Supplemental Information for Appendix A of the Common Core State Standards for English Language Arts and Literacy: New Research on Text Complexity

I. Summary Introduction

Appendix A of the Common Core State Standards (hereafter CCSS) contains a review of the research stressing the importance of being able to read complex text for success in college and career. The research shows that while the complexity of reading demands for college, career, and citizenship have held steady or risen over the past half century, the complexity of texts students are exposed to has steadily decreased in that same interval. In order to address this gap, the CCSS emphasize increasing the complexity of texts students read as a key element in improving reading comprehension.

The importance of text complexity to student success had been known for many years prior to the release of the CCSS, but its release spurred subsequent research that holds implications for how the CCSS define and measure text complexity. As a result of new research on the quantitative dimensions of text complexity called for at the time of the standards' release[1], this report expands upon the three-part model outlined in Appendix A of the CCSS in ELA/Literacy that blends quantitative and qualitative measures of text complexity with reader and task considerations. It also presents new field-tested tools for helping educators assess the qualitative features of text complexity.

II. New Findings Regarding the Quantitative Dimension of Text Complexity

The quantitative dimension of text complexity refers to those aspects—such as word frequency, sentence length, and text cohesion (to name just three)—that are difficult for a human reader to evaluate when examining a text. These factors are more efficiently measured by computer programs. The creators of several of these quantitative measures volunteered to take part in a research study comparing the different measurement systems against one another. The goal of the study was to provide state of the science information regarding the variety of ways text complexity can be measured quantitatively and to encourage the development of text complexity tools that are valid, transparent, user friendly, and reliable.[2] The six different computer programs that factored in the research study are briefly described below:

[1] The full report, *Measures of Text Difficulty*, and other resources, can be accessed on www.achievethecore.org/text-complexity.
[2] The following list of participants in the research study is not an exhaustive list of programs that exist for the purpose of measuring text complexity, nor is their inclusion intended as an endorsement of one method or program over another.

ATOS by Renaissance Learning

ATOS incorporates two formulas: ATOS for Text (which can be applied to virtually any text sample, including speeches, plays, and articles) and ATOS for Books. Both formulas take into account three variables: words per sentence, average grade level of words (established via the Graded Vocabulary List), and characters per word.

Degrees of Reading Power® (DRP®) by Questar Assessment, Inc.

The DRP Analyzer employs a derivation of a Bormuth mean cloze readability formula based on three measureable features of text: word length, sentence length, and word familiarity. DRP text difficulty is expressed in DRP units on a continuous scale with a theoretical range from 0 to 100. In practice, commonly encountered English text ranges from about 25 to 85 DRP units, with higher values representing more difficult text. Both the measurement of students' reading ability and the readability of instructional materials are reported on the same DRP scale.

Flesch-Kincaid (public domain)

Like many of the non-proprietary formulas for measuring the readability of various types of texts, the widely used Flesch-Kincaid Grade Level test considers two factors: words and sentences. In this case, Flesch-Kincaid uses word and sentence length as proxies for semantic and syntactic complexity respectively (i.e., proxies for vocabulary difficulty and sentence structure).

The Lexile® Framework For Reading by MetaMetrics

A Lexile measure represents both the complexity of a text, such as a book or article, and an individual's reading ability. Lexile® measures include the variables of word frequency and sentence length. Lexile® measures are expressed as numeric measures followed by an "L" (for example, 850L), which are then placed on the Lexile® scale for measuring reader ability and text complexity (ranging from below 200L for beginning readers and beginning-reader materials to above 1600L for advanced readers and materials).

Reading Maturity by Pearson Education

The Pearson Reading Maturity Metric uses the computational language model Latent Semantic Analysis (LSA) to estimate how much language experience is required to achieve adult knowledge of the meaning of each word, sentence, and paragraph in a text. It combines the Word Maturity measure with other computational linguistic variables such as perplexity, sentence length, and semantic coherence metrics to determine the overall difficulty and complexity of the language used in the text.

