Toward a New Readability: A Mixed Model Approach

Scott A. Crossley
(scrossley@mail.psyc.memphis.edu)
Department of English, Mississippi State University
Starkville, MS 39759

David F. Dufty
(ddufty@memphis.edu)
Department of Psychology, University of Memphis
Memphis. TN 38152

Philip M. McCarthy
(pmmccrth@memphis.edu)
Department of Psychology, University of Memphis
Memphis. TN 38152

Danielle S. McNamara
(d.mcnamara@mail.psyc.memphis.edu)
Department of Psychology, University of Memphis
Memphis. TN 38152

Abstract

This study is a preliminary examination into the use of Coh-Metrix, a computational tool that measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis, as a means of measuring English text readability. The study uses 3 Coh-Metrix variables to analyze 32 academic reading texts and their corresponding readability scores. The results show that two indices, one measuring lexical co-referentiality and one measuring word frequency, mixed with an estimate of syntactic complexity, yield a prediction of reading difficulty that is similar to traditional readability formulas. The study demonstrates that Coh-Metrix variables can contribute to a readability prediction that better reflects the psycholinguistic factors of reading comprehension.

Keywords: Readability; Corpus Linguistics; Cognitive Processing; Computational Linguistics; Discourse Analysis

Introduction

This study is an exploratory examination into the use of Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) as an improved means of measuring text readability. While traditional readability formulas such as Flesch Reading Ease (1948) and Flesch-Kincaid (1975) have been widely accepted by the reading research community, they have also been widely criticized by cognitive researchers for their inability to take into account textbase processing, situation levels (Kintsch, et al., 1990; McNamara et al., 1996) and cohesion (Graesser et al., 2004, McNamara et al., 1996). Coh-Metrix, however, offers the prospect of addressing the limitations of conventional readability measures by providing detailed analyses of language by integrating lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics (Jurafsky & Martin, 2000). In reference to cohesion indices, Coh-Metrix also analyzes co-referential cohesion, causal cohesion, density of connectives, Latent Semantic Analysis metrics, and syntactic complexity. Since Coh-Metrix considers textbase processing and cohesion, it is well suited to address many of the criticisms of traditional readability formulas.

Classic Readability

Providing students with texts that are accessible and well matched to reader abilities has always been a challenge for educators. A solution to this problem has been the creation and use of readability formulas. Since 1920 more than 50 readability formulas have been produced in the hopes of providing tools to measure text difficulty more accurately and efficaciously. Additionally, it was hoped these formulas would allow for a greater understanding of optimal text readability.

The majority of these readability formulas are based on factors that represent two broad aspects of comprehension difficulty: lexical or semantic features and sentence or syntactic complexity (Chall & Dale, 1995). According to Chall and Dale (1995), formulas that depend on these variables are successful because they are related to text simplification. For instance, when a text is written for a beginning reading audience, the text generally contains more frequent words and shorter sentences. Thus, measuring the word frequency and sentence length of a text should provide a basis for understanding how readable it is.

However, traditional readability formulas are often not based on any theory of reading or reading comprehension, but rather on empirical correlations. Therefore, their soundness is strictly predictive and they are often accused of having weak construct validity. Regardless, a number of classic validation studies have found the formulas' predictive validity to be consistently high, correlating with observed difficulty in the r = .8 range and above (Chall, 1958; Chall & Dale, 1995; Fry, 1989).

While the predictive validity of these measures seems strong, they are generally based on traditional student populations reading academic or instructional texts. This has led many proponents of readability formulas to caution against their use with literary or technical texts, or texts written to the formulas. However, the draw of readability formulas' simple, mechanical assessments has led to their widespread use for assessing all sorts of texts for a wide variety of readers and reading situations beyond those for which the formulas were invented. The widespread use of traditional formulas in spite of restricted validity has inclined many researchers within the field of discourse processing to regard them with reservation (Bruce, Rubin & Starr, 1981; Bruce & Rubin, 1988; Davison & Kantor, 1982; Rubin, 1985; Smith, 1988).

