication that is accomplished in a variety of environments, under various constraints of time, and with a variety of language resources and technological tools. (p. 3)

The Reading Framework includes particular attention to three cognitive targets (locate/recall, integrate/interpret, critique/evaluate) and two text types (literary, informational). Descriptions and examples of how these dimensions are instantiated in the NAEP reading assessment can be found in Chapter 2 (pp. 15–43) of the Reading Framework. The Writing Framework includes particular attention to three communicative purposes for writing (to persuade, to explain, to convey experience), audience, and three key features of writing (development of ideas, organization of ideas, language facility and conventions) in response to a "stand-alone" writing prompt. (There is no writing in response to text.) Descriptions and examples of how these dimensions are instantiated in the NAEP writing assessment can be found in Chapter 2 (pp. 19–42) of the Writing Framework.

Reading Framework for the 2017 National Assessment of Educational Progress

Writing Framework for the 2017 National Assessment of Educational Progress

Retrieved from https://nces.ed.gov/nationsreportcard/assessments/frameworks.aspx

APPENDIX C. EXPANDED DEFINITIONS OF CONSOLIDATED CONTENT FRAMEWORK DOMAINS AND SUBDOMAINS

READING

R1. Understanding Central Ideas and Details. Items assess central/main ideas, summaries, explicit and implicit evidence to support conclusions, development of ideas, events, dialogue, or procedures over the course of the text in a variety of literary and informational contexts.

R2. Understanding/Using Author's Craft and Text Structure. Items assess author's use and impact of craft related to point of view, style, purpose, use of literary devices/elements, genre characteristics, and text features (including embedded multimedia, visuals, and graphics).

R3. Integration and Analysis of Knowledge and Ideas. Items assess understanding of text, including literacy themes, conceptual understandings, analysis, connections, synthesis, and/or evaluation of ideas leading to deep understanding of text within and/or across a variety of literary and informational texts, including texts presented in different media or formats and analytic research strategies.

R4. Vocabulary and Language Meaning in Text. Items assess understanding of academic and discipline-specific vocabulary and language and how specific language shapes understanding and tone; includes understanding of word relations (e.g., connotation/ denotation, fact/opinion), figurative language, technical language, and multi-meaning words as well as use of context strategies and specialized reference materials.

CONVENTIONS, RESEARCH SKILLS, AND LANGUAGE

Conventions, Research Strategies, and Language. Items evaluate conventions of standard English, including grammar, usage, capitalization, punctuation, and spelling (as found in editing exercises); use of research tools and strategies; and aspects of language, such as parts of speech and affixes/roots. (coded as C= Conventions, R=Research, L=Language)

WRITING

W1. Writing Qualities and Production Without Sources. Prompts/items evaluate one or more of the following: writing for specific audiences and purposes; appropriate use of writing qualities, such as development of ideas, clarity, coherence, organization, transitions, word choice, and conventions; and application of the writing process, including planning, drafting, revision, editing, and so on.

R/W2. Reading and Writing With Sources—Research and Literary Response. Prompts/items evaluate one or more of the following: writing for specific audiences and purposes; integration, analysis, or evaluation of information/evidence from sources within the writing; appropriate use of writing qualities, such as development of ideas, clarity, coherence, organization, transitions, word choice, and conventions; and application of the writing process, including planning, drafting, revision, editing, and so on.

APPENDIX D. IMPORTANCE AND COMPLEXITY RUBRICS

Reading

Table D1. Example Importance Rubric: *Reading* Importance

FOCUS & IMPORTANCE

Consider the focus and importance of the item with regard to students developing a deep understanding of the important ideas, concepts, and content of the text(s)

Level	Description
Level 1	• Item assesses understanding of minor or unimportan**t** ideas, concepts, and/or information in the text(s)s
Level 2	• Item assesses understanding of ideas, concepts, and/or information of some importance that are related or helpful to understanding parts of the text(s)
Level 3	• Item assesses understanding of ideas, concepts, and/or information that are important to building a coherent and deep understanding of the text/s; for narratives, this is often related to plot, character development, theme; for informational text, this is often related to major concepts and key ideas

Table D2. Example Complexity Rubric: *Reading* Complexity

DEPTH & COMPLEXITY

For each item, consider the complexity of the process students use to think across the item stem, text, and distractors (for SSR items) in order to select the correct answer or earn full-credit using the test scoring rubric.

Level	Description
Level 1	• Predominantly explicit information or simple inference within sentence, single paragraph* • Concrete content in text/s, item, rubric; little/no abstract reasoning • Very small amount of source text needed (e.g., 1 sentence or within 1 paragraph) • Understanding of literary or rhetorical features not required • 0-1 distractors are plausible or attractive based on prior knowledge or the text/s (may be N/A for CR items)
Level 2	• Several inferences from text segment or in multiple spots in text/s (may include summarizing) • Mix of concrete & some abstract reasoning in item, text/s, rubric • Amount of text needed – several paragraphs, often contiguous • Some understanding of single literary or rhetorical feature • 1-2 distractors may be plausible/attractive based on prior knowledge or the text/s (may be N/A for CR items);
Level 3	• Inferences require drawing conclusions, generalizations, synthesis, or analysis (e.g. theme), or a simple level of evaluation • Predominantly abstract content or reasoning in stem, text/s, rubric • Amount of text needed – more than 3 contiguous paragraphs or several non-contiguous paragraphs/places within the text/s; may be simple cross-text question • Understanding of several literary or rhetorical features across paragraphs • One or more distractors may be plausible/attractive based on prior knowledge or the text (may be N/A for CR items; distractors may contribute to complexity)
Level 4	• Analysis with critical evaluation, more complex reasoning, more pieces of evidence and/or alternative perspectives in single text OR Level 3 inferences across 2 or more texts • Abstractness - Level 3 across 2 or more texts OR application of concepts to new idea/content • Amount of text - Level 3 across 2 or more texts OR cohesive, integrated understanding of entire selection • Deep understanding of literary or rhetorical features across multiple aspects of the text OR across multiple texts • Multiple distractors may be plausible and/or attractive based on prior knowledge or the text (may be N/A for CR items; distractors may contribute to complexity)

* References to single paragraph or multiple paragraphs refer to "typical" paragraph (in contrast to dialogue). When there is dialogue, consider "typical" amount of text in a paragraph for this age/grade level.

Conventions, Research Strategies, and Language

Table D3. Example Importance Rubric: *Conventions, Research Strategies, and Language* Importance

FOCUS & IMPORTANCE

Consider the importance and focus of the item with regard to the skill, strategy, or understanding it is intended to represent. How well does the item content and format assess an important focus of instruction and assessment in this particular aspect of writing, reading, or research at this grade level?

Level	Description
Level 1	• Targeted skill or strategy represents a relatively unimportant aspect of conventions (e.g. spelling, grammar, punctuation), language, or research strategies, etc. it is intended to represent. • Targeted skill or strategy is tested in a decontextualized way that is not authentic or readily transferable to the writing, reading, or research processes • Targeted skill or strategy is relatively unimportant to effective writing, reading, or research at this grade level
Level 2	• Targeted skill or strategy represents a somewhat important aspect of conventions (e.g. spelling, grammar, punctuation), language, or research strategies, etc. it is intended to represent. • Targeted skill or strategy is tested in a way that is somewhat related to an authentic writing, reading, or research situation. • Targeted skill or strategy somewhat important to effective writing, reading, or research at this grade level.
Level 3	• Targeted skill or strategy represents an essential aspect of conventions (e.g. spelling, grammar, punctuation), language, or research strategies, etc. it is intended to represent. • Targeted skill or strategy is tested in an authentic context that could be transferable to the writing, reading, or research process. • Targeted skill or strategy is very important to effective writing, reading, or research at this grade level.

