A FRAMEWORK TO EVALUATE

COGNITIVE COMPLEXITY IN READING ASSESSMENTS

Background

In 2013, the Council of Chief State School Officers (CCSSO), working collaboratively with state education agencies, released a set of criteria for states to use to evaluate and procure high-quality assessments.[1] The English Language Arts (ELA)/Literacy section of the document included nine content-specific criteria to evaluate the alignment of assessments to college- and career-ready ELA/Literacy standards:

B.1: Assessing student reading and writing achievement in both ELA and literacy;

B.2: Focusing on the complexity of texts;

B.3: Requiring students to read closely and use evidence from texts;

B.4: Requiring a range of cognitive demand;

B.5: Assessing writing;

B.6: Emphasizing vocabulary and language skills;

B.7: Assessing research and inquiry;

B.8: Assessing speaking and listening;

B.9: Ensuring high quality items and a variety of item types

In 2016, both the Thomas B. Fordham Institute and the Human Resources Research Organization (HumRRO) used the criteria to evaluate a set of statewide summative assessments: ACT Aspire, the Massachusetts Comprehensive Assessment System (MCAS), PARCC, and Smarter Balanced. Fordham[2] examined grades 5 and 8 assessments, while HumRRO[3] evaluated high school assessments. Reports for each of these studies included recommendations to improve the methodology. *Achieve*, in partnership with Student Achievement Partners, and in consultation with other content and assessment experts, made improvements, and in 2018 used the updated methodology to review the ACT.[4]

This brief describes efforts to address the evaluation of one of the ELA/Literacy Criteria (B.4): Requiring a range of cognitive demand, which has traditionally used Webb's Depth of Knowledge (DOK)

---

[1] Council of Chief State School Officers. (2014). The Criteria for Procuring and Evaluating High-Quality Assessments. https://ccsso.org/sites/default/files/2017-10/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf

[2] Doorey, N., & Polikoff, M. (2016). Evaluating the content and quality of next generation assessments. Thomas B. Fordham Institute. https://edexcellence.net/publications/evaluating-the-content-and-quality-of-next-generation-assessments

[3] Schultz, S., Michaels, H., Dvorak, R. & Wiley, C. (2016). Evaluating the Content and Quality of Next Generation High School Assessments: Final Report. Human Resources Research Organization. https://www.humrro.org/corpsite/sites/default/files/HQAP_HumRRO_High_School_Study_Final%20Report.pdf

[4] Achieve. (2018). Independent Analysis of the Alignment of the ACT to the Common Core State Standards. https://www.achieve.org/achieve-act-review

as its lens to consider individual assessment items. *Achieve* proposes a new English Language Arts-specific approach for reading items to measure cognitive complexity, which draws on the language of Criterion B.4.

Cognitive Complexity in Reading Items

According to CCSSO Criterion B.4, assessments "require all students to demonstrate a range of higher-order, analytical thinking skills in reading and writing based on the depth and complexity of college- and career-ready standards, allowing robust information to be gathered for students with varied levels of achievement as evidenced by the use of a generic taxonomy such as Webb's Depth of Knowledge (DOK), or, preferably, classifications specific to the discipline."

The Fordham and HumRRO studies examined this criterion through the lens of Webb's Depth of Knowledge (DOK), assigning DOK level ratings to items. In the DOK framework, an item that focuses on the recall of facts or definitions would be considered DOK Level 1 (recall) while an item that goes beyond a habitual response and requires students to make some decisions would be at DOK Level 2 (skill/concept). Items that require planning and analysis, beyond that expected in Levels 1 and 2, would be at DOK Level 3 (strategic thinking). Lastly, an item that requires complex reasoning over a period of time would be at DOK Level 4 (extended thinking).[5] The complexity of text which students analyze to answer questions is noticeably absent from the DOK framework.

Introducing the Reading Cognitive Complexity Framework

Reading researchers and assessment experts have acknowledged that text complexity is critical. In the 2005 brief *Reading Between the Lines*, ACT clearly identifies text complexity as the key factor for college and career readiness for reading:

"Performance on complex texts is the clearest differentiator in reading between students who are more likely to be ready for college and those who are less likely to be ready." (p. 6)[6]

Since *Reading Between the Lines* (2005) included the recommendations for states to explicitly define reading expectations and incorporate increasingly complex texts in their state standards (p. 8), text complexity has occupied a prominent place is most states' ELA standards. The Common Core State Standards (CCSS) for ELA prioritize text complexity in the reading standards, and Appendix A of the standards sets clear guidance for determining the complexity of a text. Of the states that have revised the CCSS, fifteen have retained clear and comprehensive guidance for text complexity, and six states' standards discuss text complexity, but vary in their approach to defining it.[7]

Because of this increased awareness of the importance of text complexity in reading, *Achieve* proposes an alternate approach to evaluating the cognitive complexity of reading items. This approach incorporates the significance of the

---

[5] 11 Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. Applied Measurement in Education, 20(1), 7-25.

[6] ACT, Inc. (2005). Reading between the lines: What the ACT reveals about college readiness in reading. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/reading_summary.pdf

[7] Achieve. (2017). Strong standards: A review of changes to state standards since the Common Core. Retrieved from https://www.achieve.org/strong-standards

research by anchoring the framework in text complexity. Evidence and reasoning are then evaluated as additional, but still important, aspects of the reading item. This framework and the accompanying tool (**TABLE 1**) are intended to allow users to determine the range of cognitive complexity present in reading assessments by completing an analysis of the individual reading items present in a test form or event. This analysis is then rolled up at both the passage/item set level and form level (**TABLES 3** and **4**), allowing users to consider the complexity of the items on the entire form.

