Coh-Metrix: Providing Multilevel Analyses of Text Characteristics

Arthur C. Graesser, Danielle S. McNamara, and Jonna M. Kulikowich

Computer analyses of text characteristics are often used by reading teachers, researchers, and policy makers when selecting texts for students. The authors of this article identify components of language, discourse, and cognition that underlie traditional automated metrics of text difficulty and their new Coh-Metrix system. Coh-Metrix analyzes texts on multiple measures of language and discourse that are aligned with multilevel theoretical frameworks of comprehension. The authors discuss five major factors that account for most of the variance in texts across grade levels and text categories: word concreteness, syntactic simplicity, referential cohesion, causal cohesion, and narrativity. They consider the importance of both quantitative and qualitative characteristics of texts for assigning the right text to the right student at the right time.

Keywords: cognition; comprehension; computer applications; content analysis; discourse processes; language processes; technology; textual analysis

Automated analyses of the language and discourse characteristics of texts have enormous practical value in education, in addition to advancing scientific theories of reading and comprehension. One use of automated text analysis is the scaling and selection of texts for students who are learning to read and reading to learn. The selection of texts is indeed one of the central concerns of teachers, principals, and superintendents as they plan for meeting the standards declared by No Child Left Behind (2001), the Race to the Top competition (see http://racetotop.com/), the Common Core State Standards (Common Core State Standards Initiative, 2010), and other high-stakes assessments. For the purpose of developing reading skills and strategies, the nature of the texts that students read is presumably just as important as the interventions and curriculum. For example, K–12 texts have progressively decreased in difficulty over the years (Adams, 1990; Common Core State Standards Initiative, 2010), so many researchers and practitioners have advocated finding ways to expose students to more difficult texts. Others believe that many text characteristics need to be considered in addition to the traditional measures of ease or difficulty, and that it is important to analyze texts on a profile of characteristics when selecting those for students to read. Moreover, these characteristics should include both qualitative attributes and quantitative scales.

Scientific investigations of reading also benefit from automated measures of text characteristics. Automated measures allow researchers to sample and manipulate texts systematically to either target or control particular reading components in their investigations of comprehension. For example, if a researcher wants to investigate the impact of text cohesion on comprehension, it is important to make sure that the texts in the various conditions vary on cohesion rather than some extraneous text characteristics that are irrelevant to cohesion (McNamara, Louwerse, McCarthy, & Graesser, 2010). An automated analysis is unquestionably more reliable and objective than approaches that involve humans annotating and rating texts by hand.

The purpose of this article is to review the automated text analysis systems that are currently in use and to identify the characteristics of language, discourse, and cognition that they appear to tap. The review includes traditional metrics of text difficulty (versus ease) and also a system that we have developed. Our system, called Coh-Metrix, analyzes texts on multiple levels of language and discourse that range from words to discourse genres (Graesser & McNamara, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Graesser, & Louwerse, in press; McNamara et al., 2010). We describe levels of language-discourse that have been proposed in multilevel theoretical frameworks and explain how the measures of Coh-Metrix are aligned with these levels. We illustrate the value of Coh-Metrix in helping us understand variations in texts in a large corpus that covers different grade levels and genres (i.e., language arts, science, social studies).

Two important caveats must be raised before describing the computer analyses of text characteristics. First, computers obviously cannot identify and scale texts on all levels of language, discourse, and meaning. Available computer systems cannot fully comprehend the deep metaphors, literary devices, and historical contexts of Shakespeare's plays, for example. Some characteristics of texts require humans to provide informed, deep, critical analyses. Second, successful text comprehension involves much more than an analysis of text characteristics alone because prior knowledge, inference mechanisms, and skills of readers are also critically important (Graesser, Singer, & Trabasso, 1994; McNamara

et al., in press; Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992; Snow, 2002). Nevertheless, a systematic understanding of text characteristics at multiple levels is needed to track the interactive comprehension processes that involve the text, reader, task, and sociocultural context (Graesser & McNamara, 2011).

Traditional Computer Metrics of Text Ease/Difficulty

The traditional approach to scaling texts is to have a single metric of text ease or difficulty. This is the approach taken by the three most popular metrics, namely Flesch-Kincaid Grade Level or Reading Ease (Klare, 1974–1975), Degrees of Reading Power (DRP; Koslin, Zeno, & Koslin, 1987), and Lexile scores (Stenner, 2006). These three metrics of text difficulty are highly correlated according to statistical analyses we have conducted (r = 0.89 to .94).

The simplicity of a single dimension of text difficulty can be useful when assigning texts for students to read. A single dimension provides a common currency of difficulty for different texts in different categories, which makes it easier for reading teachers when they strive to select texts at the appropriate level of challenge. A teacher may seek a text that is at just the right level, challenge a student with a more difficult text, or provide a text that is easy enough for the student to readily understand. A unidimensional metric provides a simple solution to this task because the dimensions are generally aligned with a common metric, namely, grade level.

The Flesch-Kincaid metrics (Grade Level and Reading Ease) are based on the length of words and length of sentences. For example, Flesch-Kincaid Grade Level is computed as 11.8 * Syllables + .39 * Words – 15.59. Word length is measured as the mean number of syllables per word (which is highly correlated with number of letters); sentence length is the mean number of words per sentence. Flesch-Kincaid Grade Level is a robust predictor of the amount of time it takes to read a passage, a result that offers an impressive validation of the metric. There are a number of theoretical explanations of the validity of this metric, but two obvious ones appeal to prior knowledge and working memory. Infrequently used words in a language tend to be longer, according to Zipf's law (Zipf, 1949), so the word length variable may be a proxy for prior knowledge about the topic. Long sentences tend to place more demands on working memory and to have more complex syntax, so sentence length may be a proxy for these factors. Flesch-Kincaid metrics are easy to compute, in addition to being robust predictors of reading time and having this cognitive theoretical grounding. However, other metrics have been pursued by researchers to capture text ease/difficulty more precisely.

DRP and Lexile scores relate characteristics of the texts to performance of readers in a cloze task. In the cloze task, the reader receives the text with words left blank and is asked to fill in the words by generating them or by selecting a word from a set of options. A text is at the reader's level of proficiency if the reader can perform the cloze task at a corresponding threshold of performance (e.g., 75%). A text is defined as easy for a population of readers if performance exceeds 75% and difficult to the extent it is lower than 75%. Grade level can be calibrated for a text by identifying the age group that converges on the 75% level of performance. DRP and Lexile scores are currently the most popular metrics of text difficulty in the United States.

