Coh-Metrix Easability Components: Aligning Text Difficulty with Theories of Text Comprehension

Danielle S. McNamara, Arthur C. Graesser, Zhiqiang Cai,
University of Memphis

and Jonna M. Kulikowich
The Pennsylvania State University

Abstract

The purpose of this project is to explore the use of advanced measures of linguistic characteristics of text and discourse provided by an automated tool called Coh-Metrix in order to assess the characteristics of texts that make them easy or difficult to read. The scaling of texts on difficulty (or ease) has enormous practical value in education in addition to advancing scientific theories of reading and comprehension. 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* and *Race to the Top*. Advances in psycholinguistics, discourse processes, and cognitive science provide a theoretical foundation for scaling texts on multiple levels. To this end, a principal components analysis was conducted on Coh-Metrix indices for 37,520 texts representative of what typical senior in high school might encounter kindergarten through 12th grade. The results indicated that eight components accounted for 67.3% of the variance, with five main components associated with text ease, referred to as Coh-Metrix Easability Components. These five components closely align with the theoretical levels text and discourse comprehension, including narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. These components show similar trends across multiple data sets, with the results supporting the underlying assumption that text difficulty can best be characterized multidimensionally. Currently, an online tool is being developed to provide the Coh-Metrix Easability Components to educators.

Contact:
Danielle S. McNamara
dsmcnamara1@gmail.com

Coh-Metrix Easability Components:
Aligning Text Difficulty with Theories of Text Comprehension

This project explores the use of advanced measures of linguistic characteristics of text and discourse in order to assess text ease or difficulty in theoretically informed ways. An automated analysis of texts, such as that provided by Coh-Metrix, on different characteristics has enormous practical value in education in addition to advancing scientific theories of reading and comprehension. 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, Race to the Top*, the *Common Core State Standards* of the National Governors Association (2009), and other high stakes assessments. The nature of the texts is presumably just as important as the interventions and curriculum to develop reading skills and strategies.

There are times when students need to be challenged by assigning them texts on difficulty levels that aggressively push the envelope on what they can handle. There also are times when students need a self-confidence boost by receiving easy texts that they can readily comprehend. In most situations, however, the texts should not be too difficult or too easy for students, but rather at an intermediate zone of difficulty. A widespread assumption is that the texts should be sensitive to the individual students' proficiency profiles. In order to strategically assign texts to students, it is necessary to scale the texts on ease or difficulty. This study explores the value of going beyond traditional measures of text readability by incorporating linguistic indices inspired by theories of text and discourse comprehension.

Advances in psycholinguistics, discourse processes, and cognitive science provide a theoretical foundation for characterizing texts on multiple levels. Multilevel theoretical frameworks have identified the representations, structures, strategies, and processes at multiple levels of language and discourse. These multilevel frameworks have typically included the following levels: *words, syntax*, the explicit *textbase*, the referential *situation model* (sometimes called the mental model), the discourse *genre and rhetorical structure* (the type of discourse and its composition), and the *pragmatic communication* level (between speaker and listener, or writer and reader). Coh-Metrix was developed to analyze and measure text on the first five levels of discourse: *words, syntax, textbase, situation model,* and *genre* (Graesser & McNamara, in press; McNamara, Louwerse, McCarthy, & Graesser, 2010).

Coh-Metrix was designed to move beyond standard readability formulas, such as Flesch-Kincaid Grade Level (Klare, 1975-1976), in measuring text difficulty. Most readability formulas rely exclusively on word length and sentence length. For example, Formula 1 shows the scoring of Flesch-Kincaid Grade Level. *Words* refers to mean number of words per sentence and *syllables* refers to mean number of syllables per word.

$$Grade\ Level = .39 * Words + 11.8 * Syllables - 15.59 \quad (1)$$

Sentence length and word length do in fact robustly predict reading time (Haberlandt & Graesser, 1985; Just & Carpenter, 1987; Rayner, 2003), but certainly there is more to text difficulty than word and sentence length. Coh-Metrix was designed to measure deeper characteristics of language and cohesion.