TABLE 1: A TABLE FOR ANALYZING THE COMPLEXITY OF A READING ASSESSMENT ITEM

Text Complexity	
Evidence	Reasoning

In order to meet the Criteria's call for "all students to demonstrate a range of higher-order, analytical thinking skills," test forms or test events must include items that represent varying levels of complexity. The framework and accompanying tables allow users to hone in on the sources of complexity in individual reading items to ensure that this range is present across the testing event.

Variables for Consideration

Three variables (Text Complexity, Evidence, and Reasoning) are worthy of considerable attention when determining the cognitive complexity of a reading item. To evaluate the cognitive complexity of individual assessment questions, reviewers must answer a key question about each of three variables. Texts are assigned a rating of low, medium, or high complexity; the other two variables receive a score of 1, 2, or 3. The key questions for each variable are outlined below, with additional explanation of what reviewers consider while rating each variable:

1. Text Complexity: To what degree does the complexity of the text fall in the range of grade-level appropriateness?

Text complexity refers to the level of challenge a text provides. Complexity is determined by evaluating two factors: a text's quantitative rating and qualitative features.[8] Combined, those two factors provide a text complexity rating that overlays all items associated with that text.

- The quantitative rating includes readability aspects such as word length, word frequency, sentence length, and text cohesion. Measured using technology, the quantitative rating provides a scale to ensure that students confront increasingly complex texts as they move through the grades.

- The qualitative features include levels of meaning or purpose, structure, language conventionality and clarity, vocabulary, and knowledge demands. Measured by competent readers, qualitative features provide a way to capture the nuance and complexity of ideas, themes, and language of a text—elements that cannot be measured with technology.

---

[8] Lapp, D., Moss, B., Grant, M., & Johnson, K. (2015). Close look at close reading: Teaching students to analyze complex texts, grades K-5. New York: ASCD.

Achieve

For the vast majority of texts, complexity is determined by a thorough quantitative and qualitative analysis; however, some texts' (i.e., poetry and plays) complexity and grade appropriateness can only be determined by a qualitative analysis. When evaluating items for their cognitive complexity, a previously conducted text complexity analysis should be used to evaluate the Text Complexity variable.

Rating Notes:

- The rating for text complexity of an individual text is static for all items associated with it.
- When evaluating items, their associated text ratings are rated as High (H) Medium (M), or Low (L) complexity.
- Some items may direct students to consider more than one text. The text complexity score for these items should reflect the complexity score of the most complex text in the set or pair.
- Items associated with texts that are inappropriately complex for the grade (either too complex, or not complex enough) based on quantitative and qualitative analysis should receive a text complexity rating of Off Grade (OG). Items associated with texts that are not appropriate for the grade level should still be evaluated for Evidence and Reasoning.

2. Evidence: To what degree does the range of evidence impact the complexity of the item?

Evidence refers to the range of evidence the item references. Generally, the broader the range of evidence (e.g., the more evidence the item requires to arrive at the correct answer), the more cognitively complex the item is. Items that focus on a limited range of evidence tend to be less cognitively demanding.

Rating Notes:

When evaluating items, evidence is rated on a scale of Low (1), Medium (2), or High (3) complexity. Note that numerical scores are used for rating both Evidence and Reasoning. This signals the shift from focusing only on the text upon which the item is based to focusing on the complexity of the content of the item itself.

3. Reasoning: To what degree is reasoning (e.g., inferencing, analyzing, synthesizing) needed to respond to the item?

Reasoning refers to how critically the item requires a respondent to deliberate from multiple dimensions to respond accurately to the item. Reasoning:

- Refers to the cognitive processes necessary for comprehension (e.g., inferencing, summarizing, analyzing, evaluating).
- Refers to how much the item requires respondents to grasp nuances and integrate information, including inferencing;[9] items that are literal or explicit are not as cognitively complex as items that are more nuanced or require more integration of information.

---

[9] It is important that the inferences required by the item remain within the four corners of the text; items that require inferences based on prior, but not common, knowledge should receive an overall rating as 0, and a reasoning score of 0. Reviewers rate the item for its level of text complexity and range of evidence required.

• Considers both distractors and the correct answer[10] in an assessment item where distractors exist. Generally, the more plausible the distractors, the more carefully a student must analyze them to differentiate between the accurate response and responses which reflect inaccuracies of the text. In contrast, clearly implausible distractors require less analysis for a student to omit them and bear a lower cognitive load.

Rating Notes:

When evaluating items, reasoning is rated on a scale of Low (1), Medium (2), or High (3) complexity. Note that numerical scores are used for rating both evidence and reasoning. This signals the shift from focusing only on the text upon which the item is based, to focusing on the complexity of the content of the item itself.