TextEvaluator by Educational Testing Service*

TextEvaluator uses natural language processing technologies to assess text standing relative to eight cognitively-based dimensions of text variation: academic vocabulary, general vocabulary, concreteness, syntactic complexity, referential cohesion, connective cohesion, degree of narrativity, and presence of interactive/conversational elements. Overall text complexity scores that incorporate this evidence are generated via three separate prediction models: one optimized for application to informational texts, one optimized for application to literary texts, and one optimized for application to mixed texts. Resulting measures can help educators determine which of eight construct-relevant sources of comprehension difficulty are likely to be most challenging within any specified text.

**ETS changed its measure from SourceRater to TextEvaluator in 2013. A third-party validation in 2014 showed that the changes to the measure yielded text complexity classifications aligned with the Common Core text complexity scale.*

Easability Indicator by Coh-Metrix

One additional program—the Coh-Metrix Easability Assessor, developed at the University of Memphis and Arizona State University—factored in the research study but was not included in the

cross analysis. It analyzes the ease or difficulty of texts on five different dimensions: narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion.[3] This measure was not included in the cross analysis because it does not generate a single quantitative determination of text complexity, but it does have use as a tool to help evaluate text systematically. The Coh-Metrix Easability Assessor creates a profile that offers information regarding the aforementioned features of a text and analyzes how challenging or supportive those features might be in student comprehension of the material.

The research that has yielded additional information and validated these text measurement tools was led by Jessica Nelson of Carnegie Mellon University, Charles Perfetti of University of Pittsburgh and David and Meredith Liben of Student Achievement Partners (in association with Susan Pimentel, lead author of the CCSS for ELA). It had two components: first, all the developers of quantitative tools agreed to compare the ability of each text analyzer to predict the difficulty of text passages as measured by student performances on standardized tests. Second, they agreed to test the tools' ability to predict expert judgment regarding grade placement of texts and educator evaluations of text complexity by examining a wide variety of text types selected for a wide variety of purposes. The first was measured by comparing student results in norming data on two national standardized reading assessments to the difficulty predicted by the text analyzer measures. The second set of data evaluated how well each text analyzer predicted educator judgment of grade level placement and how well they matched the complexity band placements used for the Appendix B texts of the CCSS. In the final phase of the work, the developers agreed to place their tools on a common scale aligned with the demands of college readiness. This allows these measures to be used with confidence when placing texts within grade bands, as the common scale ensures that each will yield equivalent complexity staircases for reaching college and career readiness levels of text complexity.[4]

The major comparability finding of the research was that all of the quantitative metrics were reliably and often highly correlated with grade level and student performance based measures of text difficulty across a variety of text sets and reference measures.[5] No one of the quantitative measures performed significantly differently than the others in predicting student outcomes.[6] While there is variance between and among the measures about where they place any single text, they all climb reliably—though differently—up the text complexity ladder to college and career readiness. Choosing any one of the text-analyzer tools from second grade through high school will provide a scale by which to rate text complexity over a student's career, culminating in levels that match college and career readiness.

In addition, the research produced a new common scale for cross comparisons of the quantitative tools that were part of the study, allowing users to choose one measure or another to generate parallel

[3] Narrativity measures whether the passage is story-like and includes events and characters. Syntactic simplicity refers to the ease of the sentence syntax. Word concreteness measures the degree to which words in the passage are imaginable versus abstract. Referential cohesion is the overlap between sentences with respect to major words (nouns, verbs, adjectives). Deep cohesion measures causal, spatial and temporal relations between events, actions, goals, and states.

[4] As a condition of participating, each developer also committed to offering (a) transparency in revealing both the text features it analyzed and the general means of analysis, (b) a program that calibrated text difficulty by grade or band level to match the Common Core Standards' expectations regarding measuring text complexity, and (c) a version of its quantitative tool that could be adapted for public access at the individual user level.

