Title
Validating Coh-Metrix

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN
1069-7977

Authors
Graesser, Art
Louwerse, Max
McNamara, Danielle S.
et al.

Publication Date
2006

Peer reviewed

Validating Coh-Metrix

Danielle S. McNamara (d.mcnamara@ mail.psyc.memphis.edu)
Yasuhiro Ozuru (y.ozuru@mail.psyc.memphis.edu)
Art Graesser (a.graesser@mail.psyc.memphis.edu)
Max Louwerse (m.louwerse@mail.psyc.memphis.edu)
Psychology Department, University of Memphis
Memphis, TN 38152 USA

Abstract

Coh-Metrix is a computer program that analyzes various text features relevant to text comprehension by incorporating techniques informed by theories of text processing, cognitive psychology, and computational linguistics. Three key classes of cohesion indices (i.e., coreference, conceptual relations, connectivity) measured by Coh-Metrix are evaluated with texts used in published studies of cohesion effects on reading comprehension. The results confirmed that Coh-Metrix successfully detects levels of cohesion in texts.

Introduction

Many studies, across a variety of paradigms and dependent measures, have shown that cohesive cues in text facilitate reading comprehension (Britton & Gulgoz, 1988; McNamara, 2001; Zwaan & Radvanksy, 1998). Cohesion is the degree to which ideas in the text are explicitly related to each other. Cohesion differs from coherence in that cohesion is an objective property of the text and coherence refers to the quality of the mental representation constructed by a reader. As such, cohesive elements are text features that tend to help readers construct a coherent mental representation of text content. Cohesion arises from a variety of sources, including explicit argument overlap and causal relationships, and can operate between sentences, groups of sentences, paragraphs, and chapters (Givón, 1995; Graesser, McNamara, & Louwerse, 2003; Kintsch, 1995). Cohesion particularly enhances comprehension of text, or coherence, for low-knowledge readers because they are less able to infer relations between the ideas in the text (Loxterman, Beck, & McKeown, 1994; McNamara, 2001; McNamara et al., 1996).

The notion that comprehension partially depends on text cohesion has led to the development of a computational tool, called Coh-Metrix (Graesser et al., 2004; cohmetrix.memphis.edu). This tool augments conventional readability formulas with computational indices of text cohesion. Coh-Metrix v1.4 has lexicons, part-of-speech classifiers, syntactic parsers, latent semantic analysis, and several other components that are widely used in computational linguistics. For example, the MRC database (Coltheart, 1981) is used for psycholinguistic information; WordNet (Miller, Beckwith, Fellbaum, Gross & Miller, 1990) for hypernymy and hyponymy relations; Latent Semantic Analysis (Landauer & Dumais, 1997) for the semantic similarities between words, sentences, and paragraphs; and the ApplePie parser (Sekine & Grishman, 1995) and the Brill (1995) part-of-speech tagger for a variety of syntactic categories. Graesser et al. (2004) provide an extensive overview of the many language features provided. Coh-Metrix currently analyzes texts on three major categories of cohesion: coreference, conceptual (LSA), and connectivity (including causal cohesion). One of the goals in the Coh-Metrix project is to compare the various measures and determine which ones best account for cohesion and coherence.

For decades, computational measures of text difficulty have been restricted to *readability* formulas. Readability formulas, such as the Flesch-Kincaid Grade Level (FKGL; Klare, 1974-1975), focus on shallow textual aspects, such as sentence and word length. Certainly these features have considerable validity as indices of text difficulty. However, such shallow aspects alone explain only a part of text comprehension, and ignore many language and discourse features that are theoretically influential at estimating comprehension difficulty.

Recent advances across various disciplines have made it possible to computationally measure characteristics of text and language. These measures can now supercede surface components of text and instead explore deeper, more global discourse attributes (Graesser et al., 2004). A number of studies have used the Coh-Metrix to distinguish between different types of texts. For example, Louwerse et al. (2004) used Coh-Metrix to distinguish significant differences between spoken and written varieties of English; and McCarthy et al. (2006) demonstrated that Coh-Metrix successfully detected authorship even though individual authors recorded significant shifts in their writing style. The goal of the current study is to validate the indices of cohesion provided in Coh-Metrix by verifying its ability to discriminate between high and low-cohesion versions of texts. To this end, we have collected 19 pairs of texts with high and low-cohesion versions from 12 published studies.