A TOOL TO EVALUATE COGNITIVE COMPLEXITY OF INDIVIDUAL READING ITEMS

Text Title:
Text Complexity: To what degree does the complexity of the text fall in the range of grade-level appropriateness? • When compared to other texts in the grade, would this text be placed at the higher end of the grade, right in the middle (average) for the grade, or on the lower end of complexity for the grade? • To what degree does this text require the reader to pay close and careful attention to understand the key ideas and important particulars? *Reminder: Text complexity is the factor that most impacts each student's interaction with individual assessment questions* (For explanations of low/medium/high ratings, see **TABLES 5** and **6**.)
(Low, Medium, High, Off Grade)

Evidence: To what degree does the range of evidence impact the complexity of the item? • To what degree does the item require a student to consider the entirety of the text presented? • If a single text is required: To what degree does the item require evidence from multiple points in a text? • If multiple texts are required: To what degree does the item require evidence from multiple points in each text? (For explanations of 1/2/3 ratings, see **TABLE 7**.)	**Reasoning:** To what degree is reasoning (e.g., inferencing, analyzing, synthesizing) needed to respond to the item? • To what extent does the student need to integrate information across the text to arrive at an accurate response? • To what degree does the item require an understanding of unstated but logical conclusions based on the evidence in the text? • To what degree does the item require students to grasp nuances in the text? • If distractors exist: To what degree does the plausibility of the distractors require students to discriminate among potential responses? (For explanations of 1/2/3 ratings, see **TABLE 8**.)
(1, 2, 3)	(1, 2, 3)

---

[10] Distractors are considered only in items where distractors exist.

TABLE 2:
SAMPLE INDIVIDUAL ITEM RATING TABLE

Text Complexity	
Evidence	Reasoning

TABLE 3:
SAMPLE PASSAGE/ITEM SET ANALYSIS TABLE[11]

Text Complexity	
Evidence	Reasoning

TABLE 4: SAMPLE TEST FORM/TEST EVENT ANALYSIS TABLE[12]

Text Complexity		Text Complexity		Text Complexity		Text Complexity	
Evidence	Reasoning	Evidence	Reasoning	Evidence	Reasoning	Evidence	Reasoning

The tables that follow are intended to provide additional detail for reviewers as they consider the definitions of low, medium, and high complexity for each individual factor. The Text Complexity tables, or some tool similar, should be completed by the assessment vendor prior to reviewing an individual item for its cognitive complexity.

---

[11] For use when ensuring that there is a range of evidence and reasoning within a passage/item set.
[12] For use when evaluating the complexity of multiple passage/item sets on an individual test form or during a single test event. It allows for a quick comparison of items across passages to ensure that items are complex for a variety of reasons.

TABLE 5: TEXT COMPLEXITY, LITERARY TEXT[13]

Text Complexity (Literary Text): To what degree does the complexity of the text fall in the range of grade-level appropriateness?

- When compared to other texts appropriate to the grade, would this text be placed at the higher end of the grade, right in the middle (average) for the grade, or on the lower end of complexity for the grade?
- To what degree does this text require the reader to pay close and careful attention to understand the key ideas and important particulars?

Low	Medium	High
Quantitatively The text is appropriately placed in the grade band as determined by a quantitative analysis and may be on the lower end of the range. **Qualitatively** The text may contain a combination of these elements: • **Organization** is clear, chronological, or predictable • **Graphics/illustrations** directly support and assist in interpreting the text, or are not necessary to understand the text; graphics are not sophisticated or complex • **Language** is explicit, literal, straightforward, and easy to understand • **Vocabulary** is contemporary, familiar, conversational • **Sentences** are mainly simple • **One level of meaning**; obvious theme • Explores a single theme; experiences are common to readers • No references or allusions to other texts or cultural elements	**Quantitatively** The text is appropriately placed in the grade band as determined by a quantitative analysis. **Qualitatively** The text may contain a combination of these elements: • **Organization** may have two or more storylines or multiple subplots; there may be few times shifts; organization may not be predictable • **Graphics** support or extend the meaning of the text • **Language** is fairly complex; may contain some abstract, ironic and/or figurative language; • **Vocabulary** is mostly contemporary, but may be fairly complex and contain vocabulary that is sometimes unfamiliar • **Sentences** may be simple, compound, or complex with several phrases or clauses and transition words; may vary in length • **Multiple levels of meaning** that may be difficult to identify; theme may be clear or implied • Explores several themes with varying levels of abstraction; experiences may or may not be common to most readers • May include some references or allusions to other texts or cultural elements	**Quantitatively** The text is appropriately placed in the grade band as determined by a quantitative analysis and may be on the higher end of the range. **Qualitatively** The text may contain a combination of these elements: • **Organization** is intricate with regard to point of view, time shifts, multiple characters, storylines, and detail • **Graphics/illustrations** are essential to understanding the text; Graphics are highly sophisticated or complex • **Language** is dense/complex; may contain abstract, ironic, or figurative language • **Vocabulary** is complex, unfamiliar, archaic, subject-specific, or overly academic • **Sentences** are mainly complex with several subordinate clauses/phrases; sentences often contain multiple concepts • **Multiple competing levels of meaning** that are difficult to identify; theme is implicit/subtle, ambiguous • Text explores complex, sophisticated, or abstract themes or ideas; references or allusions to other texts or cultural elements • Text and related materials (e.g., tables, charts, pictures, illustrations) require integration and synthesis across modalities

[13] Process used by vendors to determine a text's complexity: Vendors should first use a quantitative score to place the text in an appropriate grade band. Vendors then complete a qualitative analysis, determining which elements of the text (e.g., its structure, levels of meaning, purpose, clarity of language, etc.) provide the most complexity for readers. Texts may have elements of complexity that cross levels (e.g., a clear and simple structure coupled with complex language and syntax; multiple themes coupled with clear and familiar language). Vendors make a determination as to which elements of the complexity most impact the text to give a text a final rating for the grade-level (e.g., a low-complexity grade 6 text, a moderately complex grade 3 text).