Table D4. Example Complexity Rubric: *Conventions, Research Strategies, and Language* Complexity

DEPTH & COMPLEXITY

For each item, consider the thinking and reasoning related to writing, reading, and/or research that students need to select the correct response or earn full credit using the test scoring rubric.

Level	Description
Level 1	• Item generally requires rote skill, recall, or explicit knowledge • Item doesn't require comprehension of any text that may accompany it • 0-1 distractors are plausible or attractive based on prior knowledge or the text/s (may be N/A for CR items)
Level 2	• Item requires understanding of a single or simple feature of the writing, reading, and/or research tool to select correct answer or obtain full credit • Item may require mostly literal comprehension of text accompanying it • 1-2 distractors may be plausible/attractive based on prior knowledge or the text/s (may be N/A for CR items)
Level 3	• Item requires understanding of several features of the writing, reading, and/or research tool to select correct answer or obtain full credit • Item may require inferential comprehension of the text accompanying it to select correct answer or earn full credit • 1 or more distractors may be plausible/attractive based on prior knowledge or the text; distractors may contribute to complexity (may be N/A for CR items)
Level 4	• Item requires analysis and complex understanding of features of the writing, reading, and/or research tool to select correct answer or earn full credit • Item may require deeper understanding of the text(s) accompanying it to select correct answer or earn full credit • Multiple distractors may be plausible and/or attractive based on prior knowledge or the text; distractors may contribute to complexity (may be N/A for CR items)

Writing WITHOUT Sources

Descriptions that begin with "prompt" or "rubric" are used to evaluate test items that require students to write brief or extended responses.

Descriptions that begin with "item" are used to evaluate test items that are in SSR or multiple-SR formats.

Stimuli for short or extended writing prompts may include photos, videos, illustrations, and some limited text.

Table D5. Example Importance Rubric: *Writing WITHOUT Sources* Importance

FOCUS & IMPORTANCE

Consider the importance and focus of the prompt or item with regard to authentic high quality, grade-level expectations for writing without sources. How well does the item represent an important focus for instruction and assessment at this grade level?

Level	Description
Level 1	• Prompt does not provide an authentic grade-level appropriate, writing task • Prompt does not specify or clearly imply an appropriate purpose, genre, or audience for the writing ------------------------- – Item assesses a relatively unimportant skill, strategy, or quality of writing for the targeted grade level – Item presents skills, strategy, or quality in a way that is decontextualized from an authentic writing context – Item does not specify or clearly imply an appropriate purpose, genre, or audience for the writing.
Level 2	• Prompt provides a somewhat authentic, grade-level appropriate, writing task • Prompt may specify or clearly imply an appropriate purpose, genre, or audience but these are not needed to earn full credit for the task ------------------------- – Item assesses a somewhat important skill, strategy, or quality of writing for the targeted grade level – Items presents skill, strategy, or quality in a way that is somewhat related to authentic writing context – Item may specify or clearly imply an appropriate purpose, genre, or audience for the writing but these are not needed to respond correctly to the item
Level 3	• Prompt provides an authentic, grade-level appropriate, writing task • Prompt specifies or clearly implies an appropriate purpose, genre, or audience for the writing ------------------------- – Item assesses an essential skill, strategy, or quality of writing for the targeted grade level – Item presents skills, strategy or quality in a way that is authentic or applicable to actual writing – Item specifies or clearly implies an appropriate purpose, genre, or audience for the writing that need to be considered to respond correctly to the item

Table D6. Example Complexity Rubric: *Writing WITHOUT Sources* Complexity

DEPTH & COMPLEXITY

For each writing prompt, consider the complexity of the thinking and processes students use to work through the prompt and then craft a response that earns full credit based on the rubric used to score the item.

For each item consider the complexity of the process students use to think through the specific item stem and the distractors (for SSR items) in order to select the correct answer or earn full credit using the test scoring rubric.

Level	Description
Level 1	• Rubric does not include attention to purpose, genre, or audience • Rubric includes little attention or general, generic attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) • Space provided suggests limited response - approximately 1 paragraph or less • First draft writing with no expectations, direction, and/or time for editing and/or revision – Item generally requires rote skill, recall, or explicit knowledge – Item doesn't include attention to purpose, genre, or audience; these are not needed to select correct answer or earn full credit – 0-1 distractors are plausible or attractive based on prior knowledge or the text/s.
Level 2	• Rubric requires some attention to purpose, genre, or audience • Rubric includes specific attention to some of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level) • Space provided suggests short response- greater than a single paragraph • First draft writing with some expectations, directions, and/or time for minor editing (editing may focus on conventions) – Item requires understanding of a single or simple feature of the writing or text provided in the item to select correct answer or obtain full credit – Item includes attention to purpose, genre, or audience but these may not be relevant or need to be considered to select correct answer or earn full credit – 1 or more distractors may be plausible/attractive based on prior knowledge or the text/s; distractors may contribute to complexity
Level 3	• Rubric includes specific attention to purpose, genre, or audience • Rubric includes specific attention to most of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level) • Longer response specified with adequate space provided (a page or more) to include one or more ideas with or without paragraphing • Expectations, directions, and/or time provided for some revision and editing – Item requires understanding of a several features of writing to select correct answer or earn full credit – Item requires consideration of purpose, genre, or audience to select correct answer or earn full credit – One or more distractors may be plausible/attractive; distractors may contribute to complexity

DEPTH & COMPLEXITY

For each writing prompt, consider the complexity of the thinking and processes students use to work through the prompt and then craft a response that earns full credit based on the rubric used to score the item.

For each item consider the complexity of the process students use to think through the specific item stem and the distractors (for SSR items) in order to select the correct answer or earn full credit using the test scoring rubric.

Level 4	• Rubric includes specific attention to purpose, genre, or audience • Rubric includes specific attention to all/most of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level) • Multi-paragraph response specified or clearly implied (more than one idea required) with sufficient space provided • Expectations directions, and/or time for substantive revision and editing -- – Item requires analysis and complex understanding of features of writing to select correct answer or earn full credit – Item requires consideration of purpose, genre, or audience to select correct answer or earn full credit – Multiple distractors may be plausible and/or attractive based on prior knowledge or the text; distractors may contribute to complexity

Writing WITH Sources—Research & Literary Response

Stimuli for short or extended writing prompts or constructed response questions must include text reading. They may also include photos, videos, illustrations, and some limited text. Requirements for responding direct writers to use evidence from these stimuli in their written response.

Descriptions that begin with "prompt" are used to evaluate test items that require students to write brief or extended responses. Descriptions that begin with "item" are used to evaluate test items that are in SSR or multiple-SR formats.

Table D7. Example Importance Rubric: *Writing WITH Sources* Importance

FOCUS & IMPORTANCE
Consider the importance and focus of the prompt or item with regard to authentic high quality, grade-level expectations for writing using sources. How well does the item represent an important focus for instruction and assessment at this grade level?