Coh-Metrix Measures of Language and Cohesion

The number and particular measures that are provided by Coh-Metrix depends on the version and the type of tool. We have developed public versions of the tool that analyze individual texts and have provided between 40 and 80 validated indices. We have also developed internal versions of Coh-Metrix that analyze texts in batches and that include 600-1000 indices, many of which are redundant and many of which have not been validated (and thus we do not release them to the public). Although the specific Coh-Metrix measures vary somewhat across versions and tools, the banks of measures are quite similar.

Descriptive Indices. Coh-Metrix provides descriptive indices such as the number and length of words, sentences, and paragraphs. These indices help the user to check the Coh-Metrix output (e.g., to make sure that the numbers make sense) and also to interpret patterns of data.

Words and Sentences. Coh-Metrix assesses text difficulty at the word and sentence level, but in comparison to readability measures, does so in ways that are driven by theories of reading comprehension. For example, in addition to word frequency, Coh-Metrix assesses the degree to which the text contains concrete versus abstract words. When a text contains more concrete words (e.g., table, mountain, train) as opposed to abstract words (e.g., love, energy, law), it is easier to process. Coh-Metrix assesses words on concreteness, abstractness, parts of speech, familiarity, multiple senses of a word, age of acquisition, and many other features. At the sentence level, Coh-Metrix directly analyzes syntactic complexity or simplicity, which relates to the number of modifiers in noun phrases and the number of words before the main verb in a sentence. When a sentence has fewer noun phrases, or fewer words before the main verb, it is easier to parse and understand.

Referential Cohesion. Referential (or co-referential) cohesion refers to overlap in content words between local sentences. For example, this sentence overlaps with the previous sentence by means of the singular form of the word *sentence*. Referential cohesion is measured both locally (between nearby sentences) and globally (e.g., all sentences and paragraphs). It is measured both strictly via explicit noun overlap and via stem overlap. Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) also provides measures of semantic overlap between sentences or between paragraphs. LSA considers meaning overlap between explicit words and also words that are implicitly similar or related in meaning. For example, *child* in one sentence will have a relatively high degree of semantic overlap with *infant* and *mother* in another sentence. LSA uses a statistical technique called singular value decomposition to condense a large corpus of texts to 100-500 statistical dimensions. The conceptual similarity between any two text excerpts (e.g., word, clause, sentence, text) is computed as the geometric cosine between the values and weighted dimensions of the two text excerpts. The value of the cosine typically varies from 0 to 1. Coh-Metrix measures LSA-based cohesion in several ways, such as LSA similarity between adjacent sentences, LSA similarity between all possible pairs of sentences in a paragraph, and LSA similarity between adjacent paragraphs. When a text has greater referential cohesion, it is easier it is to understand (McNamara et al., 2010).

Lexical Diversity. Lexical diversity refers to the variety of words (*types*) that occur in a text in relation to the number of words (*tokens*). When the number of word types is equal to the total number of words (tokens), then all of the words are different. In that case, lexical diversity is at a maximum, and the text is either very low in cohesion or very short. Lexical diversity is lower (and cohesion is higher) when more words are used multiple times across the text. Coh-Metrix includes three indices of lexical diversity: type-token ratio (TTR), *vocd*, and the Measure of

Textual Lexical Diversity (MTLD). TTR is simply the number of unique words divided by the number of tokens of the words. The index produced by *vocd* is calculated through a computational procedure that fits TTR random samples with ideal TTR curves. MTLD is calculated as the mean length of sequential word strings in a text that maintain a given TTR value. TTR is correlated with text length because as the number of word tokens increases, there is a lower likelihood of those words being unique. Measures such as *vocd* and MTLD overcome that confound by using sampling methods (McCarthy & Jarvis, 2010).

Connectives. Coh-Metrix provides an incidence score (occurrence per 1000 words) for all connectives as well as different types of connectives. Indices are provided on five general types of connectives: causal (*because, so*), additive *(and, moreover)*, temporal *(first, until)*, logical *(and, or)*, and adversative/contrastive *(although, whereas)*. In addition, there is a distinction between positive connectives (*also, moreover*) and negative connectives (*however, but*).