TABLE 6: TEXT COMPLEXITY , INFORMATIONAL TEXT

Text Complexity (Informational Text) To what degree does the complexity of the text fall in the range of grade-level appropriateness?
- When compared to other texts in the grade, would this text be placed at the higher end of the grade, right in the middle (average) for the grade, or on the lower end of complexity for the grade?
- To what degree does this text require the reader to pay close and careful attention to understand the key ideas and important particulars?

Low	Medium	High
Quantitatively The text is appropriately placed in the grade band as determined by a quantitative analysis and may be on the lower end of the range. **Qualitatively** The text may contain a combination of these elements: • **Organization** clearly and explicitly connects ideas, processes, or events; chronological, sequential, or predictable • **Text features** are not essential to understanding content • **Graphics/illustrations** are simple and are not necessary to understand the text; graphics are not sophisticated or complex • **Language** is explicit, literal, straightforward, and easy to understand • **Vocabulary** is contemporary, familiar, conversational • **Sentences** are mainly simple • **Purpose** is explicitly stated, clear, concrete, and narrowly focused • No references or allusions to other texts or cultural elements	**Quantitatively** The text is appropriately placed in the grade band as determined by a quantitative analysis. **Qualitatively** The text may contain a combination of these elements: • **Organization** connects a broad range of implied or subtle ideas, processes, or events; generally sequential/chronological, but may contain multiple pathways; some discipline-specific traits • **Text features** enhance the reader's understanding of content • **Graphics** support the text, but are not essential to understanding the text • **Language** is somewhat complex; may contain some abstract, ironic, or figurative language • **Vocabulary** is mostly contemporary, but may be fairly complex that is sometimes unfamiliar • **Sentences** may be simple, compound, or complex with several phrases or clauses and transition words • **Purpose** is implied, but fairly easy to understand; more theoretical or abstract • Common practical knowledge or some discipline specific knowledge; mixes simple and more complicated, abstract ideas • May include some references or allusions to other texts, ideas, or theories	**Quantitatively** The text is appropriately placed in the grade band as determined by a quantitative analysis and may be on the higher end of the range. **Qualitatively** The text may contain a combination of these elements: • **Organization** connects an extensive range of deep, intricate, or ambiguous ideas, processes, or events; may be discipline-specific • **Text features** are essential to understanding content • **Graphics/illustrations** are intricate/extensive; essential to understanding the text; may provide additional, necessary information; graphics are highly sophisticated or complex • **Language** is dense/complex; may contain abstract, ironic, and/or figurative language • **Vocabulary** is complex, unfamiliar, archaic, subject-specific, or overly academic • **Sentences** are mainly complex with several subordinate clauses/phrases; sentences often contain multiple concepts • **Purpose** is subtle and intricate, difficult to understand; includes many theoretical or abstract elements • Extensive levels of discipline-specific or theoretical knowledge; range of challenging abstract concepts • References or allusions to other texts or outside ideas or theories • Text explores complex, sophisticated, or abstract ideas

TABLE 7: RANGE OF EVIDENCE

Range of Evidence To what degree does the range of evidence impact the complexity of the item?

- To what degree does the item require a student consider the entirety of the text presented?
- If a single text is required: To what degree does the item require evidence from multiple points in a text?
- If multiple texts are required: To what degree does the item require evidence from multiple points in each text?

1	2	3
Responding accurately to the item requires evidence that • Is limited to a single section of the text	Responding accurately to the item requires evidence that • Is found in more than one non-contiguous section or several contiguous sections of the text	Responding accurately to the item requires evidence that • Spans the entire text or multiple points in more than one text

TABLE 8: REASONING

Reasoning: To what degree is reasoning (e.g., inferencing, analyzing, synthesizing) needed to respond to the item?

- To what extent does the student need to integrate information across the text to arrive at an accurate response?
- To what degree does the item require an understanding of unstated but logical conclusions based on the evidence in the text?
- To what degree does the item require students to grasp nuances in the text?
- If distractors exist: To what degree do students have to discriminate among plausible distractors?

1	2	3
Comprehending the text and responding accurately to the item requires *reasoning* that • Requires students to locate or recall information and/or • Is largely literal and/or • Does not ask students to integrate information from the text and/or • Has more than one distractor that is either not reasonable or implausible	Comprehending the text and responding accurately to the item requires *reasoning* that • Requires students to integrate and interpret information from the text and/or • Is primarily inferential and/or • Has some distractors that are reasonable but not equally plausible	Comprehending the text and responding accurately to the item requires *reasoning* that • Requires students to critique and evaluate the text and/or • Is deeply inferential and requires a sophisticated understanding of the text and/or • Has most or all highly plausible distractors

Examples

The following example items, associated with the excerpt from *The Great Fire* by Jim Murphy (Appendix A), illustrate how the tool can be used to analyze the cognitive complexity of individual reading items. Correct answers are marked with an asterisk (*).

EXAMPLE 1:

Reread this sentence from paragraph 5 of the passage:

The answer was to make the roads and sidewalks out of wood and elevate them above the waterline, in some places by several feet.

Which two phrases in the sentence best help the reader determine the meaning of the word "elevate"?

