Assigning Grade Levels to Textbooks: Is it just Readability?

David F. Dufty (d.dufty@mail.psyc.memphis.edu)

Arthur C. Graesser (a-graesser@memphis.edu)

Max M. Louwerse (max@mail.psyc.memphis.edu)

Danielle S. McNamara (d.mcnamara@mail.psyc.memphis.edu)

Department of Psychology / Institute for Intelligent Systems
Memphis, TN 38152 USA

Abstract

We evaluated the effectiveness of new indices of text cohesion to determine the appropriate human assigned grade level of a text. In particular, we investigated the efficacy of automated text indices produced by the online tool Coh-Metrix in predicting the grade level assigned by publishers to their own textbooks. To do this, we sampled 311 school textbooks from a large database, choosing roughly equal numbers of science, narrative, and social science texts. Publisher-assigned grade levels were found to be moderately predictable by traditional approaches such as the Flesch-Kincaid Grade Level. Prediction of grade level was significantly improved by the inclusion of cohesion indices obtained by Coh-Metrix. Implications for the improvement of textbook selection are discussed.

Introduction

The problem of providing material of appropriate difficulty to a student is an old one, with a venerable research tradition. Selecting the right textbook for a student is perhaps more achievable in a personalized, one-on-one, tutoring context, because the teacher can accurately assess the learning needs of the student. The task is considerably more difficult when such context is lacking, such as when a school is planning a curriculum or when a publisher produces a new textbook. There, the writer, editor, or decision-maker is removed from the learner and must make an educated guess as to the learners' needs. And unlike the one-on-one tutor, a decision about a book by a school board, an education department, or a publisher is not made for a single student, but for a class or even hundreds of classes at once.

In practice, the decision to assign a particular textbook to a particular curriculum (and therefore, to the students enrolled in that curriculum) is a complex one. Schools, often via a school board, decide on priorities and subject areas. Teachers frequently have input into the process, contributing knowledge about the needs of the particular student population. Legislators impose constraints according to the wishes of the state. Publishers also play a role in providing the available books for consideration.

As a guide in this process, books are commonly assigned a grade level for which they are considered appropriate. This may be a precise level or a range of grade levels. The manner in which such a grade level is assigned, either by a publisher, an educational institution, or a third party, has been the subject of much research and debate (e.g., Chall, Conard, & Harris-Sharples, 1991; DuBay, 2004), and is the central issue of this paper.

Despite the existence of numerous systems for assigning grade level, in practice this decision comes down to human judgment, with factors such as the author's intentions, the needs of the education system, and the opinion of teachers, educators and publishers, all playing a part. While editors may pay some attention to automated scores, such scores are but one of many factors that are taken into consideration. The task of automated grade level detection is to predict the judgment of editors, not the other way around.

Traditional approaches. There are at least 200 readability measures for texts, all of which primarily measure two aspects of the sentence: word difficulty, usually through word length or syllable count, and sentence difficulty, typically through sentence length (DuBay, 2004). One of the oldest and best-known methods for determining the appropriate grade level of a textbook is the Flesch-Kincaid Grade Level formula (Klare, 1974-1975). Grade level is based on a formula that computes the average number of syllables per word and the average length of all the sentences. The formula is computed as:

$$FK = (.39 \times ASL) + (11.8 \times ASW) - 15.59$$

FK is the Flesch-Kincaid Grade Level, ASL is the average sentence length (the number of words divided by the number of sentences), and ASW is the average number of syllables per word (the number of syllables divided by the number of words).

This formula is elegant in its simplicity. It takes superficial characteristics of the text and provides an index of difficulty, expressed as a grade level. Furthermore, it is effective. Sentence length correlates with variables that impact the effort required to read the sentence such as syntactic complexity. Number of syllables gives an indicator

of the length of the words, which is correlated inversely with word frequency and affects reading difficulty (Zipf, 1949). We would not expect to see many four-syllable words in a kindergarten storybook. Conversely, we would not expect strings of three- or four-word sentences in a senior biology text.