All of these traditional measures of text difficulty are based on the constraints of words, sentences, and local text. While their simplicity and alignment with grade level are appealing, they do not tap the more global levels of discourse meaning, cohesion, and differences in text genre (e.g., narrative versus informational texts). Moreover, a unidimensional metric of text difficulty is not helpful for improving student reading when specific guidance is needed for diagnosing a student's particular deficit and planning remediation (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; McNamara et al., in press; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). The scaling and selection of texts would therefore benefit from an analysis of multiple levels of language and discourse.

Coh-Metrix Analyses on Multiple Levels of Language and Discourse

Coh-Metrix was developed to analyze texts on multiple characteristics and levels of language-discourse. The original inspiration in developing Coh-Metrix was to have an automated metric of text cohesion (hence the label *Coh*-Metrix). Cohesion was of particular interest because comprehension is influenced by some intriguing interactions between text cohesion and the readers' prior knowledge about the topic and their general comprehension skill (McNamara & Kintsch, 1996; Ozuru, Dempsey, & McNamara, 2009). Cohesion helps most readers comprehend most texts, but under some conditions, texts with lower cohesion stimulate more knowledgeable readers to generate more inferences and meaningful explanations. To investigate this interaction systematically, we needed objective metrics of cohesion. Although cohesion was at the heart of the original inspiration for Coh-Metrix, we recognized very early in its development a need to analyze texts at multiple language-discourse levels.

The measures of Coh-Metrix are aligned with various language-discourse levels proposed in multilevel theoretical frameworks (Graesser & McNamara, 2011; Kintsch, 1998; Snow, 2002). These theoretical frameworks identify the representations, structures, strategies, and processes at different levels of language and discourse. Five levels have frequently been proposed in these frameworks: (1) *words*, (2) *syntax*, (3) the explicit *textbase*, (4) the *situation model* (sometimes called the mental model), and (5) the discourse *genre and rhetorical structure* (the type of discourse and its composition). A sixth level of *pragmatic communication* between speaker and listener, or writer and reader, is often also included, but this level is not tightly connected with text characteristics and will not be addressed in this article. The five levels are elaborated later and in Graesser and McNamara (2011).

The focus of Coh-Metrix is on linguistic features that have been closely associated with deeper levels of comprehension than those aligned with *basic reading* components, such as the alphabet, letter-sound correspondences, lexical decoding, morphological awareness, and reading fluency (words read per minute). These basic levels are important for beginning readers and have predictable consequences for subsequent reading development at deeper levels (Adams, 1990; Pearson & Hiebert, 2010; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Stanovich, 1986). However, there also is a need for targeting *secondary reading literacy* where

students benefit more from interventions, content, and texts pitched at deeper reading components and media of 21st-century literacy (Biancarosa & Snow, 2004; Meyer et al., 2010; Moje, 2008; Resnick, 2010; Wiley, Goldman, Graesser, Sanchez, Ash, & Hemmerich, 2009; Williams, Stafford, Lauer, Hall, & Pollini, 2009). The deeper levels start with vocabulary and move beyond the words into sentence interpretations, inferences, and more global discourse structures.

This section describes the five levels of the multilevel framework that Coh-Metrix attempts to accommodate. The Coh-Metrix team has collected and evaluated hundreds of measures of text during the last decade. The measures scale texts on characteristics related to words, sentences, and connections between sentences. The large suite of measures ends up nicely funneling into factors that are aligned with Levels 1–5 of the multilevel theoretical framework (Graesser & McNamara, 2011). It is beyond the scope of this article to specify precisely how each measure is computed. Such information is available from the Help system of Coh-Metrix (http://cohmetrix.memphis.edu) and the referenced Coh-Metrix publications. However, we do briefly discuss the measures analyzed in this study and some of the major computational modules.

Words

Vocabulary knowledge has a substantial impact on reading time and comprehension (Perfetti, 2007; Rayner et al., 2001; Stanovich, 1986). School texts contain increasingly more complex and unfamiliar words starting in the late elementary years (Adams, 1990; Beck, McKeown, & Kucan, 2002). Readers with high-quality lexical representations are those with rich associations among the phonology, orthography, morphology and syllable structure of words (Perfetti, 2007). Beginning readers may benefit from texts with easy vocabulary tied to everyday oral language (Hiebert & Fisher, 2007), whereas students in secondary school may benefit from texts with greater lexical diversity and repeated exposure to these words in different naturalistic contexts (Beck et al., 2002). Therefore, it is important to analyze words on multiple characteristics that have relevance to reading development and construction of meaning. Coh-Metrix has an abundance of word measures, including those described in this section.

Parts of speech. Each word is assigned a syntactic part-of-speech category. These syntactic categories are segregated into content words (such as nouns, main verbs, adjectives, adverbs) and function words (e.g., prepositions, determiners, pronouns). Many words can be assigned to multiple syntactic categories. For example, the word *bank* can be a noun (*river bank*), an adjective (*bank shot*), or a main verb (*don't bank on it*). Coh-Metrix assigns only one part-of-speech category to each word on the basis of its syntactic context. That is, a word in isolation may have potential as many alternative parts of speech, but the surrounding syntactic context of the word narrows down the alternatives to the most likely part-of-speech category. Coh-Metrix computes the *relative frequency* of each word category (both syntactic and semantic categories, described later) by counting the number of instances of the category per 1,000 words of text.

Some syntactic categories have repercussions at deeper levels of comprehension. This is important to acknowledge because

word characteristics contribute to many levels in the multilevel framework and it is an empirical question as to how words affect the various levels. Comprehension is a highly interactive mechanism, not a set of measured variables that march in a strict hierarchy. Consider pronouns, for example. Pronouns are important for constructing cohesion in the textbase and situation model because it can be difficult to link a pronoun to another discourse unit or to the topic being discussed. This difficulty may threaten textual cohesion in ways that make comprehension more difficult. Pronouns are segregated further into first-person pronouns (*I, me, we, us*) and third-person pronouns (*he, she, it, that, those*) because this distinction has repercussions at other levels of meaning (Pennebaker, Booth, & Francis, 2007). Pronouns are more prevalent in oral discourse, which is generally easier than print (Biber, 1988; Tannen, 1982). Similarly, verbs are more prevalent in narrative discourse than in informational texts; so once again, syntactic categories may be very diagnostic of high levels of meaning. Connectives also constitute an important category because they are important in establishing cohesion of the situation model. There are different categories of connectives, including causal (*because, so*), temporal (*then, after, during*), logical (*therefore, if, or*), additive (*also, moreover*), and adversative (*on the other hand, however*), which Coh-Metrix classifies on the basis of prior research (Halliday & Hasan, 1976; Louwerse, 2001). These classifications consider semantic features in addition to the syntactic classes.