A. "The answer was"

B. "to make the roads and sidewalks"

C. "above the waterline"*

D. "in some places"

E. "by several feet"*

Text Complexity	
L	
Evidence	**Reasoning**
1	2

The **Text Complexity** score is low. The quantitative scores of 9.0 (Flesh-Kincaid) and 1000 (Lexile) place the text in the grades 6-8 band. A qualitative analysis indicates that the structure is clear and relationships between ideas are logical. Vocabulary is readily accessible, and the knowledge demands placed on students are low. The purpose is singular and clearly developed for students, so the text is most appropriate for sixth grade students. For grade 6, the text falls at the early end of the grade band.

The **Evidence** score is 1. The evidence necessary to respond accurately to the item is confined within one section of the text, specifically the single sentence from paragraph 5.

The **Reasoning** score is 2. Responding accurately to the item is requires inferencing, interpretation of the text, and consideration of distractors that are reasonable, but not equally plausible. Answer choice A, "The answer was," is less plausible than some of the others.

EXAMPLE 2:

How do the details in paragraphs 3 and 4 about the poor, middle-class and wealthier neighborhoods contribute to the development of the central ideas of the passage?

A. The paragraphs support the idea that the fire spread widely because of the amount of wood in all three areas.*

B. The paragraphs support the idea that wood was the most readily available resource because of the number of trees in that area.

C. The paragraphs support the idea that long ago people were unaware of the importance of well-constrcuted buildings.

D. The paragraphs support the idea that Chicago was different from other large cities during that time.

Text Complexity	
L	
Evidence	Reasoning
3	2

The **Text Complexity** score is low. The quantitative scores of 9.0 (Flesh-Kincaid) and 1000 (Lexile) place the text in the grades 6-8 band. A qualitative analysis indicates that the structure is clear and relationships between ideas are logical. Vocabulary is readily accessible, and the knowledge demands placed on students are low. The purpose is singular and clearly developed for students, so the text is most appropriate for sixth grade students. For grade 6, the text falls at the early end of the grade band.

The **Evidence** score is 3. Several pieces of evidence necessary to respond accurately to the item are spread across several contiguous sections of the text (paragraphs 3 and 4), and students must relate those pieces of evidence to the text as a whole.

The **Reasoning** score is 2. Responding accurately to the item is primarily inferential and requires synthesizing information and sorting among equally plausible distractors.

EXAMPLE 3:

This question has two parts. Answer Part A and then answer Part B.

Part A: Which statement best summarizes the central idea of this passage?

A. The Great Fire of Chicago was one of the most damaging fires in American history.

B. The Great Fire of Chicago quickly got out of control in some neighborhoods but not others.

C. Chicago firefighters could not put out the fire even though many people tried to help.

D. Chicago provided almost perfect conditions for a widespread and damaging fire.*

Part B: Which sentence from the passage provides the best support for the corect answer in Part A?

A. "Neighbors rushed from their homes, many carrying buckets or pots of water."

B. "Chicago in 1871 was a city ready to burn."*

C. "The situation was worst in the middle-class and poorer districts."

D. "Fires were common in all cities back then, and Chicago was no exception."

Text Complexity	
L	
Evidence	Reasoning
3	3

The **Text Complexity** score is low. The quantitative scores of 9.0 (Flesh-Kincaid) and 1000 (Lexile) place the text in the grades 6-8 band. A qualitative analysis indicates that the structure is clear and relationships between ideas are logical. Vocabulary is readily accessible, and the knowledge demands placed on students are low. The purpose is singular and clearly developed for students, so the text is most appropriate for sixth grade students. For grade 6, the text falls at the early end of the grade band.

The **Evidence** score is 3. Students must read and use the entire text to correctly respond to the item.

The **Reasoning** score is 3 because responding accurately to the item is both inferential and evaluative. Part A requires interpretation and synthesis of the text, while Part B extends student thinking by requiring a critique of the best support for Part A. Finally, both Part A and Part B have highly plausible distractors and require an evaluation of each distractor to determine the best answer choice.

SAMPLE PASSAGE/ITEM SET ANALYSIS TABLE FOR *THE GREAT FIRE*

Text Complexity
L

Item Number	Reasoning	Evidence
Example 1	1	2
Example 2	3	2
Example 3	3	3

Notes: This low complexity passage presents items that primarily require a high range of evidence and medium to high levels of reasoning. A range of complexity is present.

The following example items, associated with the speech "Is it a Crime for a Citizen of the United States to Vote?" by Susan B. Anthony (Appendix B), illustrate how the tool can be used to analyze the cognitive complexity of individual reading items at the high school level. Correct answers are marked with an asterisk (*).

EXAMPLE 4:

Anthony states that preventing women from having equal rights with men is a "direct violation of the spirit and the letter of the declarations of the framers of this government." In which sentence in her speech does Anthony most clearly disregard the "letter" (the literal meaning) of the Declaration and substitute a meaning that reflects the "spirit" of the document?

A. "They agree to abandon the methods of brute force in the adjustment of their differences, and adopt those of civilization."

B. "The Declaration of Independence, the United States Constitution, the constitutions of several states and the organic laws of the territories, all alike propose to protect the peple in the exercise of their God-given rights."

C. "Here is pronounced the right of all men, and 'consquently,' as the Quaker preacher said, 'of all women,' to a voice in the government."