Situation Model. Referential cohesion is one linguistic feature that can influence a reader's deeper understanding of text. However, there are deeper levels of meaning that go beyond the words. The expression *situation model* has been used by researchers in discourse processing and cognitive science to refer to the deeper meaning representations that involve much more than the explicit words (van Dijk & Kintsch, 1983; Graesser & McNamara, in press; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; Zwaan & Radvansky, 1998). For example, with episodes in narrative text, the situation model would include *the plot*; and, in an informational text about the circulatory system, the situation model might include *the flow of the blood*. The situation model comprises a mental representation of the deeper meaning of the text (Kintsch, 1998; Singer & Leon, 2007). The situation model may be more or less coherent to the extent that the reader draws inferences that connect the ideas in the text to each other and to information outside of the text (i.e., prior knowledge). These inferences can be induced and supported by certain cues in the text that are measured by Coh-Metrix. These include measures of causality, such as incidence scores for causal verbs, intentional verbs, and causal particles (e.g., connectives like *because*, *in order to*), and ratios of causal particles to causal verbs. There are also measures of verb overlap. These indices are indicative of the extent to which verbs (which have salient links to actions, events, and states) are repeated across the text. There are measures of temporal cohesion and logical cohesion as well. These and other measures are reflective of elements of a text that are more or less likely to support a reader's construction of a coherent situation model.

Current Study

Well over 70 studies have been conducted in our laboratory using Coh-Metrix, and the number grows steadily each year (see McNamara et al., 2010; McNamara & Graesser, in press, for reviews). In our laboratory, Coh-Metrix is often used to understand the features of texts that are used in the context of experimental studies of text comprehension. We have also conducted studies using Coh-Metrix on natural language processing. These include studies examining a wide range of topics and tasks such as deception, paraphrasing, explaining, answering questions, and writing essays. In addition, studies have been conducted to verify that the indices measure what we think they are measuring and that they are theoretically compatible with patterns of data corresponding to types of texts or human performance.

One motivation for the development of Coh-Metrix was to provide better measures of text difficulty (see e.g., Duran, Bellissens, Taylor, & McNamara, 2007), and particularly the specific sources of both challenges and scaffolds within texts. One limitation of traditional

readability measures is that they consider only features of text that tend to be predictive of readers' surface understanding of the words and separate sentences. In turn, aassessments that are used to validate or provide readability scores most often use cloze tasks. Cloze tasks (i.e., filling in missing words in text) assess comprehension only within sentences based on word associations (Shanahan, Kamil, & Tobin, 1982) and depend primarily on decoding rather than language comprehension skills (Keenan, Betjemann, & Olson, 2008). However, many comprehension models (Graesser & McNamara, in press; Kintsch, 1998; McNamara & Magliano, 2009; Van Dijk & Kintsch, 1983) propose that there are multidimensional levels of understanding that emerge during the comprehension process, including (at least) surface, textbase, and situation model levels. Readability formulas by contrast assume a uni-dimensional representation.

Uni-dimensional representations of comprehension ignore the importance of readers' deeper levels of understanding. Another limitation is that uni-dimensional metrics of text difficulty are not particularly helpful to educators when specific guidance is needed for diagnosing what is wrong and planning remediation for students. Readability formulas do not identify the particular characteristic of the text that may be challenging or *helpful* to a student. Given the multi-dimensional profile of scores per text that can be made available to teachers and their students, one major advantage of Coh-Metrix is that it provides metrics on multiple levels of language and discourse.

One reason that the difficulty or ease of a text should not be assessed at only one, uni-dimensional, level of language is because multiple levels often work to compensate for one another (McNamara, Graesser, &Louwerse, in press). For example, in a recent analysis we found that narrative texts (stories, novels) tend to have low referential cohesion and low verb cohesion. By themselves, these indices may imply that narratives are difficult to read. However, narratives also tend to be composed of more frequent words and they often have high causal cohesion and temporal cohesion, allowing the reader to form a coherent mental model of the contents of the text. There are times when language in narrative texts at the situation model levels compensates for more challenging sentences and low overlap in words and concepts. By contrast, science texts are composed of rare words, making it challenging for students to understand the concepts in the text. These challenges are sometimes offset by reducing the syntactic complexity of the text and increasing the overlap in words and concepts (i.e., referential cohesion).