Word frequency. Words that occur with a higher frequency in the English language are more familiar to the reader, are processed more quickly, and are linked to richer bodies of world knowledge (Beck et al., 2002; Haberlandt & Graesser, 1985; Perfetti, 2007). Word frequency in Coh-Metrix is computed from multiple corpora of texts that we have documented in Coh-Metrix publications. The logarithm of word frequency is computed because reading times are linearly related to the logarithm of word frequency, not to raw word frequencies (Haberlandt & Graesser, 1985; Just & Carpenter, 1987). Content words normally are considered in these computations, rather than the highly frequent function words. Moreover, it is the low-frequency words in a sentence that are an important limiting factor in comprehending sentences and text. One rare word can make the entire sentence difficult to comprehend.

Psychological ratings. Coh-Metrix measures words on characteristics in an established psycholinguistic database (Coltheart, 1981), a collection of ratings of thousands of words along several psychological dimensions: age of acquisition, meaningfulness, concreteness, imagability, and familiarity. The age-of-acquisition measure specifies the age at which a word first appears in a child's vocabulary, whereas the other measures are based on adults' ratings of content words on 7-point scales, with higher scores reflecting easier processing.

Semantic content. The semantic content of words is analyzed by semantic categories provided by WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), a large lexicon with words annotated by experts on linguistic and psychological features. Coh-Metrix tracks the content words in a text on many of these lexical features. For example, the nouns are classified as *human,*

animate, inanimate concrete, or *abstract*—distinctions that are relevant to English syntax (Jurafsky & Martin, 2008). Three features are particularly relevant to the measures reported in the present study. One of these is *polysemy*, which refers to the number of senses (core meanings) of a word. For example, the noun *bug* has at least two senses, one referring to an insect and the other referring to a recording device planted by a detective. Other important lexical features provided by WordNet classify the verbs as *causal* or *intentional*. The distinction between causality and intentionality has relevance to the nature of knowledge in situation models (Zwaan & Radvansky, 1998). Intentional verbs signal actions that are voluntarily enacted by animate agents and that are motivated by plans in pursuit of goals (such as buying groceries, telling a child to behave, or watching a television program). Causal verbs reflect events in the material world or psychological world (such as an earthquake erupting, or a person discovering a solution) that may or may not be driven by goals. The distinctions between these words have consequences at the situation model level (Graesser et al., 1994, 2004; Zwaan & Radvansky, 1998).

Syntax

Theories of syntax assign words to part-of-speech categories (e.g., nouns, verbs, adjectives, conjunctions), group words into phrases (noun phrases, verb phrases, prepositional phrases, clauses), and assign syntactic tree structures to sentences (Jurafsky & Martin, 2008). Some sentences are short and have a simple syntax with (for example) an actor-action-object template, few if any embedded clauses, and active rather than passive voice. This is the syntax of oral discourse (Tannen, 1982). However, sentences in print often have a complex, embedded syntax that places demands on the reader's working memory. For example, the following question is from the 1999 American Community Survey: "At any time during the last 12 months, were you or any member of your household enrolled in or receiving benefits from free or reduced-price meals at school through the Federal School Lunch program or the Federal School Breakfast program?" This sentence has several forms of complex syntax. First, it contains dense noun phrases with many modifiers. Second, it places a high number of words (i.e., 16) before the main verb (i.e., "enrolled") of the main clause, thus taxing the reader's working memory (Graesser, Cai, Louwerse, & Daniel, 2006). Third, it requires the reader to keep track of many combinations of meaning with logic-based words such as "and," "or," and "not."

Coh-Metrix can scale texts on the syntactic considerations just described. Sentence syntax is easier when there are shorter sentences, fewer words per noun phase, fewer words before the main verb of the main clause, and fewer logic-based words. Coh-Metrix computes two additional syntactic measures: frequency of passive voice, which is more difficult to process than active voice (Just & Carpenter, 1987); and syntactic similarity, or similarity in syntactic structure between pairs of sentences in a paragraph, which facilitates reading speed and comprehension.

Textbase

The textbase consists of the explicit ideas in the text: the meaning rather than the surface code of wording and syntax (Kintsch, 1998; van Dijk & Kintsch, 1983). *Co-reference* is an important linguistic method of connecting propositions, clauses, and sentences in the textbase (Halliday & Hasan, 1976; McNamara & Kintsch, 1996). Referential cohesion occurs when a noun, pronoun, or noun phrase refers to another constituent in the text. For example, in the sentence "When the intestines absorb the nutrients, the absorption is facilitated by some forms of bacteria," the word "absorption" in the second clause refers to the event associated with the verb "absorb" in the first clause. A referential cohesion gap occurs when the words in a sentence or clause do not connect to other sentences in the text. Such cohesion gaps at the textbase level increase reading time and may disrupt comprehension (McNamara & Kintsch, 1996; O'Brien, Rizzella, Albrecht, & Halleran, 1998).

Co-reference. Coh-Metrix tracks different types of word co-reference: *content word overlap, noun overlap, argument overlap*, and *stem overlap*. Content word overlap is the proportion of content words that are the same between pairs of sentences. Noun overlap is the proportion of sentence pairs that share one or more common nouns; argument overlap is the proportion of sentence pairs that share common nouns or pronouns (e.g., *table/table, he/he,* or *table/tables*). Stem overlap is the proportion of sentence pairs in which a noun in one sentence has a semantic unit in common with any word in any grammatical category in the other sentence (e.g., the noun *photograph* and the verb *photographed*). Some metrics consider only adjacent sentences in the text, whereas others compute co-reference of all possible pairs of sentences in a paragraph.

Lexical diversity. Lexical diversity is related to cohesion because a higher number of different words in a text means that new words need to be integrated into the discourse context. The most well-known lexical diversity index is *type-token ratio*, but a number of analogous measures are included in Coh-Metrix that control for text length (McCarthy & Jarvis, 2007). Type-token ratio is the number of unique words in a text (i.e., types) divided by the overall number of words (i.e., tokens).

