Stretching Elementary Students in Complex Text: Why? How? When?

Heidi Ann Mesmer
Virginia Tech

TextProject Article Series
February 2014

TextProject, Inc.
SANTA CRUZ, CALIFORNIA

To appear in E.H. Hiebert (Ed.) (March, 2014). Stamina*, Silent Reading, & the Common Core State Standards.* Santa Cruz, CA: TextProject.

Stretching Elementary Students in Complex Text: Why? How? When?

Heidi Anne E. Mesmer

Virginia Tech

"Stretching students in text? What does that mean? Put them on a rack?" a third-grade teacher mischievously commented at a recent professional development workshop. I had to bite my tongue because, in truth, I also find the phrase a little odd. As a teacher, I thought in terms of instructional level as I looked for reading materials that would challenge my students. But *stretch* texts? Never. That is, not until the arrival of the Common Core State Standards for the English Language Arts (CCSS/ELA) (National Governors Association Center for Best Practices [NGA Center] & Council of Chief State School Officers [CCSSO], 2010) and their focus on providing students with opportunities to read increasingly complex texts over a grade span—and so "stretch" their reading abilities.

The CCSS/ELA goal is certainly worthy. However, it also raises numerous issues for teachers who are charged with selecting materials and providing instruction that will help students achieve that goal. In this chapter, I address several of these issues. I begin with a discussion of the meaning of text complexity, both how the CCSS/ELA developers define it, as well as how it is defined from other perspectives. I then discuss what constitutes stretch text and how the introduction of the stretch notion at the elementary school level will influence reader-text matching paradigms. Next, I present a series of rationales, both good and bad, that are used to bolster arguments to stretch students in text. In the last section of the chapter, I offer an extended discussion of the factors that may contribute to or inhibit students being stretched in text. In each section, I give attention to the gaps between what we know from the research

literature and the type of information needed if students are going to reach the high aspirations of the CCSS.

What Is Text Complexity?

The CCSS/ELA document is unprecedented in its focus on text. No other standards document in recent history has addressed text with greater attention, specificity, or energy. Whereas previous documents use general terms such as "on grade level" to identify the kinds of texts students should read, the CCSS/ELA developers discuss texts in terms of levels of complexity, which they defines as "The inherent difficulty of reading and comprehending a text combined with consideration of reader and task variables" (NGA Center for Best Practices & CCSSO, 2010b, p. 43). Thus, in the CCSS/ELA document, the term *complexity* is used interchangeably with *text difficulty* (a conflation with which I differ later in this chapter).

A review of the model of text-complexity assessment articulated in the standards' Appendix A brings the CCSS/ELA's definition of the term into clearer focus. This tripartite model encompasses qualitative tools, quantitative tools, and reader and task variables that capture the complexity of a text and so make that text difficult for an individual student.

The first part of the model consists of qualitative measures and human (rather than computer) evaluators. Trained teachers or researchers apply their professional judgment to estimate the complexity of a text for target readers. According to the CCSS/ELA, the text features best evaluated using human judgment include: (a) levels of meaning in literary texts and levels of purpose in informational texts; (b) text structures (e.g., simple, well-marked structures vs. implicit and layered structures); (c) language conventionality and clarity (e.g., literal, clear language vs. figurative, academic, or domain specific vocabulary); and (d) knowledge demands (e.g., level of knowledge assumed by the text). It should be noted here that the reliability and

soundness of the qualitative component of the CCSS/ELA model, while theoretically interesting, has not been established by research.

The second part of the text-complexity model relates to reader and task factors, elements that are not inherent to the text itself. (From my perspective, these are part of the reader-text match but not really the *assessment* of *text complexity.*) Appearing to draw from the reader-text-task model found in the RAND report on reading comprehension (Snow, 2002), the CCSS/ELA developers remind us that reader variables such as motivation, background knowledge, and personal experiences will all render a text more or less difficult to a group of readers. Additionally, the standards give focus to task variables such as purpose, assignment requirements, and teacher levels of expectation, reminding us that the analysis of text complexity as it relates to reader and task is best done by teachers.

The third part of the model—the quantitative systems, or the readability formulas—is the one supported by the most validation and reliability and the longest history (Harrison, 1980; Mesmer, 2008). These traditional formulas (e.g., Dale-Chall, Fry, Spache), and their second-generation digitally calculated cousins (e.g., Lexiles, ATOS, Degrees of Reading Power) theoretically work the same way. Both estimate difficulty by using a word factor, usually an estimate of word frequency, and a syntactic factor, usually the length of sentences. Labels such as *grades, Lexiles*, or *Degrees of Reading Power* are generated for texts and used to estimate their difficulty.

Although there are five readability systems that express text difficulty in terms of grade levels, two systems, Degrees of Reading Power, and Lexiles, continue to figure prominently in text-complexity determination. The Lexile Framework is used extensively in numerous state assessments as well as in the National Assessment of Educational Progress (NAEP). The

framework is based on word frequency and sentence length, and it uses Lexiles (L) rather than grade levels as a unit of text difficulty. One Lexile is "1/1000th of the difference between the mean difficulty of mid-first grade material and the mean of difficulty of college and workplace passages" (Stenner et al., 2007).