Modern approaches. The Flesch-Kincaid Grade Level and related measures such as Flesch Reading Ease have well-known shortcomings, and for this reason there has been much activity in developing better methods for assigning grade level. One of the advantages of the Flesch-Kincaid Grade Level is that number of syllables and sentence length are easy to measure and correlate with important constructs such as word difficulty and sentence complexity. However, that ease of calculation is less of an issue today than it was previously due to advances in computational linguistics. Aspects of a text such as word frequency, syntactic complexity, and many other indices (some of which will be explored below) were prohibitively difficult to calculate 30 years ago, but can now be computed effortlessly.

Stenner et al. (1987) noted two important components of a text that need to be evaluated to determine difficulty. These are the likelihood that the reader knows the words in the text (which can be approximated by word frequency) and the syntactic complexity of the sentences (which can be approximated by sentence length). Given these insights, Stenner and colleagues developed the lexile measure, which gives a central role to word frequency when determining difficulty. The measure also includes sentence length, which has some correlation with syntactic complexity. The lexile measure is a measure of text difficulty that incorporates comprehension data and statistical regression analyses (Stenner, 1996).

The Degrees of Reading Power (DRP, Koslin, et al;, 1987) uses a modified version of the Bormuth cloze readability formula (DuBay, 2004). It uses the number of words from a list of 3000 words known to most fourth graders, the average sentence length, and the average number of letters per word. The DRP has been adopted by the College Board, and has been found to be a powerful predictor of difficulty.

How publishers assign grade levels to texts. Given the availability of easy-to-compute readability scores for texts, it is expected that publishers would take these into account when designing and releasing textbooks. However, there have been criticisms of the effectiveness of readability formulas at giving an accurate estimate of ease of comprehension (DuBay, 2004).

Chall, Conard, and Harris-Sharples (1991) conducted a survey of 34 American textbook publishers and their practices in textbook development. They found that 90% of publishers take some kind of readability formula into account when determining the grade level at which the textbook should be set. However, they also found that publishers did not rely on readability formulas alone, but that there was widespread reliance on vocabulary lists and the concepts covered in the book, as well as authors' and editor's intuitions. Ninety-two percent of publishers consulted school personnel, and over half used specialists and consultants. Likewise, Stenner et al. (1987) reported that an informal survey indicated that basal text publishers invariably respond that readability measures were not used in assigning the grade level of basal readers.

Despite extensive and varied evaluation methods, the process could clearly benefit from improvements in assignments of both difficulty and grade level to textbooks. Hubisz (2001) documents a range of errors in school science texts, such as the use of unfamiliar words, errors of fact, incorrect diagrams, and people and concepts that either are not introduced or are introduced incorrectly.

Beyond readability. Despite their strengths, there are theoretical shortcomings with using readability formulas to evaluate texts. In particular, the emphasis on shallow features like word length and sentence length means that it cannot capture deep, structural properties of the text that reflect cohesion. Cohesion is the extent to which the ideas in the text are expressed clearly and relate to one another in a systematic fashion, as opposed to providing a confusing jumble of information. Several studies have established a link between cohesion and comprehension. For example, higher cohesion has been found to facilitate comprehension and recall (Beck et al., 1991). For example, Ozuru et al. (2005) reported that students comprehended high cohesion biology texts better than low cohesion biology texts. Linderholm et al. (2000) reported that improving the causal cohesion of a text improved recall.

Nevertheless, high cohesion is not always better. The optimal level of cohesion depends on the knowledge level of the reader. McNamara and colleagues (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996) found that high knowledge readers performed better on comprehension tests after reading low cohesion texts as compared to high cohesion texts. In contrast, low knowledge readers benefited from high cohesion texts, as most researchers would expect. These results clearly indicate that there is a complex and subtle relationship between the characteristics of the text and the characteristics of the reader. Specifically, cohesion gaps may force a high knowledge reader to process the text more deeply, resulting in improved comprehension and recall.