Thus, Coh-Metrix, in contrast to traditional measures of text readability, has the potential to offer a more complete picture of the potential challenges that may be faced by a reader as well as the potential scaffolds that may be offered by the text. Coh-Metrix is motivated by theories of discourse and text comprehension. Such theories describe comprehension at multiple levels, from shallow, text-based comprehension to deeper levels of comprehension that integrates multiple ideas in the text and brings to bear information that elaborates the ideas in the text using world and domain knowledge (Graesser & McNamara, in press). Coh-Metrix assesses challenges that may occur at the word and sentence levels. In addition, it is able to assess deeper levels of language in terms of cohesion and the situation model. By doing so, it comes closer to having the capability to estimate how well a reader will comprehend a text at deeper levels of cognition.

Method

We conducted analyses on a large corpus of texts that were scaled on the Coh-Metrix measures. The sample of texts consisted of 37,520 texts from the 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. There are a number of reasons for selecting the TASA corpus other than the fact that the corpus is large and readily available. One important reason is that this corpus is representative of the texts that a typical senior in high school would have encountered from kindergarten through 12th grade. Another is that the texts are scaled on Degrees of Reading Power (DRP), which can approximately be translated into grade level estimates corresponding to traditional readability formulas (Koslin, Zeno, & Koslin, 1987; McNamara et al., in press). Each text is also assigned to a text category by the TASA researchers. Most of the text genres were classified in language arts (n = 15,991), science (n = 5349), and social studies/history (n = 10,438), but other categories were business, health, home economics, and industrial arts.

Coh-Metrix provides dozens of measures of text characteristics and most measures have a predictable impact of ease or difficulty of processing. However, many of the measures are highly correlated because of natural constraints of meaning (e.g., texts with higher lexical diversity will tend to have lower lexical referential cohesion scores) and also because we considered alternative metrics that attempt to capture a particular theoretical construct (e.g., the five types of lexical referential cohesion). The correlations and redundancy among measures motivated us to perform quantitative analyses that reduce the number of functional dimensions.

In order to discover what aspects of texts account for text complexity, a principal components analysis (PCA) with Varimax rotation was conducted on Coh-Metrix indices for the sample of 37,520 texts from the TASA corpus. We adopted PCA in our efforts to reduce our large multivariate database of Coh-Metrix indices to fewer functional dimensions. PCA is routinely used in such data reduction efforts in computational linguistics and educational data mining (e.g., Brun, Ehrmann, & Jacquet, 2007; Honkela, Hyvarinen, & Vayrynen, 2010; Nerbonne, Gooskens, Kürschner, & van Bezooijen, 2008). The final analysis included 53 Coh-Metrix measures that reduced to eight principal components. Because there were a very large number of texts and Coh-Metrix measures, a high Kaiser criterion of 2.0 was adopted to determine the number of principal components.