Another tool CCSS/ELA identifies in this part of the model is the Coh-Metrix system, which measures text cohesion through a myriad of text features (e.g., anaphora, cross-sentence referents).

The CCSS/ELA's Appendix A precisely specifies the quantitative guidelines for requisite text complexity across several grade-level bands. A recent supplement to Appendix A extended the range of these bands (Nelson, Perfetti, Liben, & Liben, 2012). At each grade-level band, a precise range of text difficulty, as measured by the various readability formulas, is prescribed. By ascending this "staircase of text complexity" (see Table 1), students are expected to arrive at college and career reading levels by high school graduation (NGA Center for Best Practices & CCSSO, 2010b, p. 8).

The CCSS/ELA staircase does not assign a text-complexity range for K–grade 1, but a default level is set by the entering value for the grades 2-3 band. First-grade children must reach the minimal level at the bottom of the 2-3 band entry (420L) by the end of the school year. Note that the levels of text complexity expected at various grades are somewhat accelerated. Whereas schools would typically expect students at the end of the grade-3 year to read at a grade-4 level, the CCSS/ELA staircase sets a level of about grade 5 or 6.

Lest anyone think the staircase is merely suggestive, the language of Reading Standard 10, the text-type standard for each grade band, indicates that the text-difficulty ranges are concrete expectations. For instance, the grade-3 informational text Standard 10 states that, by the

end of the school year, students will "Comprehend informational texts . . . at the high end of the grades 2–3 text complexity band **independently and proficiently"** (NGA Center for Best Practices & CCSSO, 2010a, p.12, emphasis added**).** Thus, although the CCSS/ELA offers three ways to assess text complexity, the quantitative tools are the most specific and the most translatable into classrooms and, more importantly, into publishers' guidelines.

Text Complexity* vs. *Text Difficulty

Note the appearance of the term *text difficulty* in the previous discussion. As I mentioned earlier, the CCSS/ELA developers use the terms *text complexity* and *text difficulty* interchangeably. However, Mesmer, Cunningham, and Hiebert (2012) differentiate the two, noting that this, distinction is important if research and development are to move forward. *Text difficulty* implies a relationship between texts and readers. And, as the researchers state, "The difficulty of a text or text feature always implies a dependent or criterion variable: the actual or predicted performance of multiple readers on a task based on that text or feature" (p. 236). *Text complexity* is simply the naturally occurring textual elements in a passage or book that can be analyzed, manipulated or otherwise studied and is, as such, an independent variable.

Text-difficulty estimates, such as those created by readability formulas, connect the complexity of a text (e.g., word frequency and sentence length) to reader performance (i.e., readers' comprehension of a text) or predicted performance (e.g., another formula's estimate of difficulty, teacher's estimates of difficulty). Conflating the terms *text complexity* and *text difficulty* confuses causes with effects. Furthermore, the confusion obscures the fact that text difficulty is not one dimensional but a numeric expression of a relationship. It is not a feature that is intrinsic to the text. The estimate of text difficulty is only as good as the relationship upon which the estimate is based, and the complexity of a text is simply what's there. If we are to

"stretch" students in text, we must depend on the very best estimates of text difficulty available, and we must better understand the impact of various text-complexity features on readers' comprehension.

What Is *Stretch* Text?

Like the third-grade teacher I mentioned in the introduction, I had not heard the term "stretch text" until recently. I was vaguely familiar with secondary school English teachers' use of stretch texts, but I had never heard the term applied to elementary school materials. In fact, although the CCSS/ELA standards are replete with the theme of challenging text, the term *stretch* is mentioned only once—in Appendix A:

> Students need opportunities to *stretch* their reading abilities but also to experience the satisfaction and pleasure of easy, fluent reading within them, both of which the Standards allow for. . . . Students deeply interested in a given topic, for example, may engage with texts on that subject across a range of complexity. Particular tasks may also require students to read harder texts than they would normally be required to. (NGA Center for Best Practices & CCCSO, 2010b, p. 9, italics mine)

However, the term has caught on as a synonym for *challenging texts,* and the concept it represents has won endorsement from several researchers (e.g., Roskos & Neuman, 2013; Shanahan, 2011).

The stretch concept is also prompting the infusion of more challenging texts into our classrooms. In fact, the CCSS/ELA developers have already produced a document to guide educational publishers in the creation of more challenging materials (Coleman & Pimentel, 2012). This document clearly specifies that the complexity of texts should be aligned with the Appendix A staircase. Thus, as educational publishers move to translate the CCSS/ELA text

parameters into their programs, these parameters will have clear implications for both educational policy and practice.

At the elementary school level, the emphasis on giving students challenging texts introduces a paradigm shift in reader-text matching that contrasts with decades-old emphases on the avoidance of reader frustration. The guidelines that most elementary teachers now use for reader-text matching are the word accuracy and comprehension levels established by Betts (1946): (a) independent: texts that students read without teacher support (word accuracy: 99–100%; comprehension: 90–100%); (b) instructional: texts that students read with teacher support (word accuracy: 95–98%; comprehension: 75–80%); and (c) frustrational—texts that are inaccessible to students with or without support (word accuracy: = /<94%; comprehension: = /<74%). These boundaries for text difficulty have become the essential guidance through the present day, with many questions asked across the years about their empirical basis (Clay, 1985; Ehri, Dreyer, Flugman, & Gross, 2007; Ekwall, Solis, & Solis, 1973; Halladay, 2012; Morgan, Wilcox, & Eldredge, 2000; Pikulski & Shanahan, 1982; Stahl & Heubach, 2005). However, if teachers were to call any texts "stretch texts," they would likely identify the instructional-level texts as such.