Latent semantic analysis. Coh-Metrix measures conceptual overlap between sentences by a statistical model of word meaning called Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007; Millis, Magliano, & Todaro, 2006). LSA is an important method of computing similarity because it considers implicit knowledge in addition to the explicit words. LSA is a mathematical, statistical technique for representing word and world knowledge, based on a large corpus of texts. The central intuition is that the meaning of a word is captured by the company of other words that surround it in naturalistic documents. Two words have similarity in meaning to the extent that they share similar surrounding words. For example, the word *hammer* is highly associated with words of the same functional context, such as *screwdriver*, *pound*, and *construction*. Meaning in this sense is very different from the meaning found by looking up words in a dictionary or thesaurus. LSA uses a statistical technique called singular value decomposition to condense a very large corpus of texts to 100–500 statistical dimensions (Landauer et al., 2007). The conceptual similarity between any two text excerpts (e.g., word, clause, sentence, text) is computed by a formula that considers the values and weighted dimensions of the

words in the two text excerpts. LSA-based cohesion is measured in several ways in Coh-Metrix, but primarily in terms of the similarity between adjacent sentences and all possible pairs of sentences in a paragraph. Another LSA measure computes how much given information (as opposed to new information) exists in each sentence in a text, as compared with the content of prior text information, e.g., $G/(N + G)$. Each of these measures varies from 0 (low cohesion) to 1 (high cohesion).

Situation Model

The situation model is the subject matter content or the narrative world that the text is describing. In narrative text, this includes characters, objects, spatial setting, actions, events, processes, plans, thoughts and emotions of characters, and other details of the story. In informational text, the situation model is the substantive subject matter being described. For example, a text on the circulatory system in biology would include (a) causal networks of the events, processes, and enabling states that transpire over time, (b) properties of components, (c) the spatial composition of the anatomy, and (d) goal-oriented actions of doctors, patients, and family who try to improve the functioning of someone's circulatory system. The situation model includes inferences that are activated by the explicit text and encoded in the meaning representation (Goldman, Graesser, & van den Broek, 1999; Graesser et al., 1994; Kintsch, 1998; McNamara & Kintsch, 1996; Wiley et al., 2009; Zwaan & Radvansky, 1998).

Zwaan and Radvansky (1998) proposed five dimensions of the situation model that apply to narrative text: causation, intentionality (goals), time, space, and protagonists. These dimensions are applicable to informational texts also, as illustrated above for a text on the circulatory system. A break in text cohesion occurs when there is a discontinuity on one or more of these situation model dimensions. Such cohesion breaks result in an increase in reading time and generation of inferences (O'Brien et al., 1998; Rapp et al., 2007; Zwaan & Radvansky, 1998). Whenever such discontinuities occur, it is important to have connectives, transitional phrases, adverbs, or other signaling devices that convey to the readers that there is a discontinuity. Cohesion is facilitated by these signaling devices, which clarify and stitch together the actions, goals, events, and states conveyed in the text (Louwerse, 2001). However, if the reader has enough background knowledge about the subject matter, the lack of cohesion devices may not disrupt comprehension and can sometimes improve comprehension by stimulating inferences (McNamara & Kintsch, 1996; O'Reilly & McNamara, 2007; Ozuru et al., 2009).

A break in cohesion occurs when there is a discontinuity on one or more of these situation model dimensions. Whenever such discontinuities occur, it is important to have connectives, transitional phrases, adverbs, or other signaling devices that convey to the readers that there is a discontinuity; we refer to these forms of signaling as *particles*. Different sets of particles are associated with causal (e.g., *because, enable*), intentional (*in order to, so that*), and temporal (*before, later*) cohesion, although some particles may be applicable to more than one type of cohesion. Cohesion is facilitated by particles that clarify and stitch together the actions, goals, events, and states conveyed in the text. For causal, intentional, and temporal cohesion, Coh-Metrix computes the ratio of cohesion particles to the relative frequency of relevant referential content (i.e., main verbs that signal state changes, events, actions, and processes). The ratio metric is essentially a conditionalized frequency of cohesion particles that stitch together the clauses with referential content.

Time is represented through morphemes associated with the main verb or helping verb that signal tense (past, present, future) and aspect (in progress, completed). Some temporal indices of Coh-Metrix consist of a repetition score that tracks the consistency of tense and aspect across a passage of text. The repetition scores decrease as shifts in tense and aspect are encountered. Where such temporal shifts occur, readers would encounter difficulties without explicit shift-signaling particles, such as the temporal adverbial (*later on*), temporal connective (*before*), or prepositional phrases with temporal nouns (*on the previous day*). A low particle-to-shift ratio is a symptom of problematic temporal cohesion.

Genre and Rhetorical Structure

Genre refers to the category of text, such as whether the text is narration, exposition, persuasion, or description (Biber, 1988; Pentimonti, Zucker, Justice, & Kaderavek, 2010). These major genre categories can be broken down into subcategories within a taxonomy. For example, subtypes of narrative text range from simple folktales to novels; subtypes of persuasion are newspaper editorials and religious sermons. Distinctive characteristics of language signal text genre (Biber, 1988). Narrative text is substantially easier to read, comprehend, and recall than informational text (Graesser & McNamara, 2011; Haberlandt & Graesser, 1985) even when the familiarity of the topics and vocabulary are controlled. The truth of the information and quality of the sources are expected to be carefully scrutinized in expository text (Bråten, Strømsø, & Britt, 2009; Wiley et al., 2009), whereas fictitious stories call for willing suspension of disbelief. Training readers to recognize genre and global text structures helps them to improve comprehension (Meyer et al., 2010). In addition to global genre, a text has a rhetorical composition that provides a more differentiated functional organization of the message. Sections, paragraphs, and sentences have discourse functions that are coherently linked to the macro-organization of the text.

Coh-Metrix analyzes the extent to which a text is classified as narrative as opposed to informational. Texts cannot always be crisply categorized as one or the other; some have elements of both. As discussed in the next section, there is a single, continuous, quantitative dimension that varies from informational to narrative, called *narrativity*. The features that contribute to the narrativity dimension include characteristics of words, sentences, and connections between sentences. These features are discussed in the next section.

Analysis of Texts in a Large Corpus With Coh-Metrix

Many questions arise concerning the added value of Coh-Metrix when compared with the traditional metrics of text difficulty. One is the question whether computer analyses will go the distance in analyzing the qualitative characteristics of text. We understand these concerns, and our modest goal is to offer some solutions that may prove helpful. The purpose of this section is to report on an analysis that applies Coh-Metrix to a large corpus

of texts. The analysis provides some perspective on the value of the system for uncovering constraints in texts for different grade levels and genres. We believe it will contribute to the collective vision of how to advance literacy and text selection in many relevant communities.