Level	Description
Level 1	• Prompt does not provide an authentic, grade-level appropriate, writing task related to the source materials (e.g. text, video, etc.) • Prompt does not specify or clearly imply an appropriate purpose, genre, or audience for the writing. --- – Item assesses a relatively unimportant skill, strategy, or quality of writing for the targeted grade level – Item presents skills, strategy, or quality in a way that is decontextualized or tangential to the source material and an authentic writing context – Item does not specify or clearly imply an appropriate purpose, genre, or audience for the writing.
Level 2	• Prompt provides a somewhat authentic, grade-level appropriate, writing task related to <u>some ideas</u> in the text source material • Prompt may specify or clearly implies an appropriate purpose, genre, or audience for the writing but these are not needed to earn full credit. --- – Item assesses a somewhat important skill, strategy, or quality of writing for the targeted grade level – Items presents skill, strategy, or quality in a way that is somewhat related to the source material & an authentic writing context – Item may specify or clearly imply an appropriate purpose, genre, or audience for the writing but these are not needed to respond correctly to the item
Level 3	• Prompt provides an authentic, grade-level appropriate, writing task related to the <u>important ideas/content</u> of the source materials • Prompt specifies or clearly implies appropriate purpose, genre, or audience for the writing. --- – Item assesses an essential skill, strategy, or quality of writing for the targeted grade level – Item presents skills, strategy or quality in a way that is authentic to the source material and applicable to actual writing – Item specifies or clearly implies an appropriate purpose, genre, or audience for the writing that need to be considered to respond correctly to the item

Table D8. Example Complexity Rubric: *Writing WITH Sources* Complexity

DEPTH & COMPLEXITY
For each writing prompt, consider the complexity of the thinking and processes students use to work through the prompt, understand the sources, and then craft a response that earns full credit based on the rubric used to score the item.

Level	Description
Level 1	• Rubric does not include attention to purpose, genre, or audience • Rubric does not include attention to an understanding the content/concepts of the source text(s) • Rubric includes little attention or general, generic attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) • Space provided suggests limited response - approximately 1 paragraph or less • First draft writing with no expectations, direction, and/or time for editing and/or revision -- – Item generally requires rote skill, recall, or explicit knowledge – Item doesn't include attention to purpose, genre, or audience; these are not needed to select correct answer or earn full credit – 0-1 distractors are plausible or attractive based on prior knowledge or the text/s.
Level 2	• Rubric requires some attention to purpose, genre, or audience • Rubric includes general reference to understanding the content/concepts of the source text(s) with little/no specificity other than requirement to use or cite text sources in the writing • Rubric includes specific attention to some of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level) • Space provided suggests short response- greater than a single paragraph • First draft writing with some expectations, direction, and/or time for minor editing (editing may focus on conventions) -- – Item requires understanding of a single or simple feature of the writing to select correct answer or earn full credit – Item includes attention to audience or purpose, genre, or audience but may not be relevant or need to be considered to select correct answer or earn full credit – Item requires mostly literal comprehension of text accompanying it to select correct answer or earn full credit – 1-2 distractors may be plausible/attractive based on prior knowledge or the text/s
Level 3	• Rubric includes specific attention to purpose, genre, or audience • Rubric includes specific references to the content/concepts of the source text(s) but may not distinguish simple vs deep understanding of text(s) communicated in writing • Rubric includes specific attention to most of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level)

DEPTH & COMPLEXITY

For each writing prompt, consider the complexity of the thinking and processes students use to work through the prompt, understand the sources, and then craft a response that earns full credit based on the rubric used to score the item.

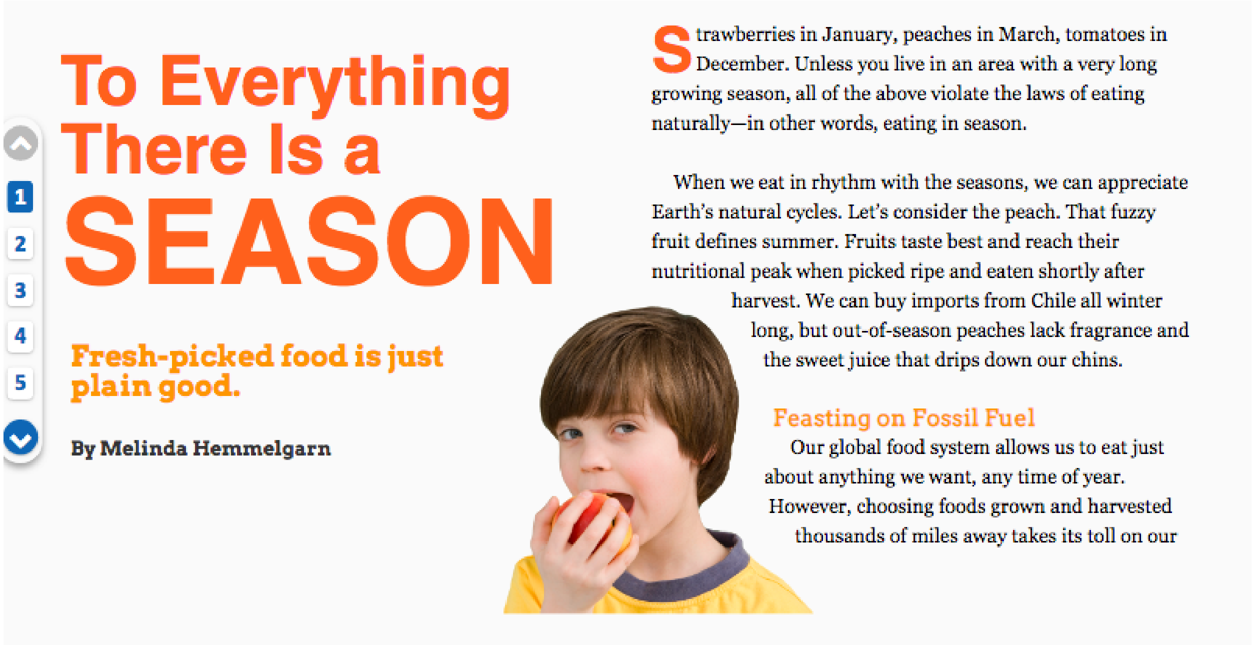
	• Longer response specified with adequate space provided (a page or more) to include one or more ideas with or without paragraphing • Expectations, directions, and/or time provided for some revision and editing – Item requires understanding of a several features of writing to select correct answer or earn full credit – Item requires consideration of purpose, genre, or audience to select correct answer or earn full credit – Item requires some inferential of the text accompanying it to select correct answer or earn full credit – 1 or more distractors may be plausible/attractive based on prior knowledge or the text/s; distractors may contribute to complexity
Level 4	• Rubric includes specific attention to purpose, genre, or audience • Rubric includes specific references to the content/concepts of the source text(s) including expectations for deeper levels of comprehension (e.g. inferring, analyzing, evaluating) to be included in the writing • Rubric includes specific attention to all/most of the following: idea development, attention to idea development, organization/coherence, word choice, sentence structure, conventions (spelling, grammar, usage) (as appropriate to grade level) • Multi-paragraph response specified or clearly implied (more than one idea required) with sufficient space provided • Expectations directions, and/or time for substantive revision and editing – Item requires analysis and complex understanding of features of writing to select correct answer or earn full credit – Item requires consideration of genre, purpose, or audience to select correct answer or earn full credit – Item requires deeper understanding of the text(s) accompanying it to select correct answer or earn full credit – Multiple distractors may be plausible and/or attractive based on prior knowledge or the text; distractors may contribute to complexity

APPENDIX E. EXEMPLAR NAEP READING AND WRITING STIMULI AND ITEMS: WITH PANELISTS' RATINGS

Reading: Grade 4

The following passage stimulus and selected items associated with it are publicly available at NCES's NAEP Question Tool (https://nces.ed.gov/nationsreportcard/nqt/).