Corpus

The texts were collected from prior experimental studies that have investigated the effect of referential and causal cohesion on comprehension of natural (multi-paragraph) text. We initially identified 34 studies on text revision that met our criteria by searching through journal articles and reviews (e.g., Britton, Gulgoz, & Glynn, 1993). We were able to obtain the texts for 13 of the 34 studies. Two studies were redundant, however, because they used the same texts. This resulted in a total of 19 pairs of texts from 12 studies, described below. Although some of these studies included more than two versions of the texts, this analysis is limited to the comparison between the highest and lowest cohesion versions.

The *Raccoon* text in the Beck, McKeown, Omason, & Pople (1984) study was from a 2nd grade text book. The revision alleviated problems in surface structure (e.g., syntactic complexity, unclear relations between reference and referent), unfamiliarity of events, and ambiguity or confusability of the content. Four social studies texts from Beck, McKeown, Sinatra, and Loxterman (1991) were from a 5th grade text book. The text revisions made connections more explicit and increased causal connections between the ideas, concepts, and events by clarifying, elaborating, explaining, and motivating important information. The *El Nino* text in the Loxterman, Beck, and MacKeown (1994) study was obtained from a 6th grade social studies text book. The revisions followed the same method used in the Beck et al. (1991) study.

The *Air War in the North*, a college level text used in Britton and Gulgoz (1991; McNamara & Kintsch, 1996), was revised based on a specific theory of text processing (Kintsch & van Dijk, 1978). Coherence breaks were repaired in the *principled* revision by providing argument overlap, presenting known ideas before new ideas in sentences, and/or making explicit any implicit references.

Revisions to *Peru and Argentina* (E. Kintsch, 1990), a 6th grade text, also followed the Kintsch and van Dijk (1978) model. Four text versions were created by disrupting the macrostructure (e.g., shifting topics) and microstructure (e.g., difficult words, longer and more complex sentence patterns, fewer connectives) of the original text.

Traits of mammals from McNamara et al. (1996; Exp. 1) was a biology text for 6th to 8th grade students. The original text was locally coherent but had a list like structure at a more global level. Thus, the revision made links more explicit between subtopics and the main topic by adding information. The *Heart Disease* text (Exp. 2) was based on an entry in a science encyclopedia for school age students. The high-cohesion version examined here included revisions at the local and global levels. Local changes included replacing pronouns with noun phrases, adding descriptive elaborations and connectives, and increasing argument overlap. Global manipulations included adding topic headers and topic sentences to link each paragraph to the rest of the text and overall topic. The *Cell Division* text from McNamara (2001) was from a middle school textbook. The manipulations were similar to the changes in McNamara et al. (1996; Exp. 2).

The Voss and Silfies (1996) study included two college level texts describing series of events in two different fictional sets of countries. Text manipulations involved elaborations of causal factors related to the events. Two social studies texts used in Linderholm et al. (2000), *Mademoiselle Germaine* (easy text) and *Project X-Ray* (difficult text) were revised based on the causal network theory of comprehension (Trabasso & van den Broek, 1985). Repairs to the causal structure/organization of the texts included arranging events in temporal order, making goals of the character explicit, and repairing coherence breaks caused by inadequate explanation, multiple causality, or distant causal relations.

A text on the *Russian revolution* in Vidal-Abarca, Martinez, and Gilabert (2000) was obtained from an 8th grade history textbook. The maximally coherent version included both argument overlap revisions (see Britton & Gulgoz, 1991) and causal construction revisions (see Trabasso & van den Broek, 1985).

Cataldo and Oakhill (2000) examined narrative stories written for children, including an original, coherent version and a scrambled version with randomly reordered sentences. A text used in Lehman and Schraw (2002; Exp. 2), *The Quest for Northwest Passage*, was a historical narrative. In this analysis, we compared the original text and the maximally incoherent text. Producing an incoherent text involved presenting selected sentences in each paragraph in the altered order to reduce local coherence. Also, the temporal flow of the story was interrupted to reduce global coherence.

Results and Discussion

The 19 pairs of high and low-cohesion texts were analyzed using Coh-Metrix 1.4. ANOVAs were conducted to test for significant differences between the high and low-cohesion text versions. The significance value is p <.05 unless otherwise stated.