The rise of cognitive models of reading has underscored not only the limitations of the traditional formulas but also the need for a measure that accounts for discourse-specific factors such as textbase and situation level processing (Kintsch et al., 1990; McNamara et al., 1996). A more inclusive assessment of text comprehensibility must go deeper than surface readability features and explain how that learner interacts with a text (Kintsch, 1994; McNamara et al., 1996; Miller & Kintsch, 1980). Most importantly for the purpose of this study, such assessment must include a measure of text cohesion, which is vital to text processing (Gernsbacher, 1997; McNamara, 2001; McNamara et al., 1996).

The limitations of classic readability formulas led to new readability theories based on cognitive and structural variables. Much of the ground work for this approach was conducted by Kintsch and Vipond (1977), who were critical of classic readability formulas in that the formulas were a-theoretical and based solely on text factors. In suggesting new variables for testing readability based on conceptuality, Kintsch and Vipond advocated the use of propositions (defined as arguments attached to predicates). Using propositional density along with classic readability measures (word frequency and sentence length) Kintsch and Vipond reported a multiple correlation of .97 between these variables and the reading difficulty scores of a limited data set. In later work (Kintsch et al., 1993), this approach was expanded to relate propositions to coherence with the idea that as the coherence of a text improved, so did the readability.

Coh-Metrix

Recent advances in various disciplines have made it possible to computationally investigate various measures of text and language comprehension that supercede surface components of language and instead explore deeper, more global attributes of language. The various disciplines and approaches that have made this approach possible include computational linguistics, corpus linguistics, information extraction, information retrieval, and discourse processing. Taken together, the improvements in these fields have allowed the analysis of many deep level factors of textual coherence to be automated, allowing for more accurate and detailed analyses of language to take place (Graesser et al., 2004).

A synthesis of the advances in these areas has been achieved in Coh-Metrix, a computational tool developed at the University of Memphis that measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis. This tool was designed with the goal of improving reading comprehension in classrooms by providing a means to improve textbook writing and to more appropriately match textbooks to the intended students (Graesser et al., 2004).

Corpus

A corpus of reading texts was selected to test the hypothesis that linguistic variables related to cognitive processing and cohesion could predict text readability. The corpus we chose was the Bormuth (1971) passage set. The Bormuth passage set is comprised of 32 academic passages that include corresponding readability scores. The passage set includes texts taken from school instructional material and includes passages from biology, chemistry, civics, current affairs, economics, geography, history, literature, mathematics, and physics. The Bormuth readabiltity scores are based on the reading difficulty scores of 285 elementary and high school students from the grades of 3rd to 12th. Bormuth used cloze scoring procedures on his 32 academic passages to test for reading difficulty. His cloze procedure deleted every 5th word of the text and the participants were expected to correctly deduce the correct word (or synonym).

The selection of these passages as the foundation for this study is based not only on the seminal work conducted by Bormuth (1971) with this passage set, but also on the work done by Chall and Dale (1995) who selected the Bormuth set to construct a new readability formula. The advantages of the Bormuth passages, as stated by Chall and Dale (1995), are based primarily on Bormuth's use of a cloze criterion as well as the fact that the Bormuth passage set was constructed using variable text content and text difficulty. Additionally, the decision by Chall and Dale (1995) to use Bormuth's passages rather than other passage sets was made after extensive evaluation and comparison of the passage characteristics and cross-validation of their readability scores to other passage sets (e.g. MacGinitie & Tretiak, 1971; Miller & Coleman, 1967; Caylor et al., 1973).

Readability Formulas and the Bormuth Passages

Bormuth (1969) used the mean cloze scores from his passage sets to create a readability formula that was based on the number of letters per word, the number of Dale-Chall words per total words (based on the 1948 Dale-Chall word list), and the number of words per sentence. Using Bormuth's 1969 formula and an updated Dale word list (from 1983), a new multiple regression comparing the text features of the Bormuth passage set and its corresponding reading difficulty scores reported a multiple correlation of .961 with an adjusted R^2 of .915 between the formula and the cloze scores.