Corpus Analyses

We conducted analyses on a large corpus of texts that were scaled on Coh-Metrix measures. There were 37,520 texts in a corpus provided by TASA (Touchstone Applied Science Associates). The texts had a mean length of 288.6 words (*SD* = 25.4), approximately the length of a paragraph. We had a number of reasons for selecting the TASA corpus other than the fact that the corpus was large and readily available. One important reason was that this corpus is representative of the texts that a typical senior in high school would have encountered from kindergarten through 12th grade. A second important reason was that TASA researchers provide measures that directly or indirectly reflect text ease/difficulty. Each text has an associated Degrees of Reading Power (DRP) score of text difficulty (Koslin et al., 1987), with an approximate grade level associated with these scores specified in McNamara et al. (in press). The TASA researchers assigned each text to a category: Most of the text genres were classified in language arts (*n* = 15,991), science (*n* = 5,349), or social studies/history (*n* = 10,438); other categories were business, health, home economics, and industrial arts. Science and other informational texts cover topics less familiar to readers than texts in language arts, which are predominately narrative. The TASA measures of DRP (approximate grade level) and genre provide an objective gold standard for analyzing text characteristics.

In Table 1 we list 53 Coh-Metrix measures that were included in the analyses. The measures are grouped into those related most to words, sentences, and connections between sentences. All of these measures were defined in the previous section, but additional details can be accessed in our publications cited in this article and in the help facility in the Coh-Metrix website (http://cohmetrix.memphis.edu). Coh-Metrix provides dozens of additional measures of text characteristics, but we excluded measures that did not predict variations in text characteristics and measures that were highly correlated ($|r| > .90$) with other similar measures. Our selection of measures was also intended to ensure sufficient coverage of the different language and discourse levels.

A principal component analysis (PCA) was performed to reduce our large database with 53 measures to a small number of functional dimensions. PCA is routinely used in such data reduction efforts. The Coh-Metrix measures in Table 1 converged on a PCA with eight principal components. For interpretation of results, our PCA followed the guidelines outlined by Der and Everitt (2002). That is, the PCA assumed an orthogonal loading of components with Varimax rotation. The Varimax solution yielded a very small percentage of cross-loadings across components that exceeded |.30| and minimal increments in the solution from oblique solutions after rotation.

The eight principal components (PCs) accounted for 67.3% of the variability among texts. The incremental variance explained by PCs 1–8 were 18.5%, 14.1%, 9.5%, 6.3%, 5.5%, 5.4%, 4.1%, and 4.0%, respectively, for the total sums of squares loadings. Descriptions of these PCs are listed in the note to Table 1 in

the order of variance explained: narrativity, referential cohesion, syntactic simplicity, word concreteness, causal cohesion, verb cohesion, logical cohesion, and temporal cohesion.

The PCs map quite favorably onto the levels of the multilevel theoretical framework articulated by Graesser and McNamara (2011). The five levels in this framework, starting from global spans of text to words are

- Genre, corresponding to narrativity (PC1)
- Situation model, corresponding to causal (PC5), verb (PC6), logical (PC7), and temporal (PC8) cohesion
- Textbase, corresponding to referential cohesion (PC2)
- Syntax, corresponding to syntactic simplicity (PC3)
- Words, corresponding to word concreteness (PC4)

One way of validating the PCA and Coh-Metrics is to examine how the measures map onto grade level and genre (see Deane, Sheehan, Sabatini, Futagi, & Kostin, 2006). Ease scores should decrease over grade level and also should systematically differ for narrative (language arts) versus informational texts (i.e., science and social studies). We evaluated the PCAs by observing how the *z*-scores of each factor varied as a function of grade level and genre. More specifically, we segregated six grade bands that have been adopted by the Common Core literacy standards of the Common Core State Standards Initiative (2010) and three genres (language arts, social studies, and science). The grade bands are *less than Grade 2, Grades* 2–3, *Grades 4–5, Grades 6–8, Grades 9–10,* and *Grade 11 and higher.* Given the very large number of texts in these analyses, all of the main effects and interactions were statistically significant at $p < 0.001$ for all eight PCs. We will therefore report the magnitude of variance explained by main effects and interactions (unadjusted eta squared, η^2) in a verbal form rather than in the form of quantitative magnitudes of eta-squared and *F*-scores. The unadjusted η^2 values are virtually indistinguishable from adjusted values because of the high sample size. We regard η^2 values of .01, .06, and .14 as being small, medium, and large effects, respectively.

The order in which the text characteristics are discussed in this section corresponds to the order in which they appear in an automated facility that is planned for dissemination through the website of the Common Core standards (http://www.corestandards.org/). We address the cohesion measures after the discussions of narrativity, syntactic simplicity, and word concreteness.

Narrativity (PC1). Narrativity captures the extent to which the text conveys a story, a procedure, or a sequence of episodes of actions and events with animate beings. This robust component is also affiliated with word familiarity, higher prior knowledge, and oral language. Informational texts on unfamiliar topics, typically appearing in print, are at the opposite end of the continuum (Biber, 1988; Tannen, 1982). Seventeen Coh-Metrix measures were loaded onto this component as primary measures, with loadings ranging from /.53/ to /.92/. Of the 119 cross-loadings reported for the PCA solution (17 variables × 7 components), only 7 values exceeded |.30|. When an overall narrative score was computed, some variables were reverse-scored so that all loadings would be positive, reflecting higher narrativity and greater ease of comprehension.

Given the robustness of narrativity, it is worthwhile to comment on the measures that contribute to this first principal

Table 1
Coh-Metrix Analysis of Printed Texts

Language-Discourse Measures	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Words								
Syllables per word	−							
Nouns	−							
Verbs	+							
Adjectives	−							
Adverbs	+							
Pronouns	+							
First-person pronouns	+							
Third-person pronouns	+							
Ratio of function words to content words			−			s+		
Connectives					+		s−	
Causal connectives					+			
Temporal connectives					+			
Logical connectives					+			
Additive connectives							+	
Adversative connectives							+	
Word frequency (logarithm)	+					s+		
Content word frequency (logarithm)	+					s+		
Minimum word frequency per sentence	+		s+					
Age of acquisition	−			s−				
Meaningfulness				+				
Concreteness				+				
Imagery				+				
Familiarity	+							
Negations	+							
Causal verbs			+					
Intentional actions, events, and particles	s+		+					
Polysemy—multiple senses of word						+		
Sentence structure								
Words per sentence				−				
Modifiers per noun phrase	−							
Words before main verb of main clause	s−			−				
Passive constructions	−							
Syntactic similarity—sentences in paragraph			+					
Connections between sentences								
Content word overlap—adjacent sentences		+						
Content word overlap—all sentences		+						
Argument overlap—adjacent sentences		+						
Argument overlap—all sentences		+	s−					
Noun overlap—adjacent sentences	s−	+						
Stem overlap—all sentences	s−	+						
Type-token ratio		−	s+					
Lexical diversity—all words		−						
Lexical diversity—verbs		−						
LSA-givenness versus newness		+						
LSA overlap—adjacent sentences	s−	+						
LSA overlap—all sentences	s−	+						
Dissimilarity of parts of speech between sentences	+	s−						
Dissimilarity of words between sentences		−						
Causal cohesion				s−	+			
Intentional cohesion					+			
Verb overlap—adjacent						+		
LSA verb overlap—adjacent sentences						+		
Temporal cohesion								+
Verb tense repetition								+
Verb aspect repetition								+