Descriptive and Readability Statistics

Table 1 shows the traditional readability statistics (i.e., Flesh-Kincaid Grade level; FKGL) for the low and high-cohesion texts. These statistics show that the high-cohesion texts tend to include more words, $F(1,18)=18.21$, sentences, $F(1,18)=7.09$, and words per sentence, $F(1,18)=17.08$. As a result, the FKGL index is higher, $F(1,18)=4.64$. Increasing cohesion requires adding words to fill in the conceptual gaps. Hence, grade level tends to increase because it is partially driven by the number of words per sentence. As such, grade level indices would predict that lower cohesion texts are easier than higher cohesion texts. This presumably is not true generally, and is definitely not true in the case of these particular texts.

Table 1. Text features related to readability indices

	Cohesion	
	Low	High
Words	507 (326)	673 (424)
Sentences	36.26 (19.16)	41.68 (23.35)
Words/Sentences	13.50 (3.97)	15.78 (3.73)
FKGL	7.76 (3.04)	8.35 (2.82)

Notes: standard deviations are in parentheses; FKGL is Flesch-Kincaid Grade level

Table 2. Coreference indices by cohesion (high, low) as a function of the type of index (noun, argument, stem), distance (adjacent, 2 sentences, 3 sentences, all distances) and weight (not-weighted or weighted).

Type	Dist	Wt	Low Coh	High Coh	*F*
Noun	Adj	Not	0.34 (.19)	0.53 (.16)	23
	2	Not	0.32 (.16)	0.47 (.14)	27
		Wtd	0.32 (.17)	0.49 (.15)	26
	3	Not	0.30 (.15)	0.44 (.13)	26
		Wtd	0.31 (.16)	0.47 (.14)	27
	All	Not	0.23 (.10)	0.33 (.13)	19
		Wtd	0.27 (.12)	0.40 (.13)	24
Arg	Adj	Not	0.40 (.20)	0.58 (.15)	19
	2	Not	0.38 (.16)	0.53 (.14)	21
		Wtd	0.38 (.17)	0.54 (.14)	21
	3	Not	0.36 (.15)	0.50 (.12)	22
		Wtd	0.37 (.16)	0.53 (.13)	22
	All	Not	0.28 (.10)	0.38 (.14)	17
		Wtd	0.32 (.13)	0.45 (.13)	20
Stem	Adj	Not	0.45 (.22)	0.61 (.16)	15
	2	Not	0.42 (.19)	0.56 (.15)	18
		Wtd	0.43 (.20)	0.57 (.16)	17
	3	Not	0.40 (.17)	0.53 (.15)	24
		Wtd	0.42 (.19)	0.56 (.15)	20
	All	Not	0.32 (.13)	0.41 (.15)	18
		Wtd	0.37 (.15)	0.48 (.15)	18

Note: standard deviations are in parentheses

Coreference

Coh-Metrix provides three general types of coreference indices. Noun overlap is overlap between nouns, with no deviation in form. Argument overlap is overlap between the noun in the target sentence and the same noun in singular or plural form in the previous sentence. Stem overlap is overlap from the noun to stems, regardless of word type (e.g., noun, verb, adjective). Thus, stem overlap could include overlap between *giver* in the target sentence and *giver*, *giving, or gave* in previous sentences. Both argument and stem overlap also include overlap between a pronoun and the same pronoun.

Coreference indices also vary by distance between the target sentence and coreferent sentences. *Adjacent* overlap includes only adjacent sentences. Distances of *two* or *three* sentences include the target sentence and the two or three previous sentences, respectively. *All distances* consist of the overlap between each sentence with all other sentences in the text – this is intended as a more global index of cohesion.

All coreference indices are average overlap between sentence pairs, with overlap for each pair being either 0.0 or 1.0. Weighted versions are also adjusted for distance: the weight for each pair is the inverse of the distance between two sentences.