[5] When running the passages through Flesch-Kincaid measures, researchers found no single answer for what the Flesch-Kincaid score was for a specific text. The score depended on which version of the Flesch-Kincaid program was run and how that particular program counted syllables, sentence length, and the like. Because Flesch-Kincaid has no 'caretaker' that oversees or maintains the formula, researchers had to make decisions about how to count syllables and sentence length as they programmed the formula to get a 'read' on text(s).

[6] Some of the quantitative measures aligned more closely with human judgment regarding where to situate a text within a complexity band, though these measures did not better predict student performance.

complexity readings for texts as students move through their K-12 school careers. This common scale is anchored by the complexity of texts representative of those required in typical first-year credit-bearing college courses and in workforce training programs. Each of the measures has realigned its ranges to match the Standards' text complexity grade bands and has adjusted upward its trajectory of reading comprehension development through the grades to indicate that all students should be reading at the college and career readiness level by no later than the end of high school.

Figure 1: Updated Text Complexity Grade Bands and Associated Ranges from Multiple Measures [7]

Common Core Band	ATOS	Degrees of Reading Power®	Flesch-Kincaid[8]	The Lexile Framework®	Reading Maturity	TextEvaluator
2nd – 3rd	2.75 – 5.14	42 – 54	1.98 – 5.34	420 – 820	3.53 – 6.13	100 – 590
4th – 5th	4.97 – 7.03	52 – 60	4.51 – 7.73	740 – 1010	5.42 – 7.92	405 – 720
6th – 8th	7.00 – 9.98	57 – 67	6.51 – 10.34	925 – 1185	7.04 – 9.57	550 – 940
9th – 10th	9.67 – 12.01	62 – 72	8.32 – 12.12	1050 – 1335	8.41 – 10.81	750 – 1125
11th – CCR	11.20 – 14.10	67 – 74	10.34 – 14.2	1185 – 1385	9.57 – 12.00	890 – 1360

III. New Tools for Evaluating the Qualitative Dimension of Text Complexity

Simultaneously with the work on quantitative metrics, additional fieldwork was performed with the goal of helping educators better judge the qualitative features of text complexity. In the CCSS, qualitative measures serve as a necessary complement to quantitative measures, which cannot capture all of the elements that make a text easy or challenging to read and are not equally successful in rating the complexity of all categories of text.

Focus groups of teachers from a variety of CCSS adoption states, and representing a wide variety of teaching backgrounds, used the qualitative features first identified in Appendix A to develop and refine an evaluation tool that offers teachers and others greater guidance in rating texts. The evaluation tool views the four qualitative factors identified in Appendix A as lying on continua of difficulty rather than as a succession of discrete "stages" in text complexity. The qualitative factors run from easy (left-hand side) to difficult (right-hand side). Few (if any) authentic texts will be at the low or high ends on all of these measures, and some elements of the dimensions are better suited to literary or to informational texts. Below are brief descriptions of the different qualitative dimensions:

[7] The band levels themselves have been expanded slightly over the original CCSS scale that appears in Appendix A at both the top and bottom of each band to provide for a more modulated climb toward college and career readiness and offer slightly more overlap between bands. The wider band width allows more flexibility in the younger grades where students enter school with widely varied preparation levels. This change was provided in response to feedback received since publication of the original scale (published in terms of the Lexile® metric) in Appendix A.

[8] Since Flesch-Kincaid has no 'caretaker' that oversees or maintains the formula, the research leads worked to bring the measure in line with college and career readiness levels of text complexity based on the version of the formula used by Coh-Metrix.