Note. PC = principal component; PC1 = narrativity; PC2 = referential cohesion; PC3 = syntactic simplicity; PC4 = word concreteness; PC5 = causal cohesion; PC6 = verb cohesion; PC7 = logical cohesion; PC8 = temporal cohesion. A plus (+) or minus (−) sign designates primary loading on a PC, in either a positive or negative direction. Secondary loading on a PC is designated by s+ or s−.

component. Some measures reflect the fact that narratives are replete with action descriptions: They tend to have more main verbs, adverbs, and intentional actions, events, and particles. In contrast, informational texts have more unfamiliar jargon that often takes the form of nouns. Many measures of narrativity capture characteristics of oral language (Biber, 1988; Clark, 1996; Tannen, 1982), which tends to be on familiar topics, with speech participants co-present (as opposed to the decontextualized language of print), and tends to involve sentence constructions that are easy for the audience to comprehend. Familiarity would explain a narrative's prevalence of short, high-frequency words representing an earlier age of acquisition. Co-presence would explain the prevalence of pronouns in narrative. The simpler syntactic constructions in narrative are characterized by shorter noun phrases, fewer words before the main verbs of main clauses, and fewer passive constructions. When we consider the measures affiliated with secondary loadings, we see that informational texts tend to have higher cohesion between sentences, as compared with narratives; cohesion apparently is one way to compensate for the greater difficulty of unfamiliar subject matter. The only measure of narrativity that is somewhat unpredictable is the higher relative frequency of negations.

Figure 1 plots narrativity z-scores as a function of the six grade levels and three genres. There were large decreases as a function of grade level and large differences in genre, whereas the interaction was small. As expected, narrativity scores were substantially higher for the language arts texts than for the two categories of informational text (science and social studies).

Syntactic simplicity (PC3). Scores are higher when sentences have fewer words and simpler, more familiar syntactic structures. At the opposite end of the continuum are structurally embedded sentences that require the reader to hold many words and ideas in working memory. Six variables were significantly loaded onto this component, with factor loadings that ranged from .52 to .85; only 2 of the 42 cross-loading values were greater than |.30|.

Syntactic simplicity z-scores showed large decreases as a function of grade level and medium-to-large differences in genre, with a small interaction. Syntactic simplicity showed the following ordering: science > social studies > language arts. From the standpoint of genre, writers use simpler syntax on the informational texts in which readers have lower prior knowledge. Simple syntax apparently compensates for the more unfamiliar and challenging subject matters in informational texts.

Word concreteness (PC4). Scores are higher when a higher percentage of content words are concrete, are meaningful, and evoke mental images—as opposed to being abstract. Three variables were significantly loaded onto this component, with factor loadings of .72 to .90 and low cross-loadings with all other components.

The z-scores for word concreteness showed medium effects of grade level and small effects of genre, with a small to medium interaction effect. Whereas science texts showed a decrease in word concreteness over grade level, there was a curvilinear function for language arts and social studies. Apparently, abstract words are used at early grade levels in the latter two genres. An exploration of the words in these earlier-grade-level texts would be worthwhile because the trend is somewhat counterintuitive.

Referential cohesion (PC2). This characteristic is the extent to which explicit content words and ideas in the text are connected with each other as the text unfolds. The noun phrases are important for providing co-reference and bridging the explicit clauses and sentences in the textbase. A referential cohesion gap occurs when a sentence has few if any words that overlap with previous sentences. There were 13 variables that loaded significantly on this dimension, with values for these loadings ranging from .52 to .86. Only 8 of the 91 cross-loading values were greater than |.30|.

Referential cohesion z-scores showed a small decrease over grade level, a moderate effect of genre, with a small interaction. The science texts had much higher referential cohesion than language arts and social studies, which were very similar. From the standpoint of the science genre, writers apparently fortify texts with higher referential cohesion, perhaps because the readers' prior knowledge is low and referential cohesion can compensate for the resulting challenges with the subject matter.

Causal cohesion (PC5). Scores are higher to the extent that clauses and sentences in the text are linked with causal and intentional (goal-oriented) connectives. Six Coh-Metrix measures were loaded onto causal cohesion, with loadings ranging from .47 to .86. Of the 42 cross-loadings (6 variables × 7 components), only 2 values exceeded |.30|.

Causal cohesion z-scores showed a small increase over grade level, a small effect of genre, and a small interaction. The science texts had slightly lower causal cohesion scores than the other two genres, which were approximately the same. The trend of increased causal cohesion scores over grade level suggests that texts that are more difficult or involve less familiar topics can benefit from discourse devices that promote higher causal cohesion. However, this explanation does not account for the lower causal cohesion scores of science texts compared with language arts and social studies. The major conclusion is that different forms of cohesion are not always positively correlated with grade-level bands and that sometimes cohesion devices can compensate for difficult topics.

The remaining three principal components (PC6, PC7, and PC8) address different aspects of situation model cohesion, namely verb cohesion, logical cohesion, and temporal cohesion. These had sensible and predictable effects on text variations, but it is beyond the scope of this article to describe them.

Brief Discussion of PCA Analyses

The results of the PCA analyses support a number of conclusions. One obvious conclusion is that orthogonal dimensions of text ease show significant changes over the grade levels and genre, with patterns that are very different for the scores plotted in Figure 1. The fact that the five dimensions are aligned with the multilevel theoretical frameworks of reading comprehension (Graesser & McNamara, 2011; Kintsch, 1998) is, of course, a reassuring confirmation of the value of the multilevel theoretical framework.