Passage



To Everything There Is a SEASON

Fresh-picked food is just plain good.

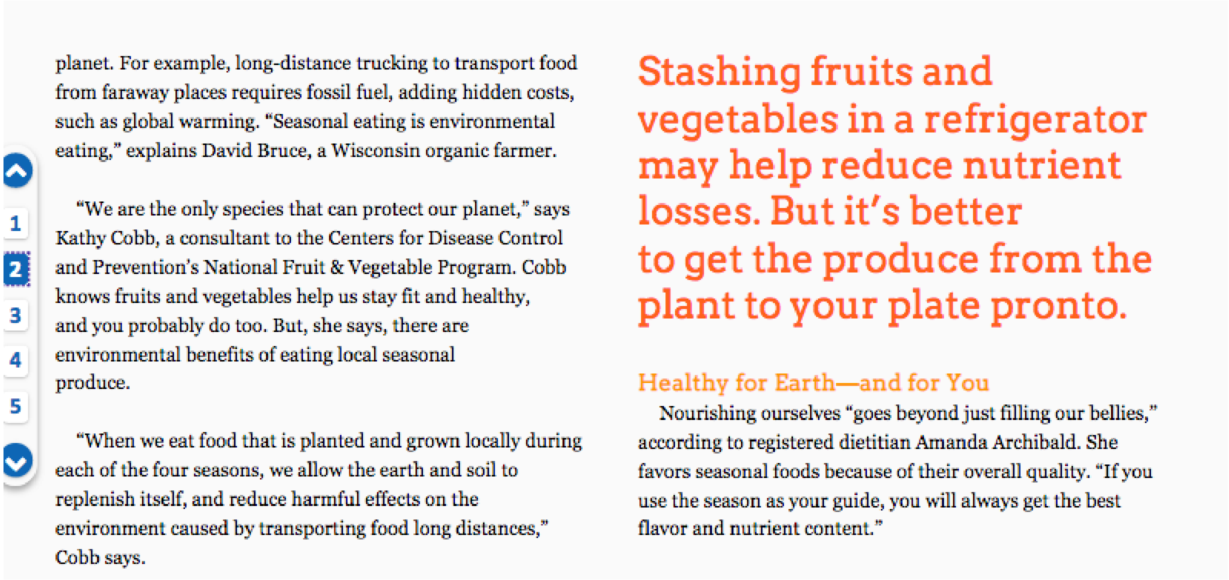
By Melinda Hemmelgarn

Strawberries in January, peaches in March, tomatoes in December. Unless you live in an area with a very long growing season, all of the above violate the laws of eating naturally—in other words, eating in season.

When we eat in rhythm with the seasons, we can appreciate Earth's natural cycles. Let's consider the peach. That fuzzy fruit defines summer. Fruits taste best and reach their nutritional peak when picked ripe and eaten shortly after harvest. We can buy imports from Chile all winter long, but out-of-season peaches lack fragrance and the sweet juice that drips down our chins.

Feasting on Fossil Fuel

Our global food system allows us to eat just about anything we want, any time of year. However, choosing foods grown and harvested thousands of miles away takes its toll on our



planet. For example, long-distance trucking to transport food from faraway places requires fossil fuel, adding hidden costs, such as global warming. "Seasonal eating is environmental eating," explains David Bruce, a Wisconsin organic farmer.

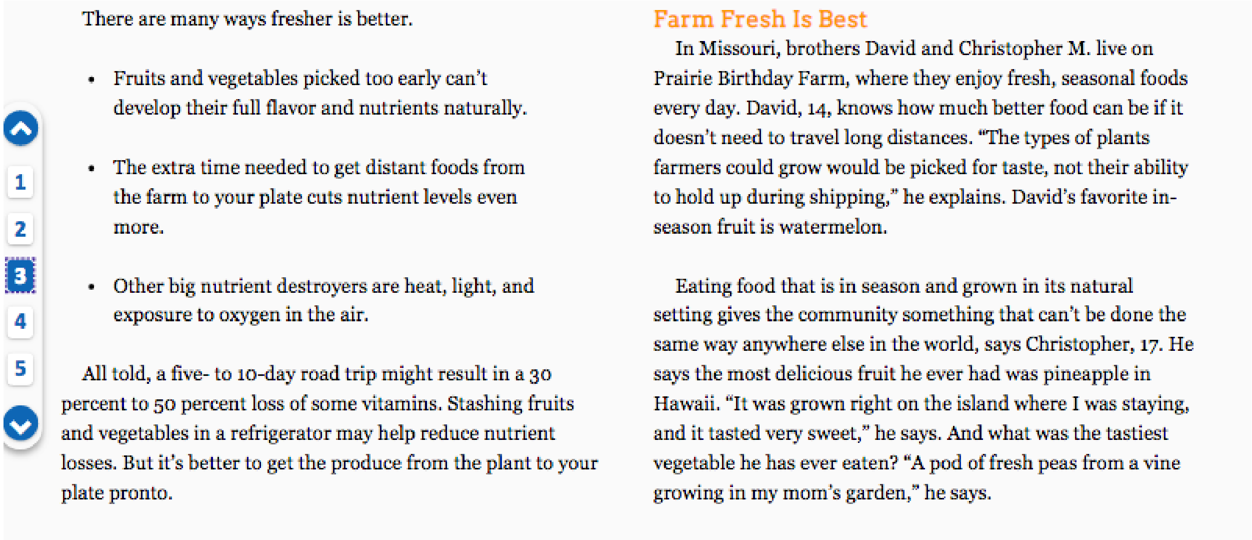
"We are the only species that can protect our planet," says Kathy Cobb, a consultant to the Centers for Disease Control and Prevention's National Fruit & Vegetable Program. Cobb knows fruits and vegetables help us stay fit and healthy, and you probably do too. But, she says, there are environmental benefits of eating local seasonal produce.

"When we eat food that is planted and grown locally during each of the four seasons, we allow the earth and soil to replenish itself, and reduce harmful effects on the environment caused by transporting food long distances," Cobb says.

Stashing fruits and vegetables in a refrigerator may help reduce nutrient losses. But it's better to get the produce from the plant to your plate pronto.

Healthy for Earth—and for You

Nourishing ourselves "goes beyond just filling our bellies," according to registered dietitian Amanda Archibald. She favors seasonal foods because of their overall quality. "If you use the season as your guide, you will always get the best flavor and nutrient content."



There are many ways fresher is better.