(1) *Structure.* Texts of low complexity tend to have simple, well-marked, and conventional structures, whereas texts of high complexity tend to have complex, implicit, and (in literary texts) unconventional structures. Simple literary texts tend to relate events in chronological order, while complex literary texts make more frequent use of flashbacks, flash-forwards, multiple points of view and other manipulations of time and sequence. Simple informational texts are likely not to deviate from the conventions of common genres and subgenres, while complex informational texts might if they are conforming to the norms and conventions of a specific discipline or if they contain a variety of structures (as an academic textbook or history book might). Graphics tend to be simple and either unnecessary or merely supplementary to the meaning of texts of low complexity, whereas texts of high complexity tend to have similarly complex graphics that provide an independent source of information and are essential to understanding a text. (Note that many books for the youngest students rely heavily on graphics to convey meaning and are an exception to the above generalization.)

(2) *Language Conventionality and Clarity.* Texts that rely on literal, clear, contemporary, and conversational language tend to be easier to read than texts that rely on figurative, ironic, ambiguous, purposefully misleading, archaic, or otherwise unfamiliar language (such as general academic and domain-specific vocabulary).

(3) *Knowledge Demands.* Texts that make few assumptions about the extent of readers' life experiences and the depth of their cultural/literary and content/discipline knowledge are generally less complex than are texts that make many assumptions in one or more of those areas.

(4) *Levels of Meaning (literary texts) or Purpose (informational texts).* Literary texts with a single level of meaning tend to be easier to read than literary texts with multiple levels of meaning (such as satires, in which the author's literal message is intentionally at odds with his or her underlying message). Similarly, informational texts with an explicitly stated purpose are generally easier to comprehend than informational texts with an implicit, hidden, or obscure purpose.

Figure 2: Qualitative Dimensions of Text Complexity

Category	Notes and comments on text, support for placement in this band	Where to place within the band?				
		Beginning of lower grade	End of lower grade	Beginning of higher grade	End of higher grade	**NOT suited to band**
Structure (both story structure or form of piece)						
Language Clarity and Conventions (including vocabulary load)						
Knowledge Demands (life, content, cultural/literary)						
Levels of Meaning/ Purpose						
Overall placement	**Justification**					

IV. Reader and Task Considerations and the Role of Teachers

While the research noted above impacts the quantitative and qualitative measures of text complexity, the third element of the three-part model for measuring text complexity—reader and task considerations—remains untouched. While the quantitative and qualitative measures focus on the inherent complexity of the text, they are balanced in the CCSS' model by the expectation that educators will employ professional judgment to match texts to particular tasks or classes of students. Numerous considerations go into such matching. For example, harder texts may be appropriate for highly knowledgeable or skilled readers, who are often willing to put in the extra effort required to read harder texts that tell a story or contain complex information. Students who have a great deal of interest or motivation in the content are also likely to handle more complex texts.

The RAND Reading Study Group, identified in the 2002 report *Reading for Understanding,* also named important task-related variables, including the reader's purpose (which might shift over the course of reading), "the type of reading being done, such as skimming (getting the gist of the text) or studying (reading the text with the intent of retaining the information for a period of time)," and the intended outcome, which could include "an increase in knowledge, a solution to some real-world problem, and/or engagement with the text."[9] Teachers employing their professional judgment, experience, and knowledge of their students and their subject are best situated to make such appraisals.

V. The Issue of Text Quality and Coherence in Text Selection

Selecting texts for student reading should not only depend on text complexity but also on considerations of quality and coherence. The Common Core State Standards emphasize that "[t]o become college and career ready, students must grapple with works of exceptional craft and thought whose range extends across genres, cultures, and centuries. Such works offer profound insights into the human condition and serve as models for students' own thinking and writing."[10] In addition to choosing high quality texts, it is also recommended that texts be selected to build coherent knowledge within grades and across grades. For example, the Common Core State Standards illustrate a progression of selected texts across grades K-5 that systematically build knowledge regarding the human body.[11] Considerations of quality and coherence should always be at play when selecting texts.