D. "For however destrucive in their happiness this government might ecome, a disenfranchised class could neither alter nor abolish it, nor institute a new one, except by the old brute force method of insurrection and rebellion."

Text Complexity	
H	
Evidence	Reasoning
2	3

The **Text Complexity** score is high. The quantitative scores of 12.9 (Flesh Kinkaid) and 1420 (Lexile) place this text in the Grade 11-CCR band. A qualitative analysis indicates that the text uses a clear and logical organizational structure, with language that is often abstract and archaic. The knowledge demands are high; students will benefit from some knowledge of the Women's Suffrage movement and the Declaration of Independence. Anthony explicitly states her purpose. The text is most appropriate for 11th grade students, and it is highly complex for grade 11.

The **Evidence** score is 2. Students must understand a specific quotation and apply that understanding to four other lines from the text.

The **Reasoning** score is 3. Students must carefully analyze Anthony's language in both the quoted section and the text. They must synthesize her meaning and apply her understanding to four equally plausible answer choices.

Why does Anthony provide a list of the wrongs government commits against women?

A. to illustrate that change is needed to fulfill the intentions of the founding documents*

B. to urge women to take their rightful place as rulers instead of dependents

C. to ask women to forgive past actions in hopes of forging a stronger nation

D. to justify the fact that when women have power they will replace the existing government

Text Complexity	
H	
Evidence	**Reasoning**
1	1

The **Text Complexity** score is high. The quantitative scores of 12.9 (Flesh Kinkaid) and 1420 (Lexile) place this text in the Grade 11-CCR band. A qualitative analysis indicates that the text uses a clear and logical organizational structure, with language that is often abstract and archaic. The knowledge demands are high; students will benefit from some knowledge of the Women's Suffrage movement and the Declaration of Independence. Anthony explicitly states her purpose. The text is most appropriate for 11th grade students, and it is highly complex for grade 11.

The **Evidence** score is 1. The evidence to correctly answer this question is contained within one contiguous section of this text.

The **Reasoning** score is 1. Students must analyze the author's language and purpose to correctly answer this question. Answer option B is less plausible, as the text does not mention women becoming "rulers." Anthony focuses entirely on the right to vote.

EXAMPLE 6:

The following question has two parts. Answer Part A and then answer Part B.

Part A: Which statement best describes Anthony's primary point of view regarding the role of government?

A. Government should define and provide basic human rights to all.
B. Government should enforce the law even when citizens are dissatisfied.
C. Government should classify different roles for different groups of people.
D. Government should ensure that basic human rights are not infringed upon.

Part B: Which two statements from the speech provide the best evidence for Anthony's point of view?

A. "We assert the province of government to be to secure the people in the enjoyment of their unalienable rights."*
B. "And when 100 or 100,000,000 peple enter into a free government, they do not barter away their natural rights; they simply pledge themselves to protect each other in the enjoyment of them, through prescribed judicial and legislative tribunals."*
C. "That whenever any form of government becomes destructive of these ends, it is the right of the people to alter or abolish it, and to institute a new government...."
D. "For however desctructive in their happiness this government might become, a disenfranchised class could neither alter nor abolish it, nor institute a new one, except by the old brute force method of insurrection and rebellion."
E. "By them, too, men, as such, were deprived of their divine right to rule, and placed on a political level with women."
F. "By the practice of those declarations all class and caste distinction will be abolished; and slave, serf, plebeian, wife, woman, all alike, bound from their subject position to the proud platform of equality."

Text Complexity	
H	
Evidence	Reasoning
3	2

The **Text Complexity** score is high. The quantitative scores of 12.9 (Flesh Kinkaid) and 1420 (Lexile) place this text in the Grade 11-CCR band. A qualitative analysis indicates that the text uses a clear and logical organizational structure, with language that is often abstract and archaic. The knowledge demands are high, students will benefit from some knowledge of the Women's Suffrage movement and the Declaration of Independence. Anthony explicitly states her purpose. The text is most appropriate for 11th grade students, and it is highly complex for grade 11.

The **Evidence** score is 3. Students must use information from the entire text to understand how Anthony's point of view develops.

The **Reasoning** score is 2. Students must analyze the author's language to understand the purpose and make connections between the language and the perspective.

SAMPLE PASSAGE/ITEM SET ANALYSIS TABLE FOR *IS IT A CRIME FOR A CITIZEN OF THE UNITED STATES TO VOTE?*

Text Complexity
H

Item Number	Reasoning	Evidence
Example 4	2	2
Example 5	1	1
Example 6	3	2

Notes: This high complexity passage presents items at each level of complexity for both evidence and reasoning. A range of complexity is present within this text/item set.

Summary

The tool in this document provides a new way to analyze reading assessment items in summative assessments. This analysis is in two steps:

> **(1) determining the complexity of the text using previously developed text complexity analyses,**

> **(2) analyzing what the item requires a student to do with the text.**

Ideally, a summative assessment will include texts of varying complexity for the grade and tasks that reflect varying degrees of complexity. Interim assessments should reflect the range of complexity appropriate to the purpose of the assessment, time of administration, etc.

We believe this new approach to evaluating item-level complexity has advantages over traditional approaches like DOK. This model provides specific feedback on item complexity and does this through the evaluation of text, range of evidence, and level of reasoning required. Additionally, this approach creates a more seamless system for assessment review by directly connecting to the expectations outlined in Criterion B.4.