Chall and Dale (1995) also formulated a new readability formula based on the Bormuth passage set. Their formula was designed using Dale's updated word list (1983), the modification of rules for unfamiliar word counts, and a simplified equation. Their final readability formula was based on three variables: number of frequent words (based on the 1983 Dale 3,000 words known by students in grade 4), number of unfamiliar words (those words not in the Dale 3,000 words), and number of sentences. These variables were rendered into a readability formula based on semantic difficulty and syntactic difficulty. Using Chall and Dale's readability formula, a new multiple regression comparing the text features of the Bormuth passage set and its corresponding reading difficulty scores reported a multiple correlation of .956 with a corresponding adjusted R^2 of .907.

Purpose

The purpose of this study is to analyze how well Coh-Metrix variables predict text readability. To accomplish this goal, a readability formula based on Coh-Metrix variables will be examined so that we can compare our results to previous ones using traditional formulas. The number of passages available (the 32 passages of the Bormuth corpus in this case) limited the number of variables that could be used without over-fitting the model. At a minimum, 10 cases of data for each predictor are considered sufficient (with conservative models using 15 to 20). Accordingly, 3 independent variables from Coh-Metrix were selected to analyze the Bormuth passages. These indices were selected based on past research pointing to syntactic complexity (Bormuth, 1969; Chall & Dale, 1995; Kintsch, 1979), word difficulty (Haberlandt & Graesser, 1985; Just & Carpenter, 1980), and co-referentiality (Kintsch & van Dijk, 1978; Rashotte & Torgesen, 1985) as important for text difficulty and readability.

Estimate of Syntactic Complexity In defining syntactic complexity, we assume that sentences with difficult syntactic composition are structurally dense, syntactically ambiguous, or ungrammatical (Graesser et al., 2004). An estimate of syntactic complexity was included as a predictor of readability because multiple reading theorists have affirmed its importance in text readability and most readability formulas have included some measure of syntactic complexity (e.g. Bormuth, 1969; Chall & Dale, 1995; Kintsch, 1979). Because longer sentences are a rough estimate of the number of propositions contained, the variable *number of words per sentence* was selected for this study.

Co-referentiality Coh-Metrix currently measures four forms of lexical co-reference between sentences: noun overlap, argument overlap, stem overlap, and content word overlap. Lexical co-referentiality was chosen as a predictor of readability because overlapping vocabulary has been found to be an important aspect in processing texts and can lead to reading gains and faster reading rates (Rashotte & Torgesen, 1985). It has been shown to aid in text comprehension and reading speed (Kintsch & van Dijk, 1978). For this, *argument overlap* was chosen to represent lexical co-referentiality. *Argument overlap* was selected as it is the most robust measure of lexical co-referentiality in that it measures how often two sentences share common arguments (nouns, pronouns, and noun phrases).

Word Frequency Coh-Metrix calculates word frequency information through CELEX frequency scores. CELEX (Baayen, Piepenbrock, & Gulikers, 1995) is the primary frequency count in Coh-Metrix and consists of frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9 million-word corpus. Word frequency is considered important to readability because frequent words are normally read more rapidly and understood better than infrequent words (Haberlandt & Graesser, 1985; Just & Carpenter, 1980). Additionally, researchers argue that quick, accurate, and automatic recognition of words help skilled readers in processing text (Silberstein, 1994).

Statistical Analysis

To calculate the readability of the Bormuth passage set, the three selected variables were used as predictors in a training set and a multiple regression equation with the 32 observed mean reading scores as the dependent variable was conducted. The statistical analyses in this part of the study included descriptive statistics for the predictors and the dependent variable. To assess the assumption of independent errors caused by outliers, Durbin-Watson statistics were conducted. In order to assess the assumption of multicollinearity, coefficient analyses were conducted.