VI. Key Considerations in Implementing Text Complexity

The tools for measuring text complexity are at once useful and imperfect. Each of the tools described above—quantitative and qualitative—has its limitations, and none is completely accurate. The question remains as to how to best integrate quantitative measures with qualitative measures when locating texts at a grade level. The fact that the quantitative measures operate in bands rather than specific grades gives room for both qualitative and quantitative factors to work in concert when situating texts. The following recommendations that play to the strengths of each type of tool—quantitative and qualitative—are offered as guidance in selecting and placing texts:

1. *It is recommended that quantitative measures be used to locate a text within a grade band because they measure dimensions of text complexity—such as word frequency, sentence length, and text cohesion (to name just three)—that are difficult for a human reader to evaluate when examining a text.* In high stakes settings, it is recommended that two or more quantitative measures be used to locate a text within a grade band for a most reliable indicator that text falls within the complexity range for that band.

2. *It is further recommended that qualitative measures be used to then locate a text in a specific grade*. Qualitative measures are neither grade nor grade band specific, nor anchored in college and career readiness levels. Once a text is located within a band with quantitative measures, they can be used

[9] RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND. The quoted text appears in pages xiii–xvi.
[10] CCSS, pg. 35.
[11] CCSS, pg. 33.

to measure other important aspects of texts—such as levels of meaning or purpose, structure, language conventionality and clarity, and knowledge demands—to further locate a text at the high or low end of the band or to a specific grade. For example, one of the quantitative measures could be used to determine that a text falls within the grades 6-8 band level, and qualitative measures could then be used to determine whether the text is best placed in grade 6, 7, or 8.

3. *There will be exceptions to using quantitative measures to identify the grade band; sometimes qualitative considerations will trump quantitative measures in identifying the grade band of a text, particularly with narrative fiction in later grades.* Research showed more disagreement among the quantitative measures when applied to narrative fiction in higher complexity bands than with informational text or texts in lower grade bands. Given this, preference should sometimes be given to qualitative measures when evaluating narrative fiction intended for students in grade 6 and above. For example, some widely used quantitative measures rate the Pulitzer Prize-winning novel *Grapes of Wrath* as appropriate for grades 2–3. This counterintuitive result emerges because works such as *Grapes* often express complex ideas or mature themes in relatively commonplace language (familiar words and simple syntax), especially in the form of dialogue that mimics everyday speech. Such quantitative exceptions for narrative fiction should be carefully considered, and exceptions should be rarely exercised with other kinds of text. It is critical that in every ELA classroom students have adequate practice with literary non-fiction that falls within the quantitative band for that grade level. To maintain overall comparability in expectations and exposure for students, the overwhelming majority of texts that students read in a given year should fall within the quantitative range for that band.

4. *Certain measures are less valid or not applicable for certain kinds of texts.* Until such time as quantitative tools for capturing the difficulty of poetry and drama are developed, determining whether a poem or play is appropriately complex for a given grade or grade band will necessarily be a matter of qualitative assessment meshed with reader-task considerations. Furthermore, texts for kindergarten and grade 1 are still resistant to quantitative analysis, as they often contain difficult-to-assess features designed to aid early readers in acquiring written language. (The Standards' Appendix B poetry and K–1 text exemplars were placed into grade bands by expert teachers drawing on classroom experience.)

VII. The Model in Action: Sample Annotated Reading Text

The following example demonstrates how quantitative and qualitative measures of text complexity can be used along with reader and task considerations to make informed decisions about whether a particular text is an appropriate challenge for particular students. The case below illustrates some of the intricacies that can arise when multiple measures are used to assess text complexity.

Example: *The Longitude Prize* (Grades 9–10 Text Complexity Band)

Excerpt

From Chapter 1: "A Most Terrible Sea"

At six in the morning I was awaked by a great shock, and a confused noise of the men on deck. I ran up, thinking some ship had run foul of us, for by my own reckoning, and that of every other person in the ship, we were at least thirty-five leagues distant from land; but, before I could reach the quarter-deck, the ship gave a great stroke upon the ground, and the sea broke over her. Just after this I could perceive the land, rocky, rugged and uneven, about two cables' length from us . . . the masts soon went overboard, carrying some men with them . . . notwithstanding a most terrible sea, one of the [lifeboats] was launched, and eight of the best men jumped into her; but she had scarcely got to the ship's stern when she was hurled to the bottom, and every soul in her perished. The rest of the boats were soon washed to pieces on the deck. We then made a raft . . . and waited with resignation for Providence to assist us.