Inadvertently, the Betts's labels and reader-text matching standards may have shaped the views of many text researchers and teachers. They may not have balanced their intense concern for avoiding reader frustration with the equally important message to encourage challenge and avoid stagnation. These older paradigms have focused on finding the "just right" text as measured at a specific point in time. However, it is possible to build capacity for readers to handle more difficult passages. Although the text-complexity staircase introduces many valid concerns, the theme of the Standard 10, to embrace challenge, is a message that is long overdue.

Unfortunately, just as the reader-text standards of the previous decades lacked empirical basis, so also does the stretch paradigm. We simply do not have an empirically based paradigm for how to challenge students in texts. We do not know exactly how far students can be pushed before they break and reading becomes incomprehensible and cognitively, psychologically, or emotionally exasperating. We do not know which text features can be ramped up and which must only be gently accelerated. We do not know at which points developmentally students can be stretched and within which contexts. Of course this all begs the question as to why the CCSS/ELA developers introduced the text-complexity standard and its surrounding verbiage. What exactly has happened to cause the standards writers to be concerned about the level of texts that students are reading?

Why Stretch Text?

The CCSS/ELA Rationale for Stretch Text

A careful reading of the CCSS/ELA materials and analysis of the themes and messages coming from the National Governor's Association Center for Best Practices suggests that text complexity or difficulty was made a key standards' focus for three reasons. The first, and perhaps most obvious reason provided is that United States students have not performed well on international measures of achievement over the last 10 years (e.g., Martin, Mullis, & Kennedy, 2007). Despite the high levels of education spending in the this country, students, particularly those from lower SES backgrounds and minority groups, are simply not achieving at acceptable levels.

The second reason for the CCSS/ELA emphasis is what I call the text-complexity gap between high school and college texts. To illustrate this gap, the CCSS/ELA developers cite a 2006 ACT report, *Reading Between the Lines,* which indicated that the success of students in a

college social science college course (i.e., grade of B or better) was predicted by the difficulty of texts to which they were exposed in high school. In other words, if the complexity of materials in high school was watered down, that limited students' abilities to achieve in college. Prior to the standards' text staircase, the difficulty level of materials required at the end of high school was much lower (1215L) than was that of materials required for college and career (1355L). CCSS/ELA Appendix A also cites studies by Chall (1977) and Hayes, Wolfer, and Wolfe (1996) that confirm the easing of text difficulty levels across secondary schools over several decades. However, this dumbing-down trend did not hold true for elementary schools (Hiebert &Mesmer, 2013). The increases in text difficulty reflected in the CCSS/ELA's staircase are an attempt to distribute text difficulty across all grades, to spread increases incrementally across grade levels. Thus, there is some evidence to support the need for increases in text difficulty, but that evidence exists as the secondary level and not at the elementary levels, where increases are also seen.

A third CCSS/ELA rationale for increasing text difficulty and emphasizing the stretch concept is provided the document's citation of a study conducted by Williamson (2006). The study followed a cohort of over 60,000 third graders, beginning in 1999, through their eighth-grade year. Using the North Caroline end-of-grade test in reading, the study tracked progress of students in Lexile levels and contrasted this progress with the levels of the typical textbooks. The results showed that the achievement of students was very close to, and perhaps limited by, the difficulty of texts. The findings logically suggested that if text levels were increased, student levels of achievement might also increase. Note that this pattern matches that seen in the ACT (2006) study above. What the findings do not indicate is the degree to which the student sample in North Carolina matched national samples. During this time period, North Carolina was

showing high levels of achievement on the NAEP, and it is possible that this sample was unlike national populations.

In sum, the reasons the CCSS/ELA developers provide for stretching students' abilities with increasingly challenging text are based on international comparisons, trends from secondary schools, and a study conducted with a sample of students from one state. The text-difficulty staircase, especially at the elementary levels, therefore has about as strong an empirical base as do the Betts (1946) guidelines. I stress, though, that whereas the reasons given in the CCSS/ELA for stretching elementary students are rather insubstantial, the standards' focus on challenge and on balancing reader-text paradigms is important.

How and When to Stretch Elementary Students

Because stretching, or challenging students in text must be based upon some starting point, I begin this section with a brief review of the basics of reader-text matching. I then discuss additional factors that relate to stretch, including text levels, text length, genre, and cohesion. I conclude the section by addressing the importance of viewing stretching students in text as a consistently applied, long-term approach rather than a hit-or-miss activity.