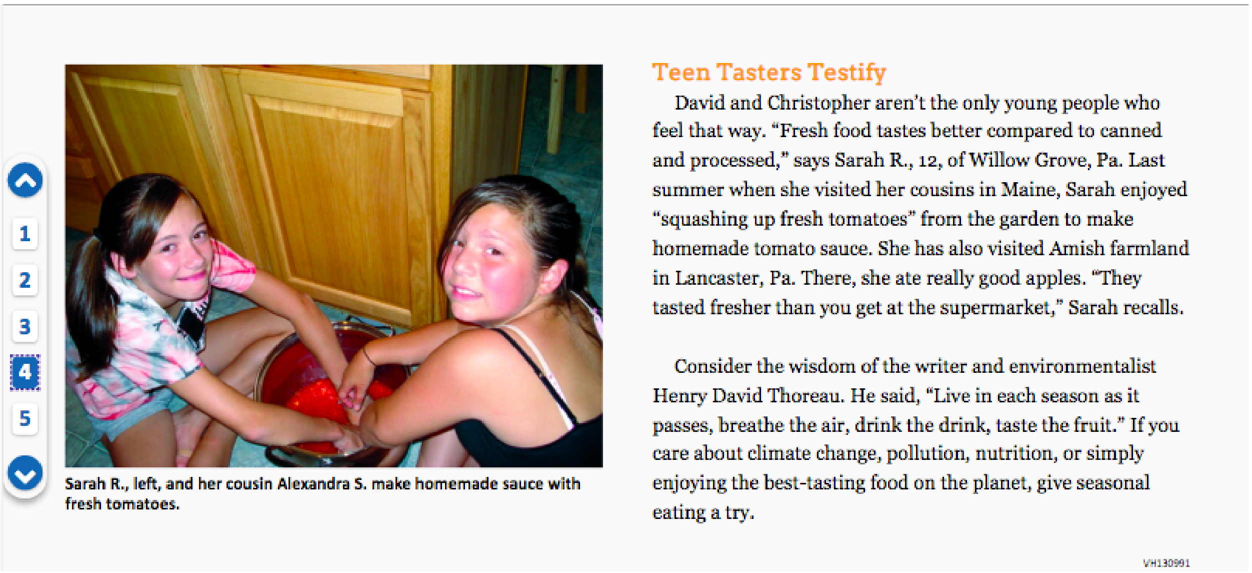
- Fruits and vegetables picked too early can't develop their full flavor and nutrients naturally.
- The extra time needed to get distant foods from the farm to your plate cuts nutrient levels even more.
- Other big nutrient destroyers are heat, light, and exposure to oxygen in the air.

All told, a five- to 10-day road trip might result in a 30 percent to 50 percent loss of some vitamins. Stashing fruits and vegetables in a refrigerator may help reduce nutrient losses. But it's better to get the produce from the plant to your plate pronto.

Farm Fresh Is Best

In Missouri, brothers David and Christopher M. live on Prairie Birthday Farm, where they enjoy fresh, seasonal foods every day. David, 14, knows how much better food can be if it doesn't need to travel long distances. "The types of plants farmers could grow would be picked for taste, not their ability to hold up during shipping," he explains. David's favorite in-season fruit is watermelon.

Eating food that is in season and grown in its natural setting gives the community something that can't be done the same way anywhere else in the world, says Christopher, 17. He says the most delicious fruit he ever had was pineapple in Hawaii. "It was grown right on the island where I was staying, and it tasted very sweet," he says. And what was the tastiest vegetable he has ever eaten? "A pod of fresh peas from a vine growing in my mom's garden," he says.



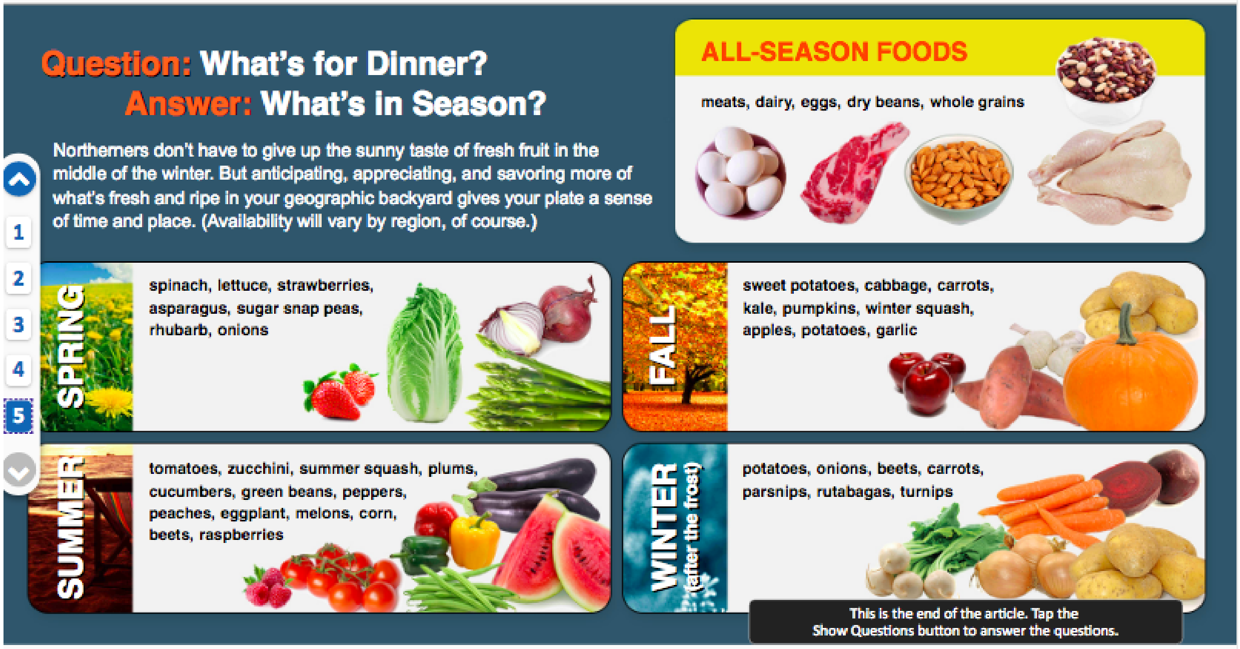
Sarah R., left, and her cousin Alexandra S. make homemade sauce with fresh tomatoes.

Teen Tasters Testify

David and Christopher aren't the only young people who feel that way. "Fresh food tastes better compared to canned and processed," says Sarah R., 12, of Willow Grove, Pa. Last summer when she visited her cousins in Maine, Sarah enjoyed "squashing up fresh tomatoes" from the garden to make homemade tomato sauce. She has also visited Amish farmland in Lancaster, Pa. There, she ate really good apples. "They tasted fresher than you get at the supermarket," Sarah recalls.

Consider the wisdom of the writer and environmentalist Henry David Thoreau. He said, "Live in each season as it passes, breathe the air, drink the drink, taste the fruit." If you care about climate change, pollution, nutrition, or simply enjoying the best-tasting food on the planet, give seasonal eating a try.

VH130991



Question: What's for Dinner?
Answer: What's in Season?

Northerners don't have to give up the sunny taste of fresh fruit in the middle of the winter. But anticipating, appreciating, and savoring more of what's fresh and ripe in your geographic backyard gives your plate a sense of time and place. (Availability will vary by region, of course.)

ALL-SEASON FOODS

meats, dairy, eggs, dry beans, whole grains

SPRING

spinach, lettuce, strawberries, asparagus, sugar snap peas, rhubarb, onions

FALL

sweet potatoes, cabbage, carrots, kale, pumpkins, winter squash, apples, potatoes, garlic

SUMMER

tomatoes, zucchini, summer squash, plums, cucumbers, green beans, peppers, peaches, eggplant, melons, corn, beets, raspberries

WINTER (after the frost)

potatoes, onions, beets, carrots, parsnips, rutabagas, turnips

This is the end of the article. Tap the Show Questions button to answer the questions.