Recently, much of modern knowledge about the structure of texts has been implemented in an on-line tool, Coh-Metrix (Graesser et al, 2004; McNamara et al., 2005). Coh-Metrix takes as input any text given by a user, and outputs around 250 indices of text difficulty and text cohesion. Coh-Metrix indices have been found to be useful in predicting comprehension (Ozuru et al., 2005). They have also been applied to a variety of problems such as determining text genre (Louwerse et al., 2004), detecting authorship (McCarthy et al., 2006) and detecting differences in publications over time (Bruss, Albers, & McNamara, 2004).

Given these advances, it is timely to revisit the question of determining the grade level of a textbook. Readability measures have been shown to be powerful predictors of text difficulty. However, there is the potential for improvement, especially in measurement of structural properties of the text such as cohesion.

Clearly, it is not feasible to include all indices provided by Coh-Metrix in a regression analysis such as the one reported here. For this reason, a handful of theoretically motivated variables were chosen as potential predictors of text grade level. Similarly, it is not possible to use every readability formula as a benchmark, so a single formula, Flesch-Kincaid Grade Level, was chosen as a representative readability index.

Method

Our goal was to examine how well new, automated indices of cohesion estimated the grade level of textbooks in comparison to more traditional indices such as the Flesch-Kincaid Grade Level. To this end, we selected a sample of textbooks from an electronic corpus of about 2400 school textbooks gathered by MetaMetrics Inc.

We sampled extracts of up to 5000 words from 311 textbooks, distributed across all grade levels from K through 12, and across three genres: narrative, science, and social science. Every book had a grade level assigned to it by the publisher at the time the book was printed. Where this grade level was a range (such as *grades 3 to 5*), for the purposes of our analysis, we assigned the middle of the range as the official grade. Books were then classified into four grade categories: kindergarten to grade 3, grades 4 to 6, grades 7 to 9, and grades 10 to 12.

We examined the accuracy with which Flesch-Kincaid Grade Level estimated the grade level assigned by the publisher of the book, and whether this could be improved with newer, automated indices of difficulty and cohesion. To do this, we analyzed all of the texts using Coh-Metrix.

Results

The correlation between Flesch-Kincaid Grade Level and Publisher-Assigned Grade Level was r=.77 (R^2 = .60). Thus, 60% of the variance in Publisher-Assigned Grade Level was accounted for by Flesch-Kincaid Grade Level. This result indicates that Flesch-Kincaid Grade Level was operating as a reasonable approximation to the publisher-assigned grade level. Nevertheless, if the publishers had used only the Flesch-Kincaid Grade Level formula to assign grades, we would have expected a correlation close to 1.

Another variable that is used in some readability formulas is word frequency. While word frequency is negatively correlated with word length, these variables are known to have independent effects on word reading times (Haberlandt & Graesser, 1985). We performed a second regression with two variables, Flesch-Kincaid Grade Level and average word frequency, as predictors of the Publisher Assigned Grade Level. This produced an R^2 value of .61, demonstrating that the inclusion of word frequency explained a mere additional 1% of variance in assigned grade level. This increase was significant, but modest.

In our third analysis, we included three indices of text cohesion from the online tool Coh-Metrix (Version 1.4). The first measure was argument overlap between adjacent sentences. Argument overlap is the number of nouns or noun stems that two sentences have in common (Kintsch & van Dijk, 1978), and is generally considered to be an indicator of the degree to which sentences cover the same material. A low score on this measure would indicate that each sentence does not relate well to the sentence before it.