Stretch Toolkit

The very first steps taken to stretch students must begin with the basics of reader-text matching. When teachers challenge students in text, they should not arbitrarily ask the whole class to read some designated stretch text. The reader-text matching process begins with knowing the students' reading levels and then having some estimate of the levels of the texts. I still find it rather commonplace to ask a group of intermediate grade teachers, "How do you find out the reading levels of your students?" and be greeted with blank stares. I always remind teachers to start somewhere. Use a basic method for establishing reading level, however imperfect it may be

(e.g., Informal Reading Inventories, STAR test), and then try to get an estimate of text difficulty in the same measurement unit. For example, if the reading level is obtained in grade levels (e.g., 2.1, 2.2, 2.3) through the STAR reading test, then the estimate of text difficulty should be given in grade-levels as well.

I use the word *estimate* very purposefully. Until the student is actually reading the text, then both the reading level and the text difficulty are estimates. Text-difficulty measures, in particular, only provide a basic approximation. The actual difficulty of a text is resolved when an actual reader is reading that text. Although knowing a student's reading level and a text's estimated difficulty level provides a good place to start, standards for stretching students to read more complex texts still have not been conclusively established.

Text Levels

Beyond the Betts's criteria for word accuracy and comprehension, little is known about exactly how much above a student's instructional level a text may be before it becomes frustrational text. We do not know what the tipping point is. Several researchers have experimented with the degree to which word accuracy levels can dip below the accepted 90%. They have found that, with a great deal of rereading and teacher support, younger students can achieve at least 85% word accuracy and still show reasonable comprehension (Ehri, Dreyer, Flugman, & Gross, 2007; Hiebert, 2005; Stahl & Heubach, 2005).

Only a few studies suggest the degree to which text levels can surpass students' instructional reading levels. Morgan, Wilcox, and Eldredge (2000) used a partner reading intervention in which low-performing second graders were randomly assigned to read at three different text levels: on-grade level (e.g., their instructional level), two grades above instructional level, and four grades above instructional level. Students read with partners for 15 minutes per

day for 95 days, and results showed the most improvements for students reading two grades above instructional level. What is not known from this study are the actual reading levels of the students prior to and after the intervention. Below-level students starting out at a preprimer level are likely quite different from below-level students starting out at a first-grade reading level.

My colleague and I (Hiebert & Mesmer, 2011) analyzed a data set of 9,535 records of third graders' text comprehension, giving attention to cases in which students were reading in texts that were above or below their targeted instructional reading level. Essentially, the difficulty of each text read by each student was labeled relative to the student's instructional level (called "target" level). The target level is the point at which readers can comprehend 70% of the material. Thus, it was possible to identify cases in which the texts read were specific amounts above or below the student's target level. Because we used the Lexile framework, we divided the records into the following categories: (a) Easy texts—101+ below level; (b) On-Level—100L below to 50L above level; (c) Stretch—51-100L above level; and (d) Difficult—101+ above level. The *stretch* text levels were one standard deviation above the targeted *on-level* designation. We also separated students into two proficiency groups, those whose targeted levels were above grade level, and those who targeted level was below the grade-level range. The CCSS/ELA defines the range of text difficulty for the 2–3 grade band as 450-790L, and so below-level readers were defined as those reading below 450L. Students reading at or above 450L were designated as on-level readers.

In the study we found main effects for the text-difficulty and reader levels, indicating that both influenced readers' comprehension. The reader by text difficulty interaction was also significant as were differences for all text and reader combinations except for the *difficult* texts. On average, all students achieved a 61% reading comprehension in *stretch* texts that averaged

76L above their target levels. Below-level readers comprehended at a lower level than on-level readers at all text levels except in the *difficult* texts where all readers comprehended at about 53%. Across reader levels, performance declined as text difficulty increased, with comprehension levels dipping below 70% in the *stretch* texts.

The patterns of response to text difficulty changes differed somewhat for below and on-level readers. On-and below-level readers had nearly identical performance in *difficult* texts that were about 200L above level but, unlike on-level students, below-level readers' performance across text-difficulty categories reflected a curvilinear pattern. Their performance dipped more sharply beginning with the *on-level* texts and sloped more steeply than the on-level readers. In fact, performance for below-level readers did not reach the designated 70% for any texts except *easy* texts.

Our exploratory work indicated that no readers could be stretched 200L above their targeted levels (about two grade levels), and that students' background, or reading history, influenced the degree to which they could be stretched. Because the levels of text were always relative to the student's own level, readers were not reading at the same text levels but in texts that were harder or easier for *them*. What happened was that below-level readers, even in on-level texts, were still not performing well. In a sense, an on –level text *was a stretch text for them.* These preliminary results suggested that the readers' background was influencing their performance in text and that readers who had a history of reading below level could not be stretched to the same degree that others could.

Text Length

Although rarely mentioned in schemes for stretching students in text, length definitely is among the elements that make a text challenging. The CCSS/ELA developers, however, make no

mention of text length other than to suggest to secondary teachers that short, dense texts are good exemplars for supporting students in close reading and answering text-based questions (NGA Center for Best Practices & CCCSO, 2010b). In the elementary grades, text length is particularly important because it changes significantly across the grades. The average length of a passage, book, or text that students to read in first grade is about 50–250 words. By fourth grade, the average length of a textbook passage, chapter, or worksheet passage is 2,000 words, an eightfold increase. Several authors have found text length (i.e., number of words) to predict text level in the Reading Recovery instructional scheme, suggesting that length factors into challenge in the primary grades (Cunningham, Spadorcia, Erickson, Koppenhaver, Sturm, & Yoder, 2005; Hatcher, 2000). Certainly as length shifts so also does a set of reading behaviors. Students in the primary grades orally read short blocks of text with supportive pictures in a matter of minutes. However, students in the intermediate grades must read extended texts silently without pictures for upwards of 20 minutes.