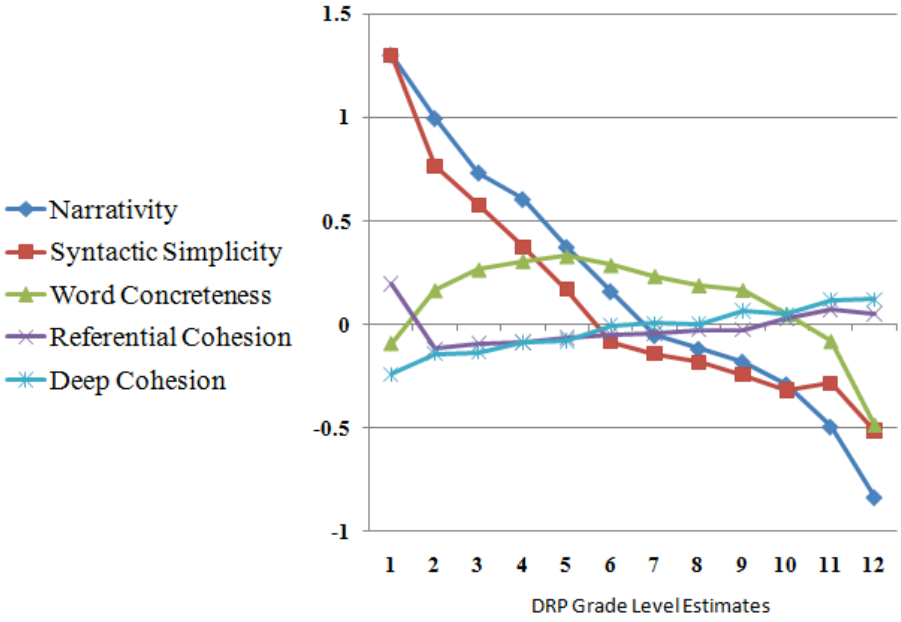
Approximately 80 measures of Coh-Metrix were examined. Of these, 53 variables were selected on the basis of multiple criteria. First, some variables had extremely high correlations with other variables ($|r| \geq 0.90$), which presents a problem of collinearity. For example, the relative frequency of pronouns and the ratio of pronouns to noun-phrase units were correlated .99. In this case, we retained the relative frequency of pronouns and discarded the ratio measure to handle the high redundancy between these two measures. Second, the distribution of scores for some measures substantially deviated from a normal distribution, being highly skewed and kurtotic. Attempts at transformations using arcsine and logarithmic equations did not correct for extreme departures from normality. Scores for these variables were discarded unless there was a compelling reason to include them, such as the measure being important theoretically or there being a lack of alternative measures.

Results

The results indicated that eight components accounted for 67.3% of the variance. These components closely aligned with the theoretical levels of narrativity (genre), syntactic simplicity, word concreteness, referential cohesion (textbase), and deep cohesion (situation model). The figure below shows the Z-scores on the first five of the Coh-Metrix Easability Components as a

function of DRP Grade Level Estimates. We selected the first 5 of the 8 components not only because these dimensions accounted for significant variance (i.e., 54%), but also because these components aligned most readily and practically with notions of the ease of processing text. As such, these five components are expected to be of practical use as teachers consider the selection of texts in light of their instructional and assessment goals.

Figure 1. *Component z-scores on narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion as a function of DRP-based grade levels.*



The five principal components map directly onto the levels of the multilevel theoretical framework articulated by Graesser and McNamara (in press). There are five basic levels in this framework after excluding pragmatic communication, a level that is more affiliated with the communication setting between reader/write or speaker/listener than to characteristics of the text per se. The five components align with these five levels in the following manner.

1. **Narrativity.** Narrative text tells a story, with characters, events, places, and things that are familiar to the reader. Narrative is closely affiliated with everyday oral conversation. This robust component is highly affiliated with word familiarity, world knowledge, and oral language. Non-narrative texts on less familiar topics would lie at the opposite end of the DRP grade-level continuum. Specifically, as grade levels increase, texts assigned are more likely to lack narrativity. As shown in Figure 1, the correlation between DRP Grade level estimates and Narrativity is $r = -.671$, $p < .001$.

2. **Syntactic Simplicity**. This component reflects the degree to which the sentences in the text contain fewer words and use simple, familiar syntactic structures, which are less challenging to process by the reader. At the opposite end of the continuum are texts that contain sentences with more words and that use complex, unfamiliar syntactic structures.

As shown in Figure 1, the correlation between DRP Grade level estimates and Syntactic Simplicity is $r = -.489$, $p < .001$.

3. **Word Concreteness**. Texts that contain content words that are concrete, meaningful, and evoke mental images are easier to process and understand than those texts which contain words that are abstract. Abstract words represent concepts that are difficult to represent visually, and as such, it is difficult for readers to generate a mental picture of what these words mean. "Running water" is easy to image. By comparison, "complex concept" arguably has no space-time image, yet we use such abstract phrases many times in discourse. Texts that contain more abstract words or phrases are more challenging to understand. As shown in Figure 1, the correlation between DRP Grade level estimates and Word Concreteness is $r = -.230$, $p < .001$.