Panelists' ratings of reading selection:

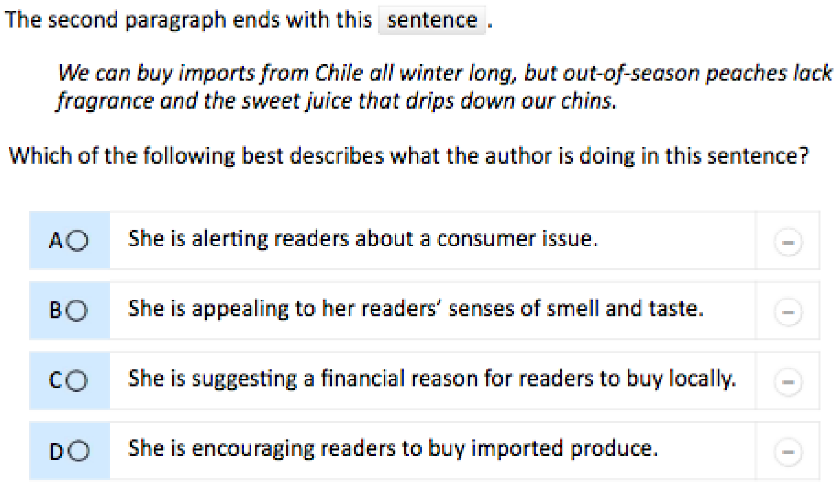
Grade-level appropriateness = 2 (on grade level)

Diversity of perspectives = 2 (some)

Engagingness = 2 (some)

Selected Items

ITEM 2



The second paragraph ends with this sentence.

We can buy imports from Chile all winter long, but out-of-season peaches lack fragrance and the sweet juice that drips down our chins.

Which of the following best describes what the author is doing in this sentence?

A ○ She is alerting readers about a consumer issue.

B ○ She is appealing to her readers' senses of smell and taste.

C ○ She is suggesting a financial reason for readers to buy locally.

D ○ She is encouraging readers to buy imported produce.

Item description:

Subdomain: Craft

Format: Single Selected Response

Cognitive Target: Critique/Evaluate

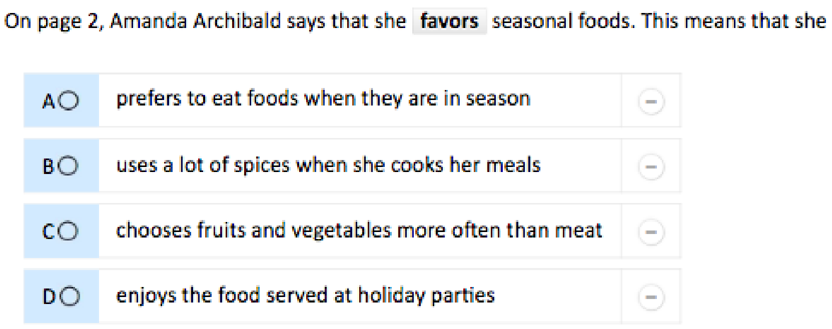
Difficulty: .64

Panelists' ratings of item:

Importance = 2: Item assesses understanding of idea that is of *some* importance to overall comprehension of the text.

Complexity = 2: Identifying the correct response requires test taker to recognize an interpretation of the purpose of a sentence that requires some abstract thinking.

ITEM 4



On page 2, Amanda Archibald says that she **favors** seasonal foods. This means that she

A ○ prefers to eat foods when they are in season

B ○ uses a lot of spices when she cooks her meals

C ○ chooses fruits and vegetables more often than meat

D ○ enjoys the food served at holiday parties

Item description:

Subdomain: Vocabulary

Format: Single Selected Response

Cognitive Target: Integrate/Interpret

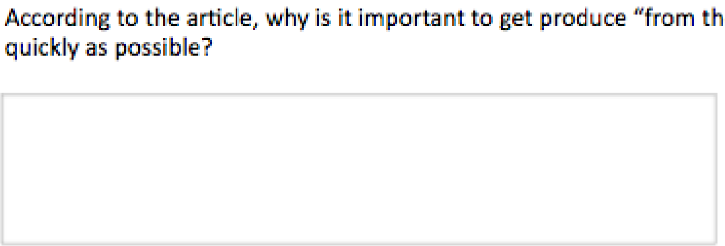
Difficulty: .93

Panelists' ratings of item:

Importance = 1: Consistent with NAEP Reading Framework, the word assessed is of minor importance to the overall understanding of the text.

Complexity = 1: Identifying the meaning of the word as it is used in the text is literal and straightforward.

ITEM 5



According to the article, why is it important to get produce "from the plant to your plate" as quickly as possible?

NAEP scoring rubric for full credit:

	Code	Description
Acceptable	**2**	Responses at this level provide a reason from the article that explains why it is important to get food from plant to plate as quickly as possible. • It's important to get produce "from the plant to your plate" quickly because fruits and vegetables picked too early can't develop their full flavor, and the extra time needed to ship food from the farm cuts nutrients. • When you eat locally grown fruits and vegetables, it makes less pollution going into the atmosphere than it would by shipping the food from a different country. • Because fresh tastes better and is more healthy for you.

Item description:

Subdomain: Key Ideas/Details

Format: Short Constructed Response

Cognitive Target: Locate/Recall

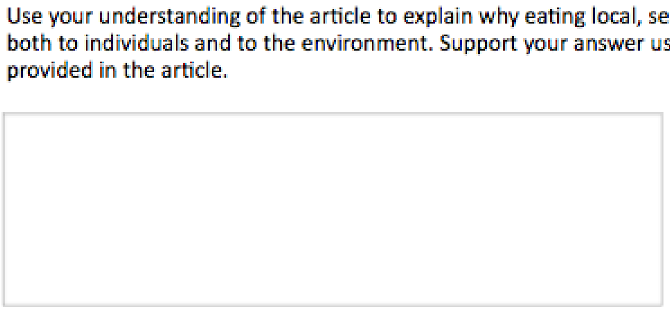
Difficulty: .66

Panelists' ratings of item:

Importance = 3: Item assesses understanding of an idea that is important to building coherent understanding of the text content.

Complexity = 1: Item/rubric requires minimal processing of information that is explicit stated in text.

ITEM 6



Use your understanding of the article to explain why eating local, seasonal food is important both to individuals and to the environment. Support your answer using specific evidence provided in the article.

NAEP scoring rubric for full credit:

	Code	Description
Full Comprehension	3	Responses at this level explain why eating local, seasonal food is important to both individuals and to the environment, and support the answer using specific evidence from the article. This evidence may be in the form of quotations, paraphrase, or accurate summary. • Eating local, seasonal food is important to individuals because it is healthier, tastier, and has much more nutritional value if it isn't shipped all the way here from Chile. Local and seasonal food is also important to the environment because it saves gas and the fumes from large shipping boats and trucks from being let into the atmosphere. • Because if you eat imported fruits and vegetables they won't taste as fresh, because they have set out for a long period of time. And when you import foods from other countries, you use large amounts of fossil fuels. • It's important to individuals because its healthier that way and important to the environment to cut down on pollution and use of fossil fuels.

Item description:

Subdomain: Integrate/Analyze

Format: Short Constructed Response

Cognitive Target: Integrate/Interpret

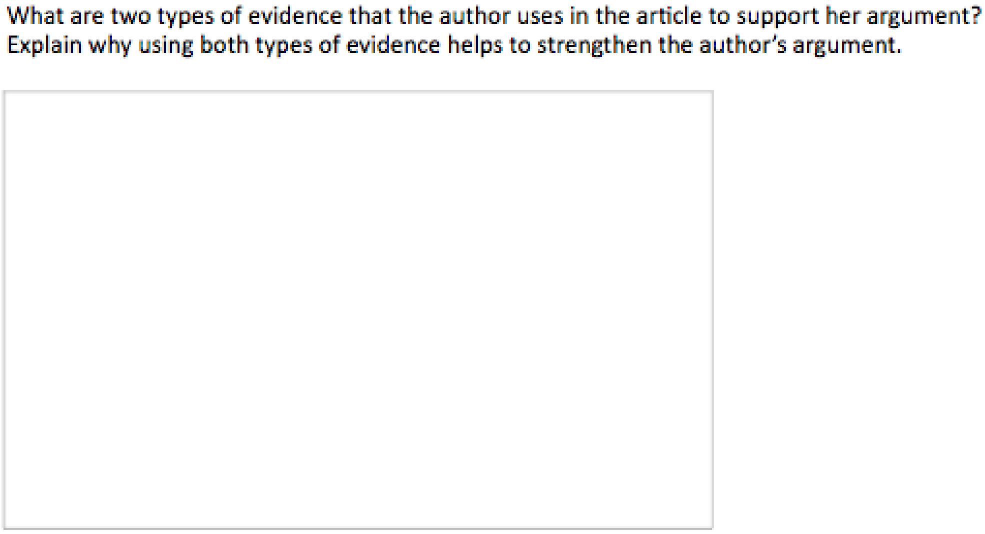
Difficulty: .31 (3 points); .86 (2 points)

Panelists' ratings of item:

Importance 3: Item assesses understanding of an idea that is important to building coherent understanding of the text content.