The second measure was the Latent Semantic Analysis (LSA). LSA is a mathematical technique that computes the similarity as a cosine between two sentences based on the similarity of the meanings of the words in the sentence, as computed as a vector in higher dimensional space (Foltz, Kintsch, & Landauer, 1998; Landauer & Dumais, 1997). The particular index was the mean similarity between all pairs of sentences; similarity was measured as a cosine between sentence vectors. This was a global measure of cohesion - a low score indicates that the various segments of the text were dissimilar to one another, suggesting that there is no clear thematic thread through the text.

The third measure was the number of causal verbs per 1000 words of text. Causal verbs are signals of the causal mental models and causal networks contained in the text, which affect comprehension (van den Broek & Trabasso, 1985). Because causal verbs are an index of causal information in the text we expect that higher-level texts with more sophisticated causal concepts to contain more causal verbs.

The fourth measure was the number of causal particles in the text. Causal particles are causal signal words and phrases, such as *in order to*, *so that* and *because*. The presence of causal particles is an indicator that the causal events in the text are coherent and explicit (Graesser et al., 2004). Therefore, to the extent that there are causal verbs, the difficulty of the text can be reduced by explicitly linking the ideas through connectives. The role of causal particles in clarifying text is dependent on the amount of causal information that needs to be clarified. Thus, theoretically, there are two forces at work on the incidence of causal particles. On the one hand, easier texts at lower grades should spell out any causal relationships, rather than leaving them to the reader to infer. On the other hand, higher grade textbooks should have more causal information in need of explanation. Therefore, we made no predictions regarding the relationship between causal particles and grade level.

As texts increase in sophistication, more abstract concepts are represented. We believed that this may be captured by the incidence of abstract nouns. Hence, abstract noun incidence per 1000 words was included as a predictor.

Similarly, we expected there to be a higher ratio of pronouns to noun phrases as writers assume greater ability

on the part of their audience to understand anaphoric references and referential chains. Hence, ratio of pronouns to noun phrases was included as a predictor.

Finally, conditionals may indicate complex reasoning. For example, two clauses joined by the word *if* indicate a logical relationship between the clauses. Since we expected more complex reasoning at higher grade levels, incidence of conditionals was included as a predictor.

The bivariate correlations between the variables and grade level are shown in Table 1. Many of the indices had highly significant correlations with grade level. Flesch-Kincaid is known to correlate with grade level.

However, indices of cohesion such as LSA and argument overlap and incidence of conditionals also capture changes in texts across different grade levels. As the grade level increases, cohesion, as measured by both these indicators, decreases. This shows that the textbooks at higher grade levels have lower cohesion, and therefore may be more difficult to read, particularly for low-knowledge readers. Causal verbs increased with increasing grade level in keeping with our prediction, demonstrating that higher grade level textbooks contain more causal information. Causal particles had no correlation with grade level.

Table 1
Bivariate Correlations for Indices of Text Difficulty and Text Cohesion with Grade Level

Variable	Correlation
Flesch-Kincaid Grade Level	0.77**
Argument overlap all distances	-0.39**
Stem Overlap, adjacent	0.05
LSA, sentence to text	-0.53**
Incidence of causal verbs	0.38**
Incidence of causal particles	0.02
Incidence of abstract nouns	-0.19**
Ratio of pronouns to noun phrases	-0.37**
Incidence of conditionals	0.16**

Notes: *$p < .05$; **$p < .01$

Abstract nouns, an index of conceptual difficulty, correlated with grade level. However, contrary to our prediction, this correlation was negative: higher grade level texts had fewer abstract nouns per 1000 words, suggesting that the texts became more concrete as they increased in grade. One possible explanation is that higher grade level texts involve increasingly narrow topics, with fewer generalities, leading to an increased use of concrete words that are specific to the topic at hand. The proportion of noun phrases that are pronouns, an index of complexity of writing style, also correlated with grade level. Again, this correlation was negative, which was in the opposite direction to our prediction. This may again be due to an increase in topic-specific information, which requires an increase in explicit noun phrase description. Alternatively, pronouns are high frequency words: the lower incidence of pronouns at higher grade levels may simply be a reflection of the overall decrease in average word frequency for higher grade level text (Stenner et al., 1987).