Thank You

Achieve, in partnership with Student Achievement Partners, EdReports, and a host of literacy, content, research, and assessment experts, and building on the work of CCSSO and Webb's DOK, has developed this draft framework and tools to provide a new way to think about determining cognitive complexity of a reading item. Critical and informative feedback was provided to *Achieve* from members of the CCSSO ELA SCASS, WestEd, Odell Education, and other reading researchers.

APPENDIX A: Excerpt from *The Great Fire* by Jim Murphy and Text Complexity Analysis

> The Great Fire of Chicago is considered the largest disaster of the 1800s. It is rumored to have started in the barn of Patrick and Catherine O'Leary.

1. A shed attached to the barn was already engulfed by flames. It contained two tons of coal for the winter and a large supply of kindling wood. Fire ran along the dry grass and leaves, and took hold of a neighbor's fence. The heat from the burning barn, shed, and fence was so hot that the O'Learys' house, forty feet away, began to smolder. Neighbors rushed from their homes, many carrying buckets or pots of water. The sound of music and merrymaking stopped abruptly, replaced by the shout of "FIRE!" It would be a warning cry heard thousands of times during the next thirty-one hours.

2. Chicago in 1871 was a city ready to burn. The city boasted having 59,500 buildings, many of them—such as the Courthouse and the Tribune Building—large and ornately decorated. The trouble was that about two-thirds of all these structures were made entirely of wood. Many of the remaining buildings (even the ones proclaimed to be "fireproof") looked solid, but were actually jerrybuilt[1] affairs; the stone or brick exteriors hid wooden frames and floors, all topped with highly flammable tar or shingle roofs. It was also a common practice to disguise wood as another kind of building material. The fancy exterior decorations on just about every building were carved from wood, then painted to look like stone or marble. Most churches had steeples that appeared to be solid from the street, but a closer inspection would reveal a wooden framework covered with cleverly painted copper or tin.

3. The situation was worst in the middle-class and poorer districts. Lot sizes were small, and owners usually filled them up with cottages, barns, sheds, and outhouses—all made of fast- burning wood, naturally. Because both Patrick and Catherine O'Leary worked, they were able to put a large addition on their cottage despite a lot size of just 25 by 100 feet. Interspersed in these residential areas were a variety of businesses—paint factories, lumberyards, distilleries, gasworks, mills, furniture manufacturers, warehouses, and coal distributors.

4. Wealthier districts were by no means free of fire hazards. Stately stone and brick homes had wood interiors and stood side by side with smaller wood-frame houses. Wooden stables and other storage buildings were common, and trees lined the streets and filled the yards.

---

[1] Built cheaply or poorly constructed

5. The links between richer and poorer sections went beyond the materials used for construction or the way buildings were crammed together. Chicago had been built largely on soggy marshland that flooded every time it rained. As the years passed and the town developed, a quick solution to the water and mud problem was needed. The answer was to make the roads and sidewalks out of wood and elevate them above the waterline, in some places by several feet. On the day the fire started, over 55 miles of pine-block streets and 600 miles of wooden sidewalks bound the 23,000 acres of the city in a highly combustible knot.

6. Fires were common in all cities back then, and Chicago was no exception. In 1863 there had been 186 reported fires in Chicago; the number had risen to 515 by 1868. Records for 1870 indicate that fire-fighting companies responded to nearly 600 alarms. The next year saw even more fires spring up, mainly because the summer had been unusually dry. Between July and October only a few scattered showers had taken place and these did not produce much water at all. Trees drooped in the unrelenting summer sun; grass and leaves dried out. By October, as many as six fires were breaking out every day. On Saturday the seventh, the night before the Great Fire, a blaze destroyed four blocks and took over sixteen hours to control. What made Sunday the eighth different and particularly dangerous was the steady wind blowing in from the southwest.

7. It was this gusting, swirling wind that drove the flames from the O'Learys' barn into neighboring yards. To the east, a fence and shed of James Dalton's went up in flames; to the west, a barn smoldered for a few minutes, then flared up into a thousand yellow-orange fingers.

Used by permission of Scholastic, Inc.

TEXT COMPLEXITY ANALYSIS

The quantitative data and quantitative analysis for *The Great Fire* are provided below:

The Great Fire	Quantitative Measure #1		Quantitative Measure #2	Grade Band Placement
	Flesch-Kinkaid: 9.0		Lexile: 1000L	6-8

Category	Notes and comments on text, support for placement in this band	Where to place within the band?					
		Early 6-mid 6	Mid 6 - early 7	Early 7 - mid 7	Mid 7 - early 8	Mid 8 - end 8	NOT suited to band
Structure (story structure or form of piece or sentence demands if notable)	The structure of the text is mostly cause and effect, showing the main reasons the Great Fire started in Chicago when it did. The relationship between the main idea and supporting details is clear.	▲					→
Language Clarity and Conventions (including vocabulary load)	The vocabulary used in the text is accessible to the average sixth grader and appropriate for grade level. The few words that may be challenging for this audience are surrounded by strong context clues that will enable students to understand the unfamiliar terms. The sentence structure varies from simple to complex but are of average length and can be dissected easily if needed.	▲					→
Knowledge Demands (life, content, cultural/literary)	The passage is self-contained, meaning that no outside knowledge is required. Students may or may not know the location of Chicago, but a lack of knowledge of that fact will not impact understanding. Also, no prior knowledge of the Great Fire is needed, as the text describes it fully. Students will need to infer that wood burns easily, but there is context in the text to support that inference.	▲					→
Levels of Meaning (chiefly literary)/ Purpose (chiefly informational)	The purpose is singular – to explain the reasons the Great Fire started.	▲					→
Quantitative and qualitative data indicate that this excerpt from *The Great Fire* is a **low-complexity** grade **6** text.		▲					→