Table 2 indicates that all of the indices showed significant differences between the high and low-cohesion texts (all p <.001). We examined whether there were differences between index type, distance, and weighting comparing the high and low-cohesion versions of the 19 texts. The mixed ANOVA included the within-text factors of cohesion (high, low), index type (noun, argument, stem), distance (all distances, 2 sentences, 3 sentences) and weight (unweighted, weighted). Adjacent indices could not be included in this analysis because a weighted version does not exist. There were main effects of cohesion, $F(1,18)=23.36$, $MSe=0.136$, index type, $F(1,18)=36.32$, $MSe=0.789$, distance, $F(1,18)=45.78$, $MSe=0.016$, and weight, $F(1,18)=41.84$, $MSe=0.004$. There was an interaction between distance and weight, $F(1,18)=42.69$, $MSe=0.001$, indicating that weighting affected the all-distance indices, $M_{unwtd}=0.32$ $M_{wtd}=0.38$, but had little effect on either two-sentence, $M_{unwtd}=0.44$, $M_{wtd}=0.46$, or three-sentence, $M_{unwtd}=0.42$ $M_{wtd}=0.44$, distances. Cohesion interacted with weight, $F(1,18)=9.11$, $MSe=0.001$, such that weighted algorithms yielded larger differences, Diff=0.144, than did unweighted algorithms, Diff=0.126. Cohesion also interacted with distance, $F(1,18)=11.25$, $MSe=0.003$, indicative of larger differences for local indices of coreference, $Diff_{2sent}=0.152$; $Diff_{3sent}=0.140$, than the all-distances algorithms, $Diff_{all}=0.110$.

In summary, the local indices (i.e., 2 and 3 sentence distances) and weighted algorithms tended to yield greater differences between the two text versions. Although cohesion did not interact with index type, $F(1,18)=2.28$, it is apparent in Table 2 that noun overlap indices yielded the most robust differences between text versions. However, all of the coreference indices successfully showed differences between the high and low-cohesion versions.

Table 3. Six LSA indices for the low and high-cohesion text versions.

	Cohesion	
LSA Index	Low	High
Sent. to Adj. Sent.	0.21 (0.11)	0.27 (0.12)
Sent. to all Sent.	0.19 (0.10)	0.24 (0.11)
Sent. to Para.	0.27 (0.14)	0.33 (0.12)
Sent. to Text.	0.34 (0.13)	0.37 (0.13)
Para. to Para.	0.36 (0.20)	0.36 (0.19)
Para. to Text	0.50 (0.19)	0.52 (0.19)

Note: standard deviations are in parentheses

Latent Semantic Analysis (LSA)

Previous studies have used LSA to measure cohesion differences (e.g., Foltz, Kintsch, Landauer, 1998). Coh-Metrix includes six types of LSA indices: adjacent sentence to sentence, sentence to all other sentences, sentence to paragraph, sentence to text, paragraph to paragraph, paragraph to text. Four of the six LSA indices showed significantly higher cohesion scores for the high as compared to the low-cohesion versions. They are: adjacent sentence to sentence, $F(1,18)=15.9$, p <.01; sentence to all sentences, $F(1,18)=9.24$; sentence to paragraph, $F(1,18)=12.28$, p<.01; and sentence to text, $F(1,18)=8.85$. The two that did not are indices of global cohesion (paragraph to paragraph, paragraph to text), which was generally not manipulated in the targeted studies.

It is notable that the LSA indices did not distinguish between the text versions as well as did the coreference indices (showing smaller effect sizes). One difference between the coreference indices and the LSA indices is that LSA is more generous in its determination of overlap. That is, a concept in a sentence is more likely to overlap with a concept in another sentence according to LSA even when there is no strict overlap in word. This trend is also observed amongst the coreference indices where the noun overlap indices tended to yielded greater differences than the stem overlap indices. Thus, the indices with the strictest indices of overlap tend to show greater differences between versions.

Connectives and Causal Cohesion

Another element of text cohesion comes from connectives. Connectives provide explicit cues to the type of relationship between ideas in a text, and thus increase text cohesion (Louwerse, 2001). Coh-Metrix provides an incidence score (occurrence per 1000 words) for four general types of connectives: causal (negative, positive), additive (negative, positive), temporal (negative, positive), and clarification. Examples of each are provided in Table 4.

Table 4. Examples of causal, additive, temporal and clarification connectives.