4. **Referential Cohesion**. This component includes Coh-Metrix indices that assess referential cohesion. High cohesion text contains words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader. Low cohesion text is typically more difficult to process because there are fewer threads that tie the ideas together for the reader. As shown in Figure 1, the correlation between DRP Grade level estimates and Referential Cohesion is $r = .028$, $p < .001$. In other words, there is little relation between readability measures such as the DRP and referential cohesion.

5. **Deep cohesion.** This dimension reflects the degree to which the text contains causal, intentional, and temporal connectives. These connectives help the reader to form a more coherent and deeper understanding of the causal events, processes, and actions in the text. As shown in Figure 1, the correlation between DRP Grade level estimates and Deep Cohesion is $r = .110$, $p < .001$. As with referential cohesion, deep cohesion also has very little relation to readability measures such as the DRP. Thus, the cohesion components assess characteristics of the text that go beyond traditional readability.

The results indicate that differences between texts can be characterized with Coh-Metrix, and the extracted components align with theories of text comprehension. In the current study, we have examined the correlations between DRP and the Coh-Metrix Easability Component Scores, observing that the cohesion components in particular capture dimensions of text that are not correlated with traditional measures of text difficulty (i.e., readability; see also McNamara et al., 2010).

Text Interpretations

Our central assumption is that text meaning is multi-dimensional and thus characterizing a text requires the consideration of multiple components. For example, a text might have challenging syntax and high text cohesion, which will have different effects for a reader than a text that has challenging syntax with low cohesion. In our view, texts should be viewed from a multi-dimensional prism that may or may not converge on a single dimension of ease.

Notably, text characteristics have both qualitative and quantitative aspects that need to be considered. An analysis of the complexity of themes, satire, humor, ethics, and cultural sensitivity are not within the horizon of computer analyses and singular dimensions, so there is

no substitute for a thoughtful critical analysis of texts by human experts. Nevertheless, we believe that the measures of Coh-Metrix should be considered in such thoughtful critical analyses. These analyses will take a long way in better understanding the nature of text.

To provide an example of the use of Coh-Metrix, we can compute percentile scores for the five components. A percentile score varies from 0 to 100%, with higher scores meaning the text is likely to be easier to read than other texts in the corpus. For example, a percentile score of 80% means that 80% of the texts are more difficult and only 20% are easier. Four such texts are presented in Figure 2, including two narrative texts and two science texts.