In perfect circumstances, a researcher will have enough data available to create separate training and testing sets and use the training set to create predictors and the testing set to calculate how well those predictors function independently. Historically, most readability studies have been statistically imperfect in that they have based their findings on the results of a single training set (i.e. Bormuth, 1971; Chall & Dale; 1995). While performance on a single training set allows conclusions regarding how well variables predict the difficulty of the texts in that set, those conclusions may not be extendible to an independent test set (Whitten & Frank, 2005). The problem, of course, is the difficulty of creating sufficiently large data sets.

With a limited data set, as in this study, the question becomes how to make the most of the available data. To address this problem, this study considers three approaches to cross-validation. The first two approaches are simple estimates of cross-validation: adjusted R^2 data is decided on. Once the number of folds has been decided, each is used for testing and training in turn. In *n*-fold cross-validation, which will be used in this study, the *n* refers to the number of instances in the data set. Each instance in turn is left out and the remaining instances are used as the training set (in this case 31) and the accuracy of the model is tested on the model's ability to predict the one remaining instance. In the case of the data at hand, predictors were taken from the training set and used in a regression analysis of the first 31 texts. The B values and the constant from that analysis were used to predict the value of the 32nd text. This text then became the first in our testing set. This process was repeated for all 32 texts, creating a testing set. The predicted values were then correlated with the actual values (the mean cloze scores) to test the model for performance on an independent testing set. All of these models (adjusted R^2, SURE estimate, and *n*-fold cross-validation) are important, because if a model can be generalized, then it is likely capable of accurately predicting the same outcome variable from the same set of predictors in a different text group (Field, 2005).

Results

Pearson Correlations

When comparing the three selected variables to the Bormuth mean cloze scores, significant correlations were reported for all indices. Correlations between the Bormuth mean cloze scores and the number of words in a sentence were significant ($N = 32$, $r = -0.908$, $p < 0.001$), as was the CELEX word frequency measures ($N = 32$, $r = 0.826$, $p < 0.05$) and the argument overlap measure ($N = 32$, $r = 0.686$, $p < 0.001$).

Multiple Regression Analysis

In order to estimate the degree to which the chosen independent variables were collectively related to predicting the difficulty of the Bormuth passages, the dependent and independent variables were investigated using a multiple regression analysis. Descriptive statistics for the dependent and independent variables are presented in Table 1, and results for the regression analysis are in Table 2. The variables were also checked for outliers and multicollinearity. For outliers, the Durbin-Watson statistic was 2.672, which is less than 3 and greater than 1, implies that there are no independent errors caused by residuals. Coefficients were checked for tolerance with all tolerance levels well beyond the .2 threshold, indicating that the model data did not suffer from multicollinearity.

The results of the forced entry multiple regression analysis indicate that the combination of syntactic complexity scores (words per sentence), CELEX frequency scores, and argument overlap scores taken together produce a multiple correlation .954 and a corresponding R^2 of .910. This signifies that the combination of the three variables alone accounts for 91% of the variance in the performance of the students on the 32 cloze tests based on the Bormuth passages.

Table 1:Descriptive Statistics

Variable	Mean	Std. Deviation	N
Predicted Mean Cloze Scores	458.209	54.699	32
Predictor			
Words per Sentence	15.872	5.154	32
CELEX Frequency	1.184	0.416	32
Argument Overlap	0.206	0.121	32

Cross validation

Two estimates of cross-validation were conducted. The adjusted R^2 for the regression analysis was .90 and the Stein's Unbiased Risk Estimate (SURE) was .89. Considering that these two estimates are very similar to the observed R^2 (.91), the estimates seem to support that the cross-validity of the model is good. As these are only estimates, though, a *n*-fold cross-validation model was constructed. A correlation between the predicted values of the testing set and the actual values revealed a significant correlation ($N = 32$, $r = 0.94$, $p < 0.001$), demonstrating that the predictors perform well on an independent testing set.