Connective Type	Examples
Causal – Positive	*a consequence of*, *after all*
Causal - Negative	*nevertheless*, *nonetheless*
Additive - Positive	*also*, *as well*, *further*
Additive - Negative	*anyhow*, *on the contrary*
Temporal - Positive	*suddenly*, *up to now*, *when*
Temporal - Negative	*until*, *until then*
Clarification	*that is to say*, *for example*

Table 5. Causal and connective indices for the low and high-cohesion text versions

	Cohesion	
	Low	High
Causal Indices		
Particles Inc.	21.40 (7.78)	28.57 (15.62)
Verbs Inc.	25.21 (12.25)	24.10 (10.42)
Vbs & Particles	46.61 (14.26)	52.68 (22.74)
Particle Verb Ratio	0.87 (0.39)	1.14 (0.46)
Connectives		
Causal – Pos	21.02 (7.94)	28.10 (15.57)
Causal – Neg	0.38 (0.87)	0.47 (1.09)
Additive – Pos	32.00 (10.69)	29.16 (10.16)
Additive – Neg	7.64 (7.26)	7.08 (4.11)
Temporal – Pos	10.36 (6.32)	11.70 (4.94)
Temporal – Neg	0.32 (0.99)	0.18 (0.60)
Clarification	0	0.37 (0.99)
All Connectives	69.29 (17.20)	73.26 (13.20)

Note: standard deviations are in parentheses

Coh-Metrix provides additional indices of causal cohesion by measuring the ratio of the incidence of causal particles to causal verbs (i.e., causal particles/(causal verbs+1). Causal verbs convey an action that impacts another entity such as the verb *impact*. Causal particles include causal connectives (e.g., *because*) as well as identified phrases that indicate causality such as the adverbial phrase *as a result*. This index is motivated by the assumption that causal cohesion is most relevant when the text refers to events and actions that involve causality. Coh-Metrix estimates causality in a text by the number of causal verbs conveying an action that impacts another entity. It is assumed that causal particles clarify the intended meaning of a causal verb. The notion is that a text that is entirely causally cohesive will provide one causal particle for every causal verb. If there are numerous causal verbs without causal particles, then the reader needs to infer the relationships between causal events/actions conveyed by each sentence.

The results indicate that the higher cohesion texts contained more causal particles, $F(1,18)=5.60$, and positive causal connectives, $F(1,18)=5.48$, and that the ratio of causal particles to verbs was greater, $F(1,18)=10.69$. However, there were no differences in the other types of connectives between the texts.

One consideration is that not all of the targeted studies intended to vary causal cohesion. Greater differences may emerge by targeting only those studies that explicitly included causal manipulations. Fourteen of the text pairs contained explicit causal cohesion manipulations. Five text pairs from four studies were identified that did not manipulate causal cohesion (Britton & Gulgoz, 1991; Cataldo & Oakhill, 2000; Kintsch, 1990; Lehman & Schraw, 2002). The interaction of cohesion and the type of study was reliable for only one index, the ratio of causal particles to causal verbs, $F(1,18)=13.43$, $p=.002$. This interaction indicates that there is no difference between the high and low-cohesion versions that did not explicitly vary causal cohesion ($F<2$), whereas the difference is quite large (Diff=0.41) for those that did, $F(1,13)=26.26$ ($M_{low}=0.80$, $SD_{low}=0.41$, $M_{high}=1.21$, $SD_{high}=0.51$). This result strengthens the conclusion that the ratio index is a successful proxy for causal cohesion.

Discussion

One purpose of this study was to validate Coh-Metrix (Graesser et al., 2004) as a tool to measure text cohesion. A second purpose was to provide a description of what aspects of text features were associated with changes in text cohesion, when that change was made intentionally.

The results showed that the largest differences between the text versions were found in coreference indices. Among the coreference indices, the most discriminative were the more local indices that include 2 or 3 sentences prior to the target sentence. Adjacent and all distance indices tended to yield smaller differences. There was also a trend such that noun overlap indices, the strictest measures, tended to yield larger differences than argument or stem overlap indices.

Likewise, although the LSA indices discriminated between text versions, the differences were smaller compared to the coreference indices. One reason may be because the compared texts were on the same topics. That is, the texts were high and low-cohesion versions of the same text. Given that LSA is designed to represent semantic similarity, the smaller differences shown by LSA may reflect the fact that the texts were highly similar semantically.

The text versions also differed in terms of causal particles, positive causal connectives, and the ratio of causal particles to causal verbs. Of particular interest to us was the ratio index. This study confirmed that the ratio successfully distinguished between high and low-cohesion texts, particularly when causal cohesion was manipulated in the texts. Thus, the results validate the causal ratio index.