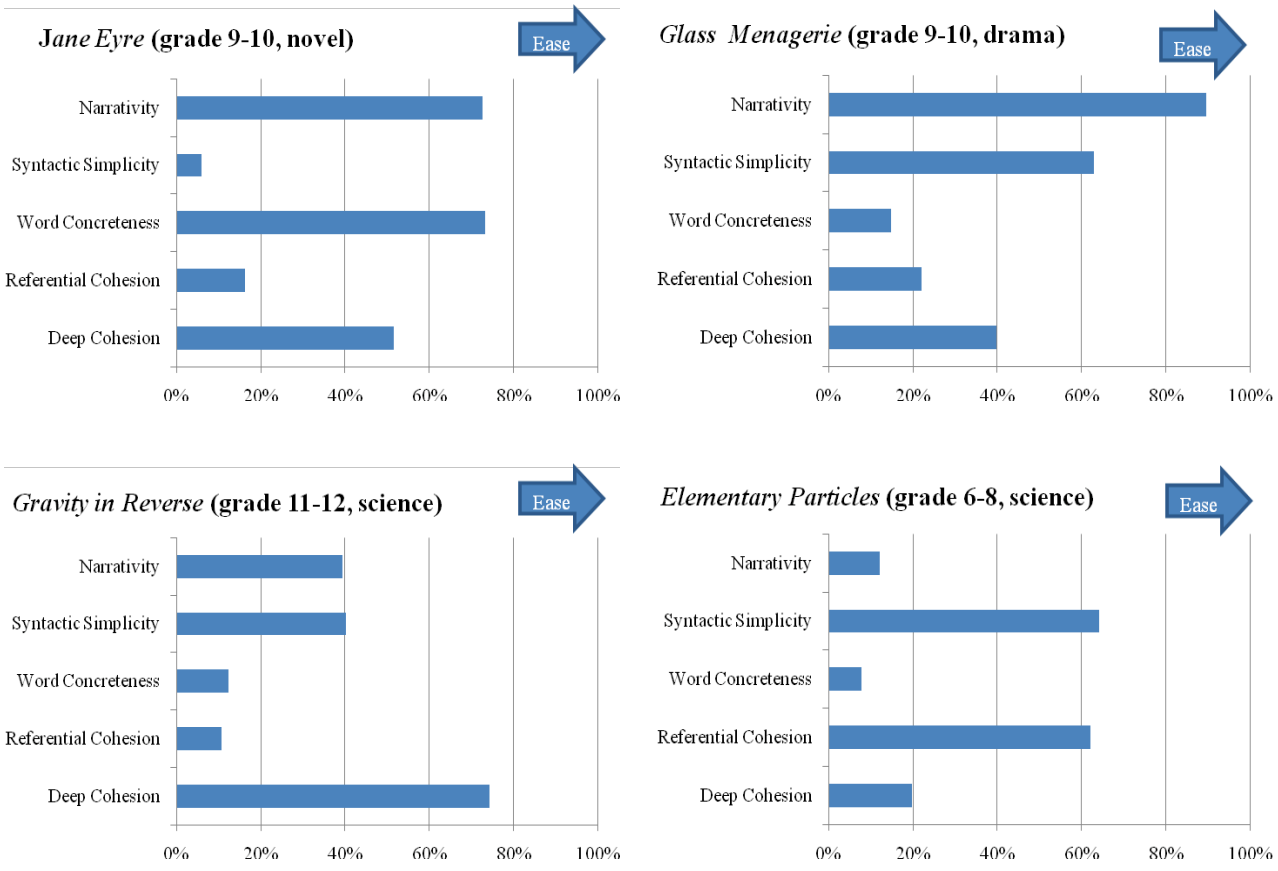
As would be predicted, the novel and drama texts are substantially more narrative than the two science texts. It is well documented that narrative is easier to read than informational texts (Bruner, 1986; Graesser, Olde, & Klettke, 2002; Haberlandt & Graesser, 1985), a fact that should be incorporated in assessments of text ease and difficulty. However, it is possible for narratives to have informational content that explains the setting or context, and it is possible for science texts to have story-like language (e.g., the journey of a water molecule through the water system). In either case, the move toward increased narrativity will generally make the text easier to understand.

The two example narrative texts (Jane Eyre and Glass Menagerie) are declared to be at least at the 9-10 grade band. However, they have very different profiles on the various dimensions. Jane Eyre is more difficult when inspecting syntax, but it is less difficult with respect to word concreteness and deep cohesion. These are different profiles of text ease even though both texts were declared to be in the Grade 9-10 band.

The two science texts represent different grade bands. Elementary Particles is grade 6-8 whereas Gravity in Reverse is grade 11-12. However, Gravity in Reverse is not uniformly more difficult on all Coh-Metrix dimensions. Gravity in Reverse is more difficult with respect to referential cohesion and syntax, but the opposite is the case with respect to narrativity and deep cohesion. Therefore, these comparative profiles should raise questions about the grade levels for which these texts might be assigned given instructional goals.

Such a *picture* of texts is hoped to provide teachers with more information about text ease as well as the potential challenges and scaffolds that might arise from certain text characteristics. This information can in turn inform their pedagogical practice. One goal is to bring to fruition the Coh-Metrix project so that educators can more fully understand the complexity and nature of the texts that they use in the classroom, particularly in concert with their pedagogical goals and the individual abilities of their students.