The bivariate relationships shown in Table 1 are informative. However, the crucial question is whether they add any explanatory value to the assignment of grade level above readability. To test this question, we performed a multiple regression with Flesch-Kincaid Grade Level and all cohesion measures described above. Because Flesch-Kincaid correlated with grade level with an R^2 value of .61, a significant increase of this value would represent a significant increase in accounting for grade level.

The regression of all of these indices against Publisher Assigned Grade Level produced a total R^2 of .68. This demonstrates that the Coh-Metrix cohesion indices produced a significant increase of .07 in the total explained variance in assigned grade level. The final regression equation is shown in Table 2.

The regression provides a very different picture than the bivariate correlations of the relationship between cohesion and grade level. First, it is worth noting that the representative readability measure, Flesch-Kincaid Grade Level, remained a highly significant predictor when included in the regression in competition with the cohesion indexes. This is not a surprise.

Table 2
Summary of Multiple Regression Analysis for Variables Predicting Grade Level of Textbook ($N = 311$)

Variable	B	*SE B*	β
Flesch-Kincaid Grade Level	0.33	0.02	0.69**
Argument overlap all distances	-0.70	0.53	-0.08
Stem Overlap, adjacent	-0.69	0.43	-0.10
LSA, sentence to text	-1.35	0.53	-0.13*
Incidence of causal verbs	-0.01	0.00	-0.07*
Incidence of causal particles	0.01	0.01	0.10*
Incidence of abstract nouns	0.00	0.00	0.04
Ratio of pronouns to noun phrases	-0.75	0.49	-0.07
Incidence of conditionals	0.01	0.02	0.03

Notes. $R^2 = .68$; *$p < .05$; **$p < .01$

Of the cohesion measures, only three contributed to the estimation of grade level beyond readability. They were: LSA, sentence to text; incidence of causal verbs; and incidence of causal particles.

LSA sentence to text had a negative bivariate relationship with grade level: similarly, in the regression equation, the relationship is negative. Higher grade level texts have more semantically diverse sentences. Incidence of causal verbs had a positive relationship with grade level, whereas the relationship in the regression equation is negative. This is known as a suppression effect: the explanatory power of the positive relationship is accounted for by other variables in the equation. Incidence of causal particles had no

relationship with grade level but in the regression equation has a positive relationship. This is because, as stated earlier, causal particles are needed in the text to the extent to which causal information (approximated by causal verbs) is present. Once causal information is controlled for, as it is in the regression equation, the amount of causal particles becomes a straightforward measure of text difficulty: a greater number of particles indicates that the causal relationships are made explicit by the writer, which would be appropriate for a younger audience.

Summary and Conclusions

The complex business of assigning textbooks to classrooms has always involved some input from quantitative indices, such as Flesch-Kincaid Grade Level, in combination with intuition, expert judgment, availability and the requirements of the state. With the advent of new technologies such as Coh-Metrix, a more rigorous approach can be undertaken. Surface characteristics such as sentence length and word length, captured by indices such as Flesch-Kincaid Grade Level, have a role to play. However, new techniques provide automatic estimates of deep constructs of the text, such as cohesion. A combination of old and new approaches provides a better estimate of the difficulty the text and can be used to evaluate its appropriateness for any particular classroom setting.

The cohesion measures evaluated here added explanatory power to grade level assignment beyond the readability measure, Flesch-Kincaid. However, they did not replace it. The Flesch-Kincaid Grade Level readability formula, while simple, remains a significant determiner of where a text lies on the grade level spectrum from K to 12.