One potential concern for this study is that low and high-cohesion texts are not in equal length. In general, increasing the cohesion of a text necessarily requires adding words; thus this has been a confound in most studies of cohesion. To somewhat alleviate that concern here, we truncated the high-cohesion texts to be equal in length to the low-cohesion texts. We found that the results and trends were generally equivalent to those that we reported here.

Overall this study validates the coreference and causal indices by showing that Coh-Metrix successfully detects cohesion manipulations intentionally made by experts in text comprehension. The high-cohesion texts were significantly different from the low-cohesion texts in conceptual overlap and causal cohesion. In contrast, more traditional indices of text difficulty, that is, the Flesch-Kincaid Grade level, incorrectly indicated that the high-cohesion text was more difficult or equally difficult, respectively. These results collectively indicate that more robust indices of text difficulty will emerge from the consideration of text cohesion.

Computational measures of cohesion are beneficial for several reasons. First, such computational indices, when compared to psychological data, may shed light on components involved in discourse processing. Second, such indices are useful in a variety of computational applications, including intelligent systems, summarization techniques, text generation, speech recognition, and question answering systems. As such, validating the cohesion indices provided in Coh-Metrix contributes to a large array of potential applications, both theoretical and practical.

Acknowledgments

This research was supported by the Institute for Education Sciences (IES R3056020018-02).

References

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (Eds.) (1993). *The CELEX Lexical Database* (CD-ROM). Univ. of Penn., PA: Linguistic Data Consortium.

Beck, I.L., McKeown, M.G., Omason, R.C., & Pople, M.T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263-277.

Beck, I.L., McKeown, M.G., Sinatra, G.M., & Loxterman, J.A. (1991). Revising social studies text from a text processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*, 251-276.

Brill, E (1995) Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics, 21,* 543-566.

Britton, B.K., & Gulgoz, S. (1991) Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-345.

Britton, B.K., Gulgoz, S., & Glynn, S. (1993). Impact of good and poor writing on learners: Research and theory. In B.K. Britton, A. Woodward, & M.R. Binkley (Eds.), *Learning from textbooks: Theory and practice* (pp.1-46). Hillsdale, NJ: Erlbaum.

Cataldo, M.G., & Oakhill, J. (2000). The effect of text organization (original vs. scrambled) on readers' ability to search for information. *Journal of Educational Psychology, 92*, 791-799.

Coltheart, M. (1981). The MRC psycholinguistic database quarterly. *Journal of Experimental Psychology*, *33A*, 497-505.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with LSA. *Discourse Processes, 25*, 285-307.

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

Graesser, A.C., McNamara, D.S., & Louwerse, M.M. (2003). What readers need to learn in order to process coherence relations in narrative and expository text. In A.P. Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford Publications.

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.

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

Kintsch W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*, 161-195.

Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104 , 211-140.

Lehman, S., & Schraw, G. (2002). The effect of text coherence on shallow and deep processing. *Journal of Educational Psychology, 94,* 738-750**.**

Linderholm, T., Everson, M.G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2001). Effects of causal text revision on more or less-skilled readers' comprehension of easy and difficult texts*.* *Cognition and Instruction, 18*, 525-556.

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

Louwerse, M.M., McCarthy, P.M., McNamara, D.S., & Graesser, A.C. (2004). Variation in language and cohesion across written and spoken registers. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 843-848). Mahwah, NJ: Erlbaum.

Loxterman, J. A., Beck, I. L., & McKeown (1994). The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly, 29*, 353-367.

McCarthy, P.M, Lewis, G.A., Dufty, D.F., & McNamara, D.S. (2006). Analyzing writing styles with Coh-Metrix. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference*, Melbourne, Florida.

McNamara, D.S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62.

McNamara, D.S., & Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better?: Interactions of texts coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction, 14*, 1-43.

Miller, G.A, Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *Five papers on WordNet*. CS Laboratory, Princeton University, No. 43.

Sekine, S., & Grishman, R. (1995). A corpus-based probabilistic grammar with only two non-terminals. In *Proceedings of the Fourth International Workshop on Parsing Technologies*.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*, 595-611.

Vidal-Abarca, E., Martinez, G., & Gilabert, R. (2000). Two procedures to improve instructional text: Effects on memory and learning. *Journal of Educational Psychology,* 92, 107-116.

Voss, J.F., & Silfies, L.N (1996). Learning from history text: the interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction, 14*, 45-68.