Discussion

A combination of three variables from Coh-Metrix predicted 91% of variance in cloze scores from the Bormuth (1971) dataset. Readability formulas based on Chall and Dale (1995) and Bormuth (1969) achieved similar results. Comparison of the adjusted R^2 value between this study and earlier studies seems unwise because of the very high correlations involved. The fact that diverse methods of measuring text difficulty all achieve correlations of .9 or above indicate that a ceiling effect is present. It seems to be the case that a variety of measures of text difficulty will achieve very high

Table 2: Regression Analysis of Three Independent Variables Predicting Reading Difficulty 2

Variable	Unstandardized Coefficient	Standardized Coefficient	Standard Error	*T*	*P*
Words per Sentence	-5.896	-0.556	.973	-6.061	0.000
Average Sentence Word Frequency	48.554	0.370	10.381	4.677	0.000
Argument Overlap	65.869	0.146	33.854	1.946	0.062

correlations on this dataset. Because of this, we can conclude that the measures used here are effective measures of text difficulty, as are readability measures such as the Dale-Chall readability score. We cannot, however, conclusively determine which of these is more effective: such a task would require a larger dataset with less variability between texts.

Reading researchers have been arguing for some time that measures of text difficulty are needed that directly take into account cognitive processing load, the individual cognitive aptitude of the learner, and how that learner interacts with a text (Kintsch, 1994; McNamara et al., 1996; Miller & Kintsch, 1980; Gernsbacher, 1997; McNamara, 2001; McNamara & Kintsch, 1996). The current study, which demonstrates that cognitively-inspired indices also provide effective measures of reading difficulty, is a step in that direction. In particular, this study addresses two salient concerns about readability formulas raised by Chall and Dale (1995). First, the formula relies on both traditional factors as well as cognitive and structural factors, but not on one approach alone. Second, the formula is not difficult to apply or more time consuming because it is automated.

While the foundations of this study were classic readability studies such as Bormuth's (1971) and Chall and Dale's (1995), the approaches are dissimilar as they are partially based on current theories of cognition. Using Bormuth's mean cloze scores, the three Coh-Metrix variables, one employing lexical co-referentiality, one estimating syntactic complexity, and one measuring word frequency, yielded an accurate prediction of reading difficulty of Bormuth's classic passage. The results are encouraging because the analysis incorporated variables that are directly related to cognitive processes of reading and show strong correlations to text readability using variables that are not all tied to superficial aspects of reading, as past readability formulas have.

Moreover, the readability formula presented here is exploratory and only considers three indices out of the hundred or so available through Coh-Metrix. These additional indices will allow future researchers options for incorporating measurements of cohesion such as anaphoric resolution, temporal and spatial information, psycholinguistic measurement of word information, and indices of causality, to name but a few. Traditional readability formulas such as Bormuth's and Dale and Chall's do not allow for such an extension nor the opportunity for deeper level analysis of text language features.

Conclusion

This study is limited by the small number of passages in the Bormuth dataset, and by the high correlations obtained using the dataset, which suggest a ceiling effect. Furthermore, the passages are from the genre of academic writing. This genre has many unique characteristics, such as the requirement for referencing, objectivity, and other conventions. This may limit its generalizability as a testbed for reading formulas, especially for less formal, genres such as children's or adolescent fiction. To determine which text measure is the best from several competing measures (such as Coh-Metrix variables, the Dale-Chall readability formula, etc.) a larger study would need to be conducted, using more passages, and choosing passages from genres more relevant to general reading. In addition, the Bormuth passage set used cloze scores as its readability criteria. Cloze scoring, by its nature, appears connected to sentence length and word frequency because excised frequent words would be easier to estimate and shorter sentences would allow for the inference of words based on limited part of speech likelihood. Thus passage difficulty based on cloze scoring likely correlates highly with readability formulas that measure these variables such as Flesch Kincaid (1975) and Flesch Reading Ease (1948). Future studies that consider more robust cognitive variables would likely benefit from readability assessments based on recall or comprehension scores and not cloze scores. Future work should also consider participants' background knowledge and account for how this knowledge interacts with text readability (McNamara et al., 1996).