In spite of its potential contribution to text difficulty, only a few studies have examined text length as a factor in reading performance (Calfee & Hiebert, 2011; Hiebert, Wilson, & Trainin, 2010; Mesmer & Hiebert, 2013). Calfee and Hiebert (2011) found, for example, that length could be a variable explaining the different levels of achievement for California fourth graders on the NAEP and the California State Test (CST). The percentage of students who scored at or above the proficient level on the CST was 38% higher than for the NAEP. Significant length differences characterize these two tests, with the NAEP passages being about 800–1,000 words and the CST passages running 350–400 words.

Hiebert, Wilson, and Trainin (2010) investigated how students of different proficiency levels performed at different points in a lengthy reading assessment passage. They found that

Students who were in the two lower quartiles showed reasonable rates and levels of comprehension in the beginning portions of the passage but had significantly depreciated comprehension scores in the latter portions. These findings suggest that stamina effects were at play, as students tired toward the end of the passage.

My colleague and I (Mesmer & Hiebert, 2013) manipulated text length and difficulty to identify the degree to which length and difficulty interacted and how students of differing proficiency levels were impacted by this combination (i.e., at-level and below level). Three different sets of text passages were designed at three difficulty levels (400L, 600L, 800L). Within each difficulty level, one passage was 200 words and one 1,000 words. Topics were kept consistent across a difficulty level (e.g., schools and community helpers for 400L-level texts; budgets and money for 600L-level texts; and natural resources and oil for 800L-level texts). Using a within-subjects design, we required all students to read all passages, with comprehension being the outcome variable. Findings showed that as texts became more difficult and longer, comprehension decreased. Students comprehended the short versions at every difficulty levels better than long versions. The results suggested that length compounded the effects of difficulty, rendering texts of the same difficulty level harder.

Existing research, then, suggests that, as educators continue to work to find ways to stretch elementary students in texts, they need to pay attention to text length . It is especially important for them to understand how to support students as they confront the length shifts at various developmental junctures, such as the movement toward reading chapter books in late first grade or early second grade and the shift from predominately short narrative to lengthy expository texts in late third grade/early fourth grade. At this later point, of course, genre also presents challenges to students.

Genre

When it comes to stretching, or challenging elementary students in text, genre appears to have multiple points of impact. Although there is a great deal of debate in the field about exactly where to draw the genre lines, it appears that a convenient way to think about genre is to divide texts into narratives or expository works (Mesmer, Cunningham, & Hiebert, 2012). What is known about genre is that expository texts tend to be dense with new, unknown vocabulary, often the type of domain-specific Tier 3 words that represent complicated concepts or processes (Beck, McKeown, & Kucan, 2013; Fang, 2006). In narrative texts, students also are introduced to new vocabulary, but frequently the words are Tier 2 words that enhance the meaning of the text, express degree, or modify the core of a sentence. Note the text examples in Table 2. The words in the narrative text are, in some cases, easily inferable, as they are compound words or ancillary to the passage. In the expository text, however, the identifying *poaching* as either a process or and action and recognizing the noun *ivory* are essential to understanding the text. If a reader does not know the meaning of these words, getting the gist of the passage is not possible.

Both Hiebert (2008) and Mesmer (2008) write about how readability formulas can artificially inflate the difficulty of expository texts due to the texts' repetition of infrequently occurring words. Readability formulas count each infrequent word, whether or not it is a repeated elsewhere in the passage, as an occurrence of a "hard" word. Thus, in the Table 2 expository text example, the word *ivory* is counted as a difficult word each time it occurs, despite the fact that the repetition of the word actually provides the student with support and practice. This artifact of the formulas especially should cause teachers to review expository texts carefully themselves before completely trusting the difficulty-level estimates delivered by the formulas.

A great deal more research needs to be conducted in understanding exactly how genre operates within text complexity models, and this is true of models for stretching or challenging students (Mesmer et al., 2012). Genre may possibly be best represented by multivariate approaches that characterize the many text features that represent the label. In addition, the text features that present challenge in each genre may differentially apply to various outcomes. For instance, prior knowledge may operate more in the expository format than the narrative format. There is clearly a second generation of research in works that will move the elementary text diet beyond simply including various genres to challenging students appropriately in those genres.

Text Cohesion

Traditionally, the estimation of text difficulty has rarely gone beyond evaluating the difficulty of individual words and individual sentences. The aforementioned readability formulas, including the second-generation formulas, theoretically treat each word and sentence separately. The formulas give no consideration to the ways that the words and sentences in a text relate to each other. Their frequencies and lengths are only joined when entered into the equations. However, recent work using a tool called Coh-Metrix is beginning to change this approach by focusing attention on text cohesion (Graesser, McNamara, & Kulikowich, 2011).