A second conclusion from Figure 1 is that the grade levels are primarily correlated with narrativity and syntax. Texts at higher grade levels tend to have more complex syntax, as would be expected. Texts at higher grade levels are also less likely to be narrative texts and more likely to contain features typical of
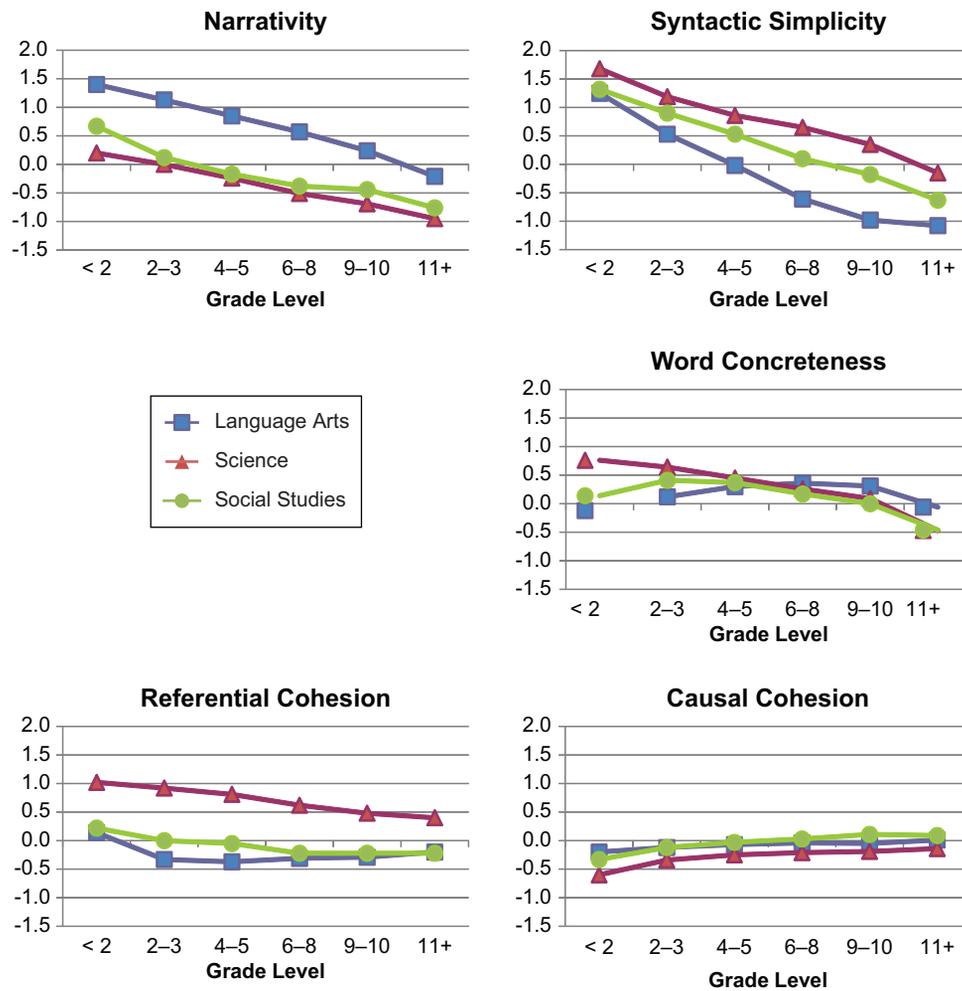
FIGURE 1. *Z-scores on text ease as a function of grade-level bands and text genre (language arts, science, and social studies). 11+ = 11th grade through college and career ready.*

informational texts on science, social studies, and other technical topics. Texts in school systems shift from narrative to informational texts as a function of grade, with Grades 3–5 spanning a critical period of transition from narrative to informational texts.

A third conclusion, however, is that text cohesion has a small variation over grade level (as defined by DRP), with a slight decrease for referential cohesion within most text genres and a slight increase for causal cohesion. Somehow the popular text complexity metrics (Flesch-Kincaid, DRP, Lexile) are not sensitive to picking up variations in text cohesion. These standard measures apparently are sensitive to local text constraints of words and sentences but not to cohesion between sentences and deeper levels of comprehension. We do know from studies in discourse processing that referential and causal cohesion both have a substantial impact on reading time and comprehension (Goldman et al., 1999; O'Brien et al., 1998; Zwaan & Radvansky, 1998), and also that Coh-Metrix is sufficiently sensitive to detect these particular cohesion manipulations (McNamara et al., 2010). Our hope is that future research with Coh-Metrix, computational linguistics, and empirical studies of reading will help us disentangle these mechanisms.

A fourth conclusion is that syntax and cohesion can sometimes compensate for difficult subject matter. When material is difficult, with unfamiliar ideas and words, authors tend to write the text with more simple syntax and more cohesion devices, perhaps to compensate for the difficulty of struggling with the ideas. These compensatory mechanisms explain the potential counterintuitive results that informational texts have more simple syntax and higher referential cohesion than language arts texts, and that causal cohesion tends to increase over grade level (McNamara et al., in press).

A fifth conclusion is that there sometimes is a curvilinear relationship between word concreteness and grade level. No one would be surprised by decreases in concreteness over grade level, but eyebrows will rise at the low concreteness scores at early grade levels. Perhaps early readers have many words that are more abstract than those in everyday usage. Examples of abstract nouns are *thing* and *person*, as opposed to basic level nouns (*ball, teacher*) and subordinate nouns (*basketball, principal*). Examples of abstract verbs are *go, do,* and *make*, compared with the more concrete *run, help,* and *juggle*. Follow-up research needs to be conducted to analyze the words in greater detail, as in the case of Beck et al.'s (2002) distinction between words that are Tier 1 (everyday frequent words, such as *home, came, dinner*), Tier 2 (low-to-medium frequency words used in more adult texts, such as *compost, fertilization, logistics*), and Tier 3 (words that are more

technical and need to be defined if most readers are to comprehend a passage, such as *spelunker, hydrogenous*).

Final Thoughts: Selecting Texts for Students

Automated analyses of texts can be used to guide the selection of texts according to particular pedagogical objectives. Sometimes students need to be challenged by texts on difficulty levels that aggressively push the envelope of what they can handle. At the other end of the continuum, sometimes students need a self-confidence boost by receiving easy texts that they can readily comprehend. For those who advocate Vygotsky's zone of proximal development, the texts should not be too difficult or too easy for students, but should occupy an intermediate zone of difficulty. Perhaps it is best for students to receive a "diet" that is balanced on the difficulty dimension, with a bias toward the intermediate zone.