Achieve

APPENDIX B: "Is it a Crime for a Citizen of the United States to Vote?" by Susan B. Anthony and Text Complexity Analysis

1. Friends and Fellow-citizens: I stand before you tonight, under indictment for the alleged crime of having voted at the last Presidential election, without having a lawful right to vote. It shall be my work this evening to prove to you that in thus voting, I not only committed no crime, but, instead, simply exercised my citizen's right, guaranteed to me and all United States citizens by the National Constitution, beyond the power of any State to deny.

2. Our democratic-republican government is based on the idea of the natural right of every individual member thereof to a voice and a vote in making and executing the laws. We assert the province of government to be to secure the people in the enjoyment of their unalienable rights. We throw to the winds the old dogma that governments can give rights. Before governments were organized, no one denies that each individual possessed the right to protect his own life, liberty and property. And when 100 or 1,000,000 people enter into a free government, they do not barter away their natural rights; they simply pledge themselves to protect each other in the enjoyment of them, through prescribed judicial and legislative tribunals. They agree to abandon the methods of brute force in the adjustment of their differences, and adopt those of civilization.

3. Nor can you find a word in any of the grand documents left us by the fathers that assumes for government the power to create or to confer rights. The Declaration of Independence, the United States Constitution, the constitutions of the several states and the organic laws of the territories, all alike propose to protect the people in the exercise of their God-given rights. Not one of them pretends to bestow rights.

4. "All men are created equal, and endowed by their Creator with certain unalienable rights. Among these are life, liberty and the pursuit of happiness. That to secure these, governments are instituted among men, deriving their just powers from the consent of the governed."

5. Here is no shadow of government authority over rights, nor exclusion of any from their full and equal enjoyment. Here is pronounced the right of all men, and "consequently," as the Quaker preacher said, "of all women," to a voice in the government. And here, in this very first paragraph of the declaration, is the assertion of the natural right of all to the ballot; for, how can "the consent of the governed" be given, if the right to vote be denied. Again:

6. "That whenever any form of government becomes destructive of these ends, it is the right of the people to alter or abolish it, and to institute a new government, laying its foundations on such principles, and organizing its powers in such forms as to them shall seem most likely to effect their safety and happiness."

7. Surely, the right of the whole people to vote is here clearly implied. For however destructive in their happiness this government might become, a disfranchised class could neither alter nor abolish it, nor institute a new one, except by the old brute force method of insurrection and rebellion. One-half of the people of this nation to-day are utterly powerless to blot from the statute books an unjust law, or to write there a new and a just one. The women, dissatisfied as they are with this form of government, that enforces taxation without representation, — that compels them to obey laws to which they have never given their consent, — that imprisons and hangs them without a trial by a jury of their peers, that robs them, in marriage, of the custody of their own persons, wages and children, — are this half of the people left wholly at the mercy of the other half, in direct violation of the spirit and letter of the declarations of the framers of this government, every one of which was based on the immutable principle of equal rights to all. By those declarations, kings, priests, popes, aristocrats, were all alike dethroned, and placed on a common level politically, with the lowliest born subject or serf. By them, too, men, as such, were deprived of their divine right to rule, and placed on a political level with women. By the practice of those declarations all class and caste distinction will be abolished; and slave, serf, plebeian, wife, woman, all alike, bound [leap] from their subject position to the proud platform of equality.

PUBLIC DOMAIN

TEXT COMPLEXITY ANALYSIS

The quantitative data and quantitative analysis for "Is it a Crime for a Citizen of the United States to Vote?" are provided below:

Excerpt from "Is It a Crime for a Citizen of the United States to Vote?"	Quantitative Measure #1 Flesh Kinkaid: 12.9		Quantitative Measure #2 Lexile: 1420L	Grade Band Placement Grades 11-12		
Qualitative Analysis	Excerpt from "Is It a Crime for a Citizen of the United States to Vote?"		Where to place within the band?			
Category	Notes and comments on text, support for placement in this band	Early 9 to Mid 9	End 9 to Early 10	Mid to End 10	Early to Mid 11	End 11	NOT suited to band
Structure (both story structure or form of piece)	The structure is relatively straightforward; the organization is generally evident and logical. However, connections among some ideas, processes, or events are implicit and subtle.			▲			
Language Clarity and Conventions (including vocabulary load)	The language is generally complex, with abstract and archaic language commonly used. The academic vocabulary and domain-specific words are not usually defined within the text. Also, the text consists of many complex sentences with subordinate phrases and clauses.					▲	
Knowledge Demands (life, content, cultural/literary)	The subject matter of the text may involve some understanding of the history of the Women's Suffrage Movement, the Declaration of Independence, and the conditions for women at this point in U.S. history.				▲		
Levels of Meaning (chiefly literary)/ Purpose (chiefly informational)	The primary purpose of the text is stated explicitly.				▲		
Quantitative and qualitative data indicate that this excerpt from "Is it a Crime for a Citizen of the United States?" to vote is a **highly complex** grade **11** text.					▲		