Text cohesion is the degree to which the words and ideas are represented both within and across sentences (Givón, 1995). When a text is cohesive, a thread that runs through it that allows the reader to construct a connected gist of the main ideas. Texts with *coherence marking* have ideas repeated and introduced at a pace that optimally mixes new and previously stated information. The examples in Table 3 illustrate elementary-level texts with different levels of cohesion. Note that in the most cohesive texts, words and phrases are repeated at a rapid clip so that these words connect sentences. In the least cohesive example, there are almost no repetitions

of words, and in the medium-level text, there are repetitions but their spacing is across paragraphs more than sentences.

Text cohesion is actually created by a number of different text features. Graesser et al. (2011) have calculated treat cohesion measuring many different elements of text. In a recent piece Graesser et al. (2011) collect text information to address a number of indicators of cohesion including, narrativity, syntactic simplicity, word concreteness, causal cohesion, and referential cohesion. In a series of line graphs they show how each of these indicators of text cohesion look across the grades in the texts of language arts, social science, and science. The findings suggest, among other things, that narrativity is highest in the language arts texts of the earliest grades and that referential cohesion is highest in science texts. As one might expect, syntactic simplicity is highest in the earliest texts, particularly in the science genre.

The introduction of cohesion into the estimation of text difficulty contributes greatly to the theoretical foundation upon which a paradigm of challenge might be based. However, current tools for measuring cohesion are quite complicated and much more needs to be explained as to how cohesion can be pragmatically applied in classrooms. There is not a great deal of information about how differences in cohesion marking impact elementary grade students. What has been established is that text cohesion interacts with prior knowledge and student ability (Graesser et al., 2011).

Programmatically Addressing Challenge

A paradigm for understanding how to stretch students in text must move beyond an isolated, drive-by approach to a more consistent, programmatic one. Stretching students cannot and should not be a Friday afternoon read aloud and discussion. It must be infused into the text choices made over weeks, months, and years. Certainly the arguments put forth for challenge in

the CCSS/ELA document suggest that it is the accumulated effects of text that resulted in lower ACT scores or grades in college (ACT, 2006). So then must the approach to stretching students in text also be longitudinal, across days, weeks, months, and years.

In presenting a framework for selecting texts for use in the early grades, Mesmer et al. (2012) have proposed four elements to examine: (a) content (e.g., words, concepts, sentences, ideas, genre), (b) sequence in which the content is presented, (c) pace of presentation, and (d) repetition of content. As researchers develop a theory of text complexity that contributes to the important notion of stretching students, each of these elements must be addressed. How length, difficulty, genre, cohesion, and text levels are balanced and introduced across a unit of study or a developmental period will support or inhibit fruitful "stretching." It will be focused and consistent efforts at presenting students with challenging texts that stretch their capacity and will ultimately have the kinds of effects intended by the CCSS/ELA writers.

Summary

The CCSS/ELA Standard 10 has introduced a major shift in reader-text matching paradigms that promises to balance the intense focus on the avoidance of frustration with the importance of challenge. Nonetheless this introduction raises some very important issues. As Shanahan (2011) expressed, "We have tended to overgeneralize from younger readers (for whom easier text allows a more systematic focus on decoding) to older readers (who may do better with more intellectually challenging texts). Now, I fear that the Common Core is over-generalizing in the other direction. Harder beginning reading books may stop many young readers in their tracks." (p. 21). I have this same fear, especially in light of the fact that the rationale for increasing text difficulty is based largely on studies of secondary students. Patterns in the data

that are frequently cited to support claims of textbook simplification, when carefully examined, do not actually hold true for elementary students (Chall, 1977; Hayes, Wolfer, & Wolfe 1996).

In the past, the education community's reactions to inappropriate text-related standards have been extreme. Either teachers (or more likely district supervisors) knuckle down and insist that every student in a given grade read texts of a certain level, or teachers abandon ship altogether and default to reading aloud anything that might be considered challenging. While the reader-text matching standards of Betts (1946) should indeed be questioned, I caution educators to remember that stretch text should not be frustrational. Stretch, whatever research decides it may be, is optimal challenge, not heart-wrenching exasperation. Shanahan (2011) notes the opposite response to challenge that might occur, "When the books get hard, the usual responses have been to move kids to easier books, to stop using textbooks, or to read the texts to the students" (p. 20). How very ironic it would be it the text standards designed to challenge students in actuality water down their exposure to challenging texts. I caution educators to resist what I call the "read aloud solution," whereby all challenging texts are read aloud. Instead, a blend of scaffolded challenge reading with some read aloud should characterize stretching students in the elementary school.

To be clear, the research that would support a paradigm of challenge in the elementary grades is simply not robust. Not enough empirical data exists to suggest exactly how students should be stretched in text. In this chapter, I identify text and other factors that may be considered in future research. At a very basic level, however, teachers must know the reading levels of their students and estimates of the difficulty of the texts that they wish to use. While a basic tenant of reader-text matching, frequently this obvious approach is overlooked. Additional factors that may impact students' abilities to be stretched include text levels, text length, genre,

and cohesion. All of these are malleable factors that can be manipulated and designed into text. In addition to giving focus to individual features or factors, research must also give attention to the development of text programs that challenge students across weeks, months, years, and particular developmental periods. Ultimately it is the texts to which students are exposed over a long period of time that influence their abilities to tackle ever-more demanding text.