—From an account of the wreck of HMS *Litchfield* off the coast of North Africa, 1758

The *Litchfield* came to grief because no one aboard knew where they were. As the narrator tells us, by his own reckoning and that of everyone else they were supposed to be thirty-five leagues, about a hundred miles, from land. The word "reckoning" was short for "dead reckoning"—the system used by ships at sea to keep track of their position, meaning their longitude and latitude. It was an intricate system, a craft, and like every other craft involved the mastery of certain tools, in this case such instruments as compass, hourglass, and quadrant. It was an art as well.

Latitude, the north-south position, had always been the navigator's faithful guide. Even in ancient times, a Greek or Roman sailor could tell how far north of the equator he was by observing the North Star's height above the horizon, or the sun's at noon. This could be done without instruments, trusting in experience and the naked eye, although it is believed that an ancestor of the quadrant called the astrolabe—"star-measurer"—was known to the ancients, and used by them to measure the angular height of the sun or a star above the horizon.

Phoenicians, Greeks, and Romans tended to sail along the coasts and were rarely out of sight of land. As later navigators left the safety of the Mediterranean to plunge into the vast Atlantic—far from shore, and from the shorebirds that led them to it—they still had the sun and the North Star. And these enabled them to follow imagined parallel lines of latitude that circle the globe. Following a line of latitude—"sailing the parallel"—kept a ship on a steady east-west course. Christopher Columbus, who sailed the parallel in 1492, held his ships on such a safe course, west and west again, straight on toward Asia. When they came across an island off the coast of what would later be called America, Columbus compelled his crew to sign an affidavit stating that this island was no island but mainland Asia.

Dash, Joan. *The Longitude Prize*.
New York: Farrar, Straus and Giroux (2000).

Figure 3: Annotation of The Longitude Prize

Qualitative Measures

Structure

The text is moderately complex and subtle in structure. Although the text may appear at first glance to be a conventional narrative, Dash mainly uses narrative elements in the service of illustrating historical and technical points. The long quote adds to the structural challenge.

Language Conventionality and Clarity

Language is used literally and is relatively clear, but numerous archaic, domain-specific, and otherwise unfamiliar terms are introduced in the course of citing primary historical sources and discussing the craft, art, and science of navigation. The quote further adds an archaic language burden.

Knowledge Demands

The text assumes relatively little prior knowledge regarding seafaring and navigation, but some general sense of the concepts of latitude and longitude, the nature of sailing ships, and the historical circumstances that promoted exploration and trade is useful to comprehending the text.

Purpose

The single, relatively clear purpose of the text (not fully apparent in the excerpt but signaled by the title) is to recount the discovery of the concept of longitude. But this is not readily apparent from the excerpt.

Quantitative Measures

Various readability measures of The Longitude Prizeare largely in agreement that the text is appropriate for the grades 9–10 text complexity band. The Coh-Metrix analysis notes that the text is primarily informational in structure despite the narrative opening. (Recall from "Why Text Complexity Matters," above, that research indicates that informational texts are generally harder to read than narratives.) While the text relies on concrete language and goes to some effort to connect central ideas for the reader, it also contains complex syntax and few explicit connections between words and sentences.

Reader-Task Considerations

These are to be determined locally with reference to such variables as a student's motivation, knowledge, and experiences as well as purpose and the complexity of the task assigned and the questions posed.

Recommended Placement

Various quantitative measurements place The Longitude Prizeinto the grades 9–10 text complexity band; the qualitative analysis would indicate there are enough complex features to warrant its placement in the tenth grade.

ATOS: 10.5
DRP®: 66
Lexile®: 1300L
Reading Maturity: 8.67
TextEvaluator: 1010