While much work remains to be done, the current study contributes to the field of text readability by demonstrating that a synthesis of traditional readability measures and indices of cohesion is a viable method for evaluating text difficulty. This work has immediate transfer potential in that it allows for more theoretically grounded approaches to readability that are easily accessible and immediately applicable. As such, this study provides educators with another tool from which to select appropriate texts to match the needs of their students.

Acknowledgments

The research was supported in part by the Institute for Education Sciences (IES R305G020018-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

References

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

Bruce, B., & Rubin, A. (1988). Readability formulas: Matching tool and task. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 5-22). Hillsdale, New Jersey: Erlbaum.

Bruce, B., Rubin, A., & Starr, K. (1981). Why readability formulas fail. Reading Education Report No. 28.

Urbana: University of Illinois Center for the Study of Reading. (ERIC Doc. No. ED 205 915)

Bormuth, J. R. (1969). *Development of readability analyses*. Final Report, Project No. 7-0052, Contract No. 1, OEC-3-7-070052-0326. Washington, DC: U. S. Office of Education.

Bormuth, J.R. (1971). *Development of standards of readability: Toward a rational criterion of passage performance.* Final report, U.S. Office of Education, Project No. 9-0237. Chicago: University of Chicago.

Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. (1973). *Methodologies for determining reading requirements of military occupational specialties*. Technical report No. 73-5. Alexandria, Virginia: Human Resources Research Organization.

Chall, J. (1958). Readability: An appraisal of research and applications. Columbus, Ohio: Ohio State University Press

Chall, J. & Dale, E. (1995). *Readability revisited: The New Dale-Chall Readability Formula*. Cambridge, Massachusetts: Brookline Books.

Davison, A., & Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly. 17,* 187-209.

Field, A. (2005). *Discovering Statistics Using SPSS*. London: Sage Publications, Ltd.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

Fry, E. (1989). Reading formulas: Maligned but valid. *Journal of Reading, 32*, 292-297.

Gernsbacher, M. (1997). Coherence cues mapping during comprehension. In Costermans, J., and Fayol, M., (Eds.), *Processing Interclausal Relationships. Studies in the Production and Comprehension of Text* (3-22). Mahwah, New Jersey: Erlbaum.

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

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.

Jurafsky, D., & Martin, J.H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kincaid, J.P., Fishburne, R.P., Rogers, R. L., & Chissom, B.S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel.* Research Branch Report 8-75, Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN

Kintsch, W. (1979). On Modeling Comprehension. *Educational Psychologist, 14*, 3-14.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist , 49*, 294-303.

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

Kintsch, W., Britton, B.K., Fletcher, C.R., Mannes, S.M., Nathan, M.J. (1993). A comprehension-based approach to learning and understanding. *The psychology of learning and motivation, 30*, 165-214.

Kintsch, W. & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In LG. Nilsson (Ed.) *Perspectives on Memory Research* (pp. 329-366). Hillsdale, New Jersey: Erlbaum.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*, 133–159.

MacGinitie, W. & Tretiak, R. (1971). Sentence depth measures as predictors of reading difficulty. *Reading Research Quarterly, 6,* 364-376.

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

Miller, G.R. & Coleman, E. B. (1967). A set of thirty-six prose passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior, 6*, 851-854.

Miller, J.R., & Kintsch,W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Behavior: Human Learning and Memory, 6*, 335-354.

Rashotte, C.A. & Torgesen, J.K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly, 20*, 180-188.

Rubin, A. (1985). How useful are readability formulas? In J. Osborn, P. T. Wilson, & R. C. Anderson (Eds.), *Reading education: Foundations for a literate America* (pp. 61-77). Lexington, MA: Lexington Books.

Smith, F. (1988). *Understanding reading: a psycholinguistic analysis of reading and learning to read*. Hillsdale, New Jersey: Erlbaum.

Silberstein, S. (1994). *Techniques and resources in teaching reading*. New York: Oxford University Press.

Whitten, I.A., & Frank, E. (2005). *Data Mining*. San Francisco: Elsevier.