In the title of a recent *Reading Today* article, Shanahan (2011) asks: Are we going to lower the fences or teach kids to climb? The metaphor is important. For too long, we have been overly concerned about the height of the fences and not about teaching kids to climb. To use another metaphor, I think that stretching students in texts might be like adjusting the uneven bars in a gym. When gymnasts are at a certain level in their training, they are expected to mount the bars using a spring board or other device to begin their routines. This means that the bar is typically above their head and several feet ahead of them. They must run and bounce on the spring board and reach for the bar to begin the routine. Sometimes they fall on the mats beneath, but eventually they can complete the routine smoothly. Throughout a meet or workout, you will see coaches raise and lower the bars to accommodate different gymnasts' heights, No one expects the bar to be set the same for a gymnast who is 4'3" as it is for a gymnast who is 4'8". The same is true with stretching students in texts. We do want them strive to reach the bar, but we should set the bar relative to their specific needs. As I have argued elsewhere (Mesmer & Hiebert, 2013), stretching students in text is a dynamic activity, and one that cannot be dictated by static text-difficulty standards. The duty of researchers is to continue to create information to guide and support teachers as they work to develop stronger readers in our elementary schools.

References

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading.* Iowa City, IA: Author.

Beck, I.L., McKeown, M.G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Betts, E.A. (1946). *Foundations of reading instructio*n. New York: American Book Company.

Calfee, R., & Hiebert, E.H. (2011). *Using cohort analyses to examine long-term effects of reading initiatives in California.* Paper presented at the annual meeting of the Society for the Scientific Study of Reading, Montreal, Quebec.

California Department of Education (1996). *Guide to the California Reading Initiative of 1996.* Sacramento, CA: Author.

Clay, M. (1985). *The early detection of reading difficulties.* Portsmouth, NH: Heinemann.

Coleman, D., & Pimentel, S. (2012). *Revised publishers' criteria for the Common Core State Standards in English Language Arts and Literacy, grades 3–12*. Retrieved from the Common Core Standards Initiative at www. corestandards. org/assets/Publishers_Criteria_for_3-12. pdf.

Cunningham, J.W., Spadorcia, S.A., Erickson, K.A., Koppenhaver, D.A., Sturm, J.M., & Yoder, D.E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly, 40*(4), 410–427.

Ehri, L.C., Dreyer, L.G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language minority students who are struggling readers in first grade. *American Educational Research Journal, 44*(2), 414–448.

Ekwall, E., Solis, J.K., & Solis, E. (1973). Investigating informal reading inventory scoring criteria. *Elementary English*, *50*, 271–274, 323.

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education, 28*(5), 491–520.

Givón, T. (1995). *Functionalism and grammar*. Philadelphia: John Benjamins.

Graesser, A.C., McNamara, D.S., & Kulikowich, J.M. (2011). Coh-Metrix providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234.

Halladay, J.L. (2012). Revisiting key assumptions of the reading level framework. *The Reading Teacher*, *66*(1), 53–62.

Harrison, C. (1980). *Readability in the classroom*. Cambridge, UK: Cambridge University Press.

Hatcher, P.J. (2000). Predictors of Reading Recovery book levels. *Journal of Research in Reading, 23*(1), 67–77. doi: 10.1111/1467-9817.00103

Hiebert, E. H. (2008). The word zone fluency curriculum: An alternative approach. In M.R. Kuhn & P. J. Schwanenflugel (Eds.), *Fluency in the classroom* (pp. 154–170). New York, NY: Guilford Press.

Hiebert, E.H., & Mesmer, H.A. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher, 42*(1), 44–51.

Hiebert, E.H., Wilson, K.M., & Trainin, G. (2010). Are students really reading in independent reading contexts? An examination of comprehension-based silent reading rate. In E.H. Hiebert & D.R. Reutzel (Eds.), *Revisiting Silent Reading: New Directions for Teachers and Researchers*. Newark, DE: International Reading Association.

Martin, M.O., Mullis, I.V., & Kennedy, A.M. (2007). *Progress in International Reading Literacy Study (PIRLS, 2006)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Mesmer, H.A.E. (2008). *Tools for matching readers to texts: Research- based practices*. New York, NY: Guilford Press.

Mesmer, H.A. & Hiebert, E.H. (2011). The impact of difficulty and length on third graders' comprehension. Unpublished manuscript.

Mesmer, H.A., Cunningham, J.W., & Hiebert, E.H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly, 47*(3), 235–258.

Morgan, A., Wilcox, B.R., & Eldredge, J.L. (2000). Effect of difficulty levels on second-grade delayed readers using dyad reading. *Journal of Educational Research*, *94*, 113–119.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010a). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects.* Washington, DC: Authors. Retrieved from www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010b). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*, Appendix A. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/Appendix_A.pdf

Pikulski, J.J., & Shanahan, T. (1982). Informal reading inventories: A critical analysis. In J.J. Pikulski & T. Shanahan (Eds.), *Approaches to the informal evaluation of reading (*pp. 94–116). Newark, DE: International Reading Association.