A note of caution should be given, however, regarding the use of existing grade classifications to evaluate new methods of determining grade level. Given the historical predominance of readability indices in determining what grade level a textbook should be assigned to, existing classifications may have been made with indices such as the Flesch-Kincaid Grade Level as a factor in the classification process. For this reason, the current analysis may overstate the importance of Flesch-Kincaid Level as a valuable determiner of grade level. However, the results presented here indicate that this is not a completely circular trap: clearly, publishers have been using many factors to determine grade level, and this process can be more closely modeled with a variety of indices than with readability alone.

An unanswered question arising from this analysis is whether cohesion measures have anything to contribute beyond the explanatory power of more modern readability formulas, such as the DRP or the lexile. The results do show, though, that cohesion does predict publisher-assigned grade level, and that cohesion in combination with a readability formula (in this case Flesch-Kincaid Grade Level), predict publisher-assigned grade level better than either readability alone or cohesion alone. This suggests that cohesion will not replace readability as a diagnostic tool, but nonetheless has a role to play in the evaluation of text difficulty.

Regardless, the ultimate answer must come from readers themselves. The variables that provide the most utility in matching reader to text will be discovered by experimental studies of comprehension and memory. To this end, the Coh-Metrix project at the University of Memphis is conducting a series of studies to determine the effectiveness of indices of cohesion and readability in predicting performance on comprehension for a variety of texts.

Therefore, the analysis presented here is not a conclusive answer to the question of determining the appropriate text for a particular learner or set of learners. However, they do indicate that the task cannot be completed with readability formulas alone.

Acknowledgements

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

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.

Bruss, M., Albers, M. J., & McNamara, D. (2004). Changes in scientific articles over two hundred years: A Coh-Metrix analysis. In *Proceedings of the 22nd Annual International Conference on Design of Communication: the Engineering of Quality Documentation* (pp. 104-109). New York: ACM Press.

Chall, J. S., Conard, S. S., & Harris-Sharples, S. (1991). *Should textbooks challenge students? The case for easier or harder textbooks*. New York: Teachers College Press.

DuBay. W. H. (2004). *The Principles of Readability.* Costa Mesa, CA: Impact Information.

Foltz, P., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25*, 285–307.

Graesser, A.C., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: 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.

Hubisz, J. (2001) Review of Middle School Physical Science Texts. Final Report. The David and Lucile Packard Foundation. Grant No. 1998-4248. Retrieved January 30, 2006, from http://www.science-house.org/middleschool/reviews/hubisz.pdf

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch W. & van Dijk T A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-94.

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

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

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Linderholm, T., Everson, M.G., van den Broek, Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more and less skilled readers' comprehension of easy and difficult text. *Cognition and Instruction, 18,* 525-556.

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

McCarthy, P.M., Lewis, G.A., Dufty, D.F., & McNamara, D.S. (2006). Analyzing writing styles with Coh-Metrix. In *Proceedings of the 19th Florida Artificial Intelligence Research Society International Conference (FLAIRS).* Menlo Park, California: AAAI Press.

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

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

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

McNamara, D.S., Louwerse, M.M., Cai, Z., & Graesser, A. (2005). Coh-Metrix version 1.4. Retrieved [August 1, 2005], from http//:cohmetrix.memphis.edu

Ozuru, Y., Dempsey, K., Sayroo, J., & McNamara, D. S. (2005). Effects of text cohesion on comprehension of biology texts. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1696-1701). Hillsdale, NJ: Erlbaum.

Stenner, A. J. (1996). *Measuring Reading Comprehension with The Lexile Framework:* Durham, N. C.: Metametrics, Inc. presented at the California Comparability Symposium, October 1996. Retrieved January 30, 2006, from http://www.lexile.com/DesktopDefault.aspx?view=re

Stenner, A. J., Smith, D. R., Horabin, I., & Smith III, M. (1987). *Fit of the Lexile Theory to Sequenced Units from Eleven Basal Series.* Durham, N. C.: Metametrics, Inc. Retrieved January 30, 2006, from http://www.lexile.com/DesktopDefault.aspx?view=re

Trabasso, T., and van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of memory and language, 24,* 612-630.

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