Our vision is that texts can be scaled on the handful of Coh-Metrix scales, overall metrics of text ease (Flesch-Kincaid, Lexile, DRP), and qualitative descriptions by educational researchers, discourse experts, teachers, and students. The first two approaches have been automated, but the third has not. In either case, texts can be recommended or assigned on the basis of this complex profile of text characteristics.

Consider the types or combinations of texts that might be assigned:

1. *Challenging texts with associated explanations.* Some assigned texts are considerably beyond students' ability level. In such cases, students need comments by a teacher, tutor, group, or computer that explains technical vocabulary and points of difficulty. Students are greatly stretched by exposure to difficult content, strategies, and associated explanations.

2. *Texts at the zone of proximal development.* Some assigned texts are slightly above the difficulty level that students can handle. These texts gently push the envelope—they are not too easy or too difficult, but just right.

3. *Easy texts to build self-efficacy.* Easy texts are assigned to build reading fluency and self-efficacy. Struggling readers can lose self-confidence, self-efficacy, and motivation when beset with a high density of texts that they can barely handle, if at all.

4. *A balanced diet of texts at varying difficulty.* Texts may be assigned according to a distribution of alternatives 1, 2, and 3 above, mostly in the zone of proximal development. The balanced diet benefits from exposure to challenging texts, texts that gradually push the envelope, and texts that build self-efficacy. This approach also includes texts in different genres.

5. *Texts tailored to develop particular reading components.* Texts may be assigned adaptively in a manner that is sensitive to the student's complex profile of reading components. The texts attempt to rectify particular reading deficits or to advance particular reading skills.

These five approaches have all been explored by reading and discourse researchers, but it is beyond the scope of this article to comment on which approach best serves particular populations of readers.

Perhaps the central argument of this article is that Coh-Metrix has advantages over unidimensional metrics of text difficulty when one is considering the fifth of the approaches just listed. Coh-Metrix can play a role in assigning texts that expand the profile of skills at the level of words, syntax, text cohesion, and genre, whereas the unidimensional metrics do not differentiate particular reading components. When readers have problems with vocabulary, they need remediation on vocabulary. When they have problems with syntax or semantic composition of sentences, they need a very different form of remediation. And the same holds for different forms of text cohesion and discourse genre. Coh-Metrix can help us identify problems and solutions at a finer grain size within a multilevel framework of text and discourse comprehension.

NOTE

REFERENCES

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.

Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy* (Report to the Carnegie Corporation of New York). Washington, DC: Alliance for Excellent Education.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*, 6–28.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Common Core State Standards Initiative. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Council of Chief State School Officers and National Governors Association. Retrieved from http://www.corestandards.org/

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: Algorithm-guided individualized reading instruction. *Science*, *315*, 464–465.

Deane, P., Sheehan, K. M., Sabatini, J., Futagi, Y., & Kostin, I. (2006). Differences in text structure and its implications for assessment of struggling readers. *Scientific Studies of Reading*, *10*, 257–275.

Der, G., & Everitt, B. S. (2002). *A handbook of statistical analyses using SAS*. Boca Raton, FL: Chapman & Hall/CRC.

Goldman, S., Graesser, A. C., & van den Broek, P. (Eds.). (1999). *Narrative comprehension, causality, and coherence*. Mahwah, NJ: Lawrence Erlbaum.

Graesser, A. C., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A Web facility that helps survey

methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, *70*, 3–22.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 193–202.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, *114*, 357–374.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hiebert, E. H., & Fisher, C. W. (2007). The critical word factor in texts for beginning readers. *Journal of Educational Research*, *101*(1), 3–11.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Upper Saddle River, NJ: Prentice-Hall.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Klare, G. R. (1974–1975). Assessing readability. *Reading Research Quarterly*, *10*, 62–102.

Koslin, B. I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York: College Entrance Examination Board.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum.

Louwerse, M. M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*, 291–315.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*, 459–488.

McNamara, D. S., Graesser, A., & Louwerse, M. (in press). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini & E. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences*. Lanham, MD: Rowman & Littlefield Education.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*, 292–330.

Meyer, B. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., & Meier, C., et al. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly*, *45*(1), 62–92.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–244.

Millis, K., Magliano, J., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, *10*, 225–240.

Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy*, *52*(2), 96–107.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2001). Retrieved from http://www2.ed.gov/nclb/landing.jhtml

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1200–1210.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*, 121–152.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*, 228–242.

Pearson, P. D., & Hiebert, E. F. (2010). National reports in literacy: Building a scientific base for practice and policy. *Educational Researcher*, *39*, 286–294.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC2007: Linguistic inquiry and word count*. Austin, TX: LIWC.

Pentimonti, J. M., Zucker, T. A., Justice, L. M., & Kaderavek, J. N. (2010). Informational text use in preschool classroom read-alouds. *The Reading Teacher*, *63*, 656–665.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*, 91–110.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 289–312.

Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31–74.

Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher*, *39*, 183–197.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–406.

Stenner, A. J. (2006). *Measuring reading comprehension with the Lexile framework*. Durham, NC: Metametrics, Inc. Paper presented at the California Comparability Symposium, October 1996. Retrieved from http://www.lexile.com/DesktopDefault.aspx?view=re

Tannen, D. (1982). *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: Ablex.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *27*, 255–265.

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology*, *101*, 1–20.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

AUTHORS

ARTHUR C. GRAESSER is a professor in the Department of Psychology and co-director of the Institute for Intelligent Systems at the University of Memphis, 400 Innovation Drive, 202 Psychology Building, Memphis, TN 38152; *a-graesser@memphis.edu*. His research focuses on cognitive science, discourse processing, computational linguistics, and

the development of learning, language, and discourse technologies (including AutoTutor and Coh-Metrix).

DANIELLE S. McNAMARA is a professor in the Department of Psychology and director of the Institute for Intelligent Systems at the University of Memphis, 400 Innovation Drive, 202 Psychology Building, Memphis, TN 38152; *dsmcnamara1@gmail.com*. Her research focuses on cognitive science, memory, knowledge acquisition, reading comprehension, writing, and the development of educational technologies that incorporate principles of cognition (including Coh-Metrix, iSTART, and Writing Pal).

JONNA M. KULIKOWICH is a professor in the Department of Educational Psychology at Pennsylvania State University, 274 Chambers Building, University Park, PA 16802; *jmk35@psu.edu*. Her research interests are in multilevel modeling for academic development and domain learning, reading comprehension, applied statistics, and measurement of variables in reading research.