Figure 2. *Coh-Metrix Easability Component percentile scores on narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion for four representative texts.*



Conclusions

Our goal in the Coh-Metrix project has been to provide indices for the characteristics of texts on multiple levels of analysis, including word characteristics, sentence characteristics, and the discourse relationships between ideas in text, within one tool. Our objective has been to go beyond traditional measures of readability that focus on surface characteristics of texts, which in turn tend to principally affect surface comprehension. Indeed, the validation of traditional readability algorithms using word and sentence characteristics (e.g., number of letters or number of words) has been almost exclusively done using assessments that primarily tap surface comprehension (e.g., cloze tests). Coh-Metrix can be used to better understand differences between texts and to explore the extent to which linguistic and discourse features successfully distinguish between text types. Why? Because analysis of Coh-Metrix indices supports a multi-dimensional, rather than a uni-dimensional, operational definition of text ease and difficulty.

Our work with Coh-Metrix has culminated most recently in the development of the Coh-Metrix Easability Components. These components provide a more complete picture of text ease (and hence potential challenges) that may emerge from the linguistic characteristics of texts. The

Easability components provided by Coh-Metrix go beyond traditional readability measures by providing metrics of text characteristics on multiple levels of language and discourse. Moreover, they are well aligned with theories of text and discourse comprehension (e.g., Graesser et al., 1994; Kintsch, 1998).

We are currently developing a tool to provide the Coh-Metrix Easability Components to educators. Funded by the Gates Foundation, and aligned with the Common Core State Standards, we are developing a tool that allows an educator to access graphical representations for text excerpts that are entered online, as well as for a library of texts chosen for the Core Standards project. The tool will provide a graph (similar to those in Figure 2) as well as a verbal interpretation of the graph for each text of interest.

Coh-Metrix offers numerous opportunities to researchers, teachers, and students across a variety of fields. With the unlimited wealth and diversity of texts that technology such as the Internet brings (Wiley et al., 2010; Leu, O'Byrne, Zawilinski, McVerry, & Everett-Cacopardo 2009), computer assessments of text characteristics stand to increase dramatically in the next few years. Coh-Metrix offers what may be the best approach for such assessments to be conducted in light of a multi-dimensional understanding of why a text is easy or difficult to comprehend.

References

Brun, C., Ehrmann M., & Jacquet G., (2007). *A hybrid system for named entity recognition*. Proceedings of the 4th International Workshop SemEval, ACL 2007. Prague, Czech Republic.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Duran, N. D., Bellissens, C., Taylor, R. S., & McNamara, D. S. (2007). Quantifying text difficulty with automated indices of cohesion and semantics. In D. S. McNamara & G. Trafton (Eds.)*, Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 233-238). Austin, TX: Cognitive Science Society.

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

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 231-263). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

Honkela, T., Hyvärinen, A., & Väyrynen. J. (2010). Word ICA: Emergence of linguistic representations for words by independent component analysis. *Natural Language Engineering, 16,* 277-308.

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

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281-300.

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

Klare, G. R. (1975). *A manual for readable writing.* Glen Burnie, MD: REMco.

Klare, G. R. (1976). A second look at the validity of the readability formulas. *Journal of Reading Behavior, 8,* 159-152.

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

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). (Eds.). *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Erlbaum.

Leu, D. J., O'Byrne, I., Zawilinski, J., McVerry, G., & Everett-Cacopardo, H. (2009). Expanding the new literacies conversation. *Educational Researcher, 38*, 264-269.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods,* 381-392.

McNamara, D. S., & Graesser, A. C. (in press). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

McNamara, D. S., Graesser, A. C., & Louwerse, M. 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: R&L Education.

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.

McNamara, D. S., & Magliano, J. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation*, Vol. 51.

Nerbonne, J., Gooskens, C., Kürschner, S., & van Bezooijen, R. (2008). Language variation studies and computational humanities. *International Journal of Humanities and Arts Computing*, *Special Issue on Language Variation*, 1-18.

Rayner, K. (2003). Eye movements in reading. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences.* Amsterdam: Pergamon Press.

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 17*, 229-255.

Singer, M., & Leon, J. (2007). Psychological studies of higher language processes: Behavioral and empirical approaches. In F. Schmalhofer & C. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, NY: 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, 46*, 1060-1106.

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