Complexity 2: Full-credit responses require test takers to integrate fairly concrete/literal information from several places in the text. (Note: Difficulty is likely due to the requirement of multiple [three] pieces of information to receive full credit.

ITEM 7



What are two types of evidence that the author uses in the article to support her argument? Explain why using both types of evidence helps to strengthen the author's argument.

NAEP scoring rubric for full credit:

	Code	Description
Extensive	4	Responses at this level identify two types of evidence that the author uses in the article to support her argument and explain why using both strengthens her argument. • The author used teenagers and facts to defend her argument. Using teen taste testers helps because she has opinions of others, not just her own, as support that locally grown produce tastes better. Facts help support her argument because she is proving that not only do the fruits taste better but they also have more nutrients than canned and processed fruits. • One type of evidence that the author uses is quotes from a registered dietitian. This helps with her argument by giving the sense of a respected practice reassuring her argument. The second is the use of teenage kids. The author used their opinions so the reader can relate to the people in the story.

Item description:

Subdomain: Craft

Format: Extended Constructed Response

Cognitive Target: Critique/Evaluation

Difficulty: .08 (4 points); .42 (3 points)

Panelists' ratings of item:

Importance 3: Item assesses understanding of how text is organized and how organization helps strengthen the author's argument. These understandings are essential to building a coherent and deep understanding of the text.

Complexity 3: Full-credit responses require test takers to engage in analysis and evaluation about fairly abstract ideas and provide multiple (four) pieces of information in their responses.

Writing: Grade 8

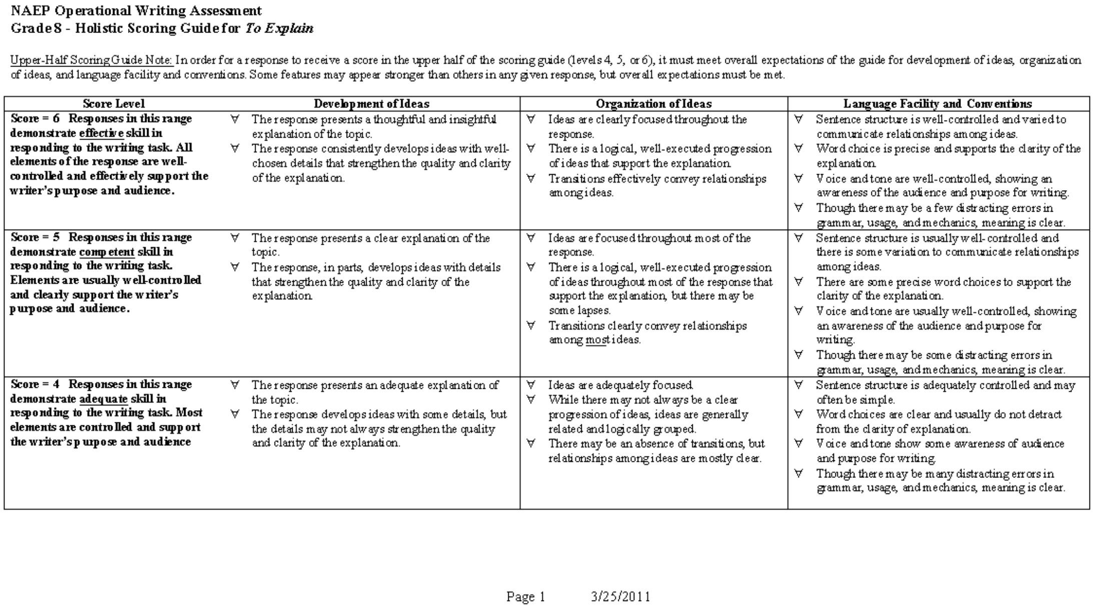
Examples that were used in this study for writing were not released by NAEP and, hence, cannot be included as originally planned in this appendix. Therefore, we have included only background information on two of the prompts scored for this study as well as released samples from 2011 NAEP Writing that were not included in this study. The purpose of providing the 2011 examples is to illustrate for readers what is meant by NAEP's particular approach to writing assessments which is "writing without sources."

In this appendix, we offer information on two different types of NAEP prompts, "to explain" and "to persuade". For type each we provide:

- the scoring rubrics from two of the 2017 prompts that were used in this study
- the panel ratings for those two 2017 prompts
- examples from the two publicly released 2011 prompts as the general structure for the prompt is similar to those used in Writing Grade 8 2011.

PROMPT "To Explain"

2017 "To Explain" – Rubric



NAEP Operational Writing Assessment
Grade 8 - Holistic Scoring Guide for *To Explain*

Upper-Half Scoring Guide Note: In order for a response to receive a score in the upper half of the scoring guide (levels 4, 5, or 6), it must meet overall expectations of the guide for development of ideas, organization of ideas, and language facility and conventions. Some features may appear stronger than others in any given response, but overall expectations must be met.

Score Level	Development of Ideas	Organization of Ideas	Language Facility and Conventions
Score = 6 Responses in this range demonstrate effective skill in responding to the writing task. All elements of the response are well-controlled and effectively support the writer's purpose and audience.	∀ The response presents a thoughtful and insightful explanation of the topic. ∀ The response consistently develops ideas with well-chosen details that strengthen the quality and clarity of the explanation.	∀ Ideas are clearly focused throughout the response. ∀ There is a logical, well-executed progression of ideas that support the explanation. ∀ Transitions effectively convey relationships among ideas.	∀ Sentence structure is well-controlled and varied to communicate relationships among ideas. ∀ Word choice is precise and supports the clarity of the explanation. ∀ Voice and tone are well-controlled, showing an awareness of the audience and purpose for writing. ∀ Though there may be a few distracting errors in grammar, usage, and mechanics, meaning is clear.
Score = 5 Responses in this range demonstrate competent skill in responding to the writing task. Elements are usually well-controlled and clearly support the writer's purpose and audience.	∀ The response presents a clear explanation of the topic. ∀ The response, in parts, develops ideas with details that strengthen the quality and clarity of the explanation.	∀ Ideas are focused throughout most of the response. ∀ There is a logical, well-executed progression of ideas throughout most of the response that support the explanation, but there may be some lapses. ∀ Transitions clearly convey relationships among most ideas.	∀ Sentence structure is usually well-controlled and there is some variation to communicate relationships among ideas. ∀ There are some precise word choices to support the clarity of the explanation. ∀ Voice and tone are usually well-controlled, showing an awareness of the audience and purpose for writing. ∀ Though there may be some distracting errors in grammar, usage, and mechanics, meaning is clear.
Score = 4 Responses in this range demonstrate adequate skill in responding to the writing task. Most elements are controlled and support the writer's purpose and audience	∀ The response presents an adequate explanation of the topic. ∀ The response develops ideas with some details, but the details may not always strengthen the quality and clarity of the explanation.	∀ Ideas are adequately focused. ∀ While there may not always be a clear progression of ideas, ideas are generally related and logically grouped. ∀ There may be an absence of transitions, but relationships among ideas are mostly clear.	∀ Sentence structure is adequately controlled and may often be simple. ∀ Word choices are clear and usually do not detract from the clarity of explanation. ∀ Voice and tone show some awareness of audience and purpose for writing. ∀ Though there may be many distracting errors in grammar, usage, and mechanics, meaning is clear.