Roskos, K., & Neuman, S.B. (2013). Common Core, commonplaces, and community in teaching reading. *The Reading Teacher, 66*(6), 469–473.

Shanahan, T. (2011). Common Core Standards: Are we going to lower the fences or teach kids to climb? *Reading Today*, *29*(1), 20-21.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Washington, DC: Rand Corporation.

Stahl, S., & Heubach, K. (2005). Fluency-oriented reading instruction. *Journal of Literacy Research, 37,* 25–60.

Stenner, A.J., Burdick, H., Sanford, E.E., & Burdick, D.S. (2007). *The Lexile framework for reading* (Technical Report). Durham, NC: MetaMetrics.

Texas Department of Education (1997). 1997 Proclamation of the State Board of Education Advertising for Bids on Instructional Materials. Austin, TX: Author.

Williamson, G.L. (2006). *Aligning the journey with a destination: A model for K–16 reading standards.* Durham, NC: MetaMetrics, Inc.

Table 1

Common Core State Standards' Revised Guidelines on Ranges of Text Complexity (Nelson, Perfetti, Liben, & Liben, 2012)

Common Core Band	ATOS	Degrees of Reading Power®	Flesch-Kincaid[8]	The Lexile Framework®	Reading Maturity	SourceRater
2nd – 3rd	2.75 – 5.14	42 – 54	1.98 – 5.34	420 – 820	3.53 – 6.13	0.05 – 2.48
4th – 5th	4.97 – 7.03	52 – 60	4.51 – 7.73	740 – 1010	5.42 – 7.92	0.84 – 5.75
6th – 8th	7.00 – 9.98	57 – 67	6.51 – 10.34	925 – 1185	7.04 – 9.57	4.11 – 10.66
9th – 10th	9.67 – 12.01	62 – 72	8.32 – 12.12	1050 – 1335	8.41 – 10.81	9.02 – 13.93
11th – CCR	11.20 – 14.10	67 – 74	10.34 – 14.2	1185 – 1385	9.57 – 12.00	12.30 – 14.50

Table 2

Examples of Texts

Expository[1]	Narrative Text[2]
DOZENS OF AFRICAN ELEPHANTS **SLAUGHTERED.** That headline has become all too common. Last month, *poachers* killed at least 86 elephants in Chad and in Cameroon. Both countries are in a *region* of Africa that has lost more than 60% of its elephants to illegal hunters in the past decade, according to a recent study from the Wildlife Conservation Society. In 2012 alone, experts say, 30,000 elephants were killed in countries across Africa. "We're seeing the highest levels of *poaching* since our record-keeping began," Crawford Allan, of the World Wildlife Fund, told TFK. Why are so many elephants being killed?	My four friends and I had come across a loose *floorboard* at the back of the classroom, and when we *pried* it up with the *blade* of a *pocketknife,* we discovered a big hollow space *underneath.* This, we decided, would be our own secret hiding place for sweets and other small treasures such as *conkers,* and monkey-*nuts,* and birds' eggs. Every afternoon, when the last lesson was over, the five of us would wait until the classroom had emptied, then we would lift up the *floorboard* and *examine* our secret *hoard,* perhaps adding to it or taking something away.

[1] The Price of Ivory. *Time for Kids*

[2] From Roald Dahl (xxxx). *Boy*. City, State: Publisher.

Table 3

Examples of Expository Texts at Three Levels of Referential Cohesion from the Common Core State Standards Exemplars in Appendix B

LOW	**MEDIUM**	**HIGH**
This island is covered with snow. No trees grow. Nothing has green leaves. The land is white as far as you can see. Then something small and round and black pokes up out of the snow. A black nose sniffs the air. Then a smooth white head appears. A mother polar bear heaves herself out of her den. A cub scrambles after her. When the cub was born four months ago, he was no bigger than a guinea pig. Blind and helpless, he snuggled in his mothers fur. He drank her milk and grew, safe from the long Arctic winter. (From *Where Polar Bears Live? Thomson, 2010)*	**Horses** move in four natural ways, called gaits or paces. They **walk**, **trot**, canter, and gallop. The **walk** is the slowest **gait** and the **gallop** is the fastest. When a **horse walks**, each hoof leaves the ground at a different time. It moves one **hind** leg first, and then the front leg on the same side; then the other **hind** leg and the other front leg. When a **horse walks**, its body swings gently with each stride. When a **horse trots**, its **legs** move in pairs, left front leg with right **hind** leg, and right front leg with left **hind** leg. (From *Horses*, Simon, 2006)	Most **plants** make **seeds**. A **seed** contains the beginning of a new **plant**. **Seeds** are different shapes, sizes, and colors. All **seeds** grow into the same kind of **plant** that made them. Many plants grow **flowers**. **Flowers** are where most **seeds** begin. A **flower** is made up of many parts. At the bottom of the pistil are tiny egg cells called ovules. In the center of the **flower** is the **pistil**. The sticky part at the top of the **pistil** is the **stigma**. The parts of the **flower** around the pistil are the stamens. Stamens make yellow powder called pollen. (*From Seed to Plant*, Gibbons, 1991)