Page 1 3/25/2011

2017 "To Explain" – Panelists' ratings of writing prompt

Importance = 3: Prompt provides an authentic, grade-level-appropriate writing task that specifies an appropriate purpose and audience.

Complexity = 4: Based on rubric criteria and instantiations as seen in anchor papers.

Grade-level appropriateness = 2 (on grade level)

Grade-level engagingness = 2 (some)

Diversity of perspectives = 2 (some)

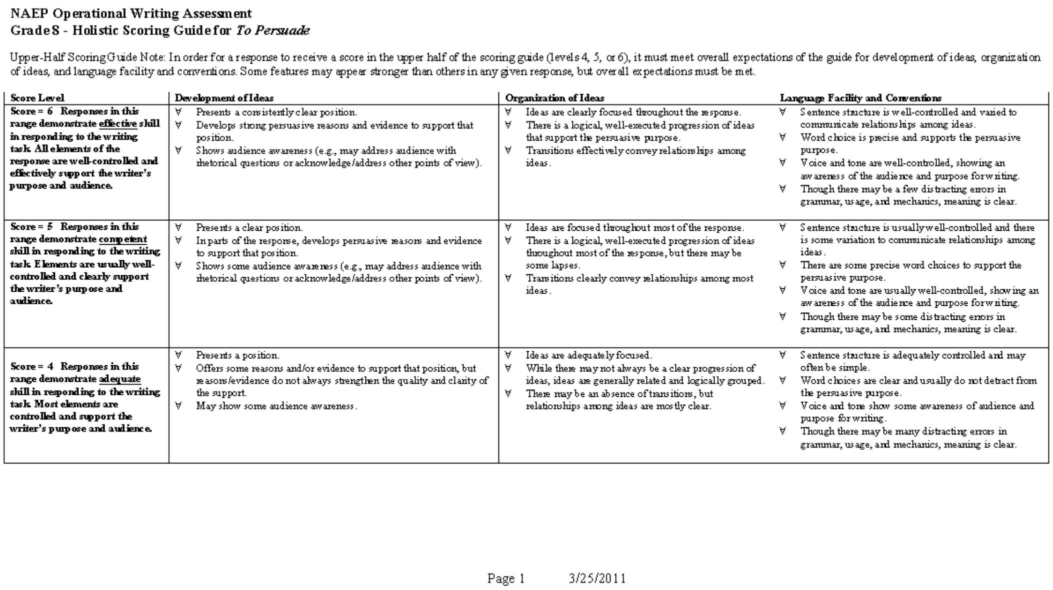
2011 "To Explain" – Writing prompt

"A magazine for young people is asking its readers to submit articles about changes in people's lives.

Write an article to send to the magazine. In your article, write about a time when the way you thought or felt about something changed. For example, maybe you began to like someone you at first disliked, or maybe you lost interest in an activity you used to find interesting. In your article, explain what the change was and why it happened."

PROMPT "To Persuade"

2017 "To Persuade" – Rubric



NAEP Operational Writing Assessment
Grade 8 - Holistic Scoring Guide for *To Persuade*

Upper-Half Scoring Guide Note: In order for a response to receive a score in the upper half of the scoring guide (levels 4, 5, or 6), it must meet overall expectations of the guide for development of ideas, organization of ideas, and language facility and conventions. Some features may appear stronger than others in any given response, but overall expectations must be met.

Score Level	**Development of Ideas**	**Organization of Ideas**	**Language Facility and Conventions**
Score = 6 Responses in this range demonstrate effective skill in responding to the writing task. All elements of the response are well-controlled and effectively support the writer's purpose and audience.	∀ Presents a consistently clear position. ∀ Develops strong persuasive reasons and evidence to support that position. ∀ Shows audience awareness (e.g., may address audience with rhetorical questions or acknowledge/address other points of view).	∀ Ideas are clearly focused throughout the response. ∀ There is a logical, well-executed progression of ideas that support the persuasive purpose. ∀ Transitions effectively convey relationships among ideas.	∀ Sentence structure is well-controlled and varied to communicate relationships among ideas. ∀ Word choice is precise and supports the persuasive purpose. ∀ Voice and tone are well-controlled, showing an awareness of the audience and purpose for writing. ∀ Though there may be a few distracting errors in grammar, usage, and mechanics, meaning is clear.
Score = 5 Responses in this range demonstrate competent skill in responding to the writing task. Elements are usually well-controlled and clearly support the writer's purpose and audience.	∀ Presents a clear position. ∀ In parts of the response, develops persuasive reasons and evidence to support that position. ∀ Shows some audience awareness (e.g., may address audience with rhetorical questions or acknowledge/address other points of view).	∀ Ideas are focused throughout most of the response. ∀ There is a logical, well-executed progression of ideas throughout most of the response, but there may be some lapses. ∀ Transitions clearly convey relationships among most ideas.	∀ Sentence structure is usually well-controlled and there is some variation to communicate relationships among ideas. ∀ There are some precise word choices to support the persuasive purpose. ∀ Voice and tone are usually well-controlled, showing an awareness of the audience and purpose for writing. ∀ Though there may be some distracting errors in grammar, usage, and mechanics, meaning is clear.
Score = 4 Responses in this range demonstrate adequate skill in responding to the writing task. Most elements are controlled and support the writer's purpose and audience.	∀ Presents a position. ∀ Offers some reasons and/or evidence to support that position, but reasons/evidence do not always strengthen the quality and clarity of the support. ∀ May show some audience awareness.	∀ Ideas are adequately focused. ∀ While there may not always be a clear progression of ideas, ideas are generally related and logically grouped. ∀ There may be an absence of transitions, but relationships among ideas are mostly clear.	∀ Sentence structure is adequately controlled and may often be simple. ∀ Word choices are clear and usually do not detract from the persuasive purpose. ∀ Voice and tone show some awareness of audience and purpose for writing. ∀ Though there may be many distracting errors in grammar, usage, and mechanics, meaning is clear.

Page 1 3/25/2011

2017 "To Persuade" – Panelists' ratings of writing prompt

Importance = 3: Prompt provides an authentic, grade-level-appropriate writing task that specifies an appropriate purpose, genre, and audience.

Complexity = 3: Based on rubric criteria and instantiations as seen in anchor papers.

Grade-level appropriateness = 2 (on grade level)

Grade-level engagingness = 2 (some)

Diversity of perspectives = 2 (some)

2011 "To Persuade" – Writing prompt

"Some of your friends perform community service. For example, some tutor elementary school children and others clean up litter. They think helping the community is very important. But other friends of yours thing community service takes too much time away from what they need or want to do. Your principal is deciding whether to require all students to perform community service.

Write a letter to your principal in which you take a position on whether students should be required to perform community service. Support your position with examples."