Independent evaluation finds ATOS a valid and reliable tool for estimating text complexity

The Development of ATOS

The Renaissance Readability Formula

Written for Renaissance Learning by
Michael Milone, Ph.D.

Renaissance™ Learning
Advanced Technology for Data-Driven Schools

RENAISSANCE LEARNING, INC.
P.O. Box 8036
Wisconsin Rapids, WI 54495-8036
(800) 338-4204
www.renlearn.com
answers@renlearn.com

4/12

Contents

Tables

Figure

Introduction

Approximately 30,000 schools worldwide use Accelerated Reader (AR) Enterprise, making it the most popular educational software. AR software is used to monitor students' reading progress while motivating them to read more. A central factor that has contributed to the success of this tool is the accuracy with which students can be matched with texts that reflect their level of reading achievement and interests. The accuracy of this match is due in part to the process that is used to analyze the readability of books. This document explains the science underlying this process.

The ATOS readability formula was developed as a result of collaboration between Renaissance Learning and a diverse group of outside readability experts who had access to unmatched resources.[1] Among the resources was the Accelerated Reader Reading Practice Database, a remarkable accumulation of student performance on progress-monitoring assessments. The database has received national attention since 2008 when it first served as the basis for the landmark annual study, *What Kids Are Reading: The Book-Reading Habits of Students in American Schools* (Renaissance Learning, 2012b). Over the years, this study has been reported on extensively by various media outlets, including the *Washington Post*, *New York Daily News*, and *Huffington Post*.

In research complementing the release of the new 2010 Common Core State Standards, Student Achievement Partners conducted an evaluation of the validity of ATOS and five other measures of text complexity (Nelson, Perfetti, Liben, & Liben, 2011). The evaluation involved obtaining text complexity ratings for reference measures of both text characteristics and student performance. The results indicated that ATOS and the other measures were similarly capable of accurately and reliably predicting the reference measures of text complexity. A report entitled *Text Complexity: Accurate Estimates and Educational Recommendations* (Renaissance Learning, 2012a) describes the study and how ATOS can be applied in educational settings to help students achieve the goals of reading complex texts independently, proficiently, and with understanding.

Readability and Its Importance

Readability is one of the longest-standing constructs in contemporary education. As far back as the 1920s, researchers were interested in ways of improving the comprehension of content-area textbooks (Zakaluk & Samuels, 1988). Lively and Pressey (1923) are credited with the first readability formula, and since their pioneering efforts, a number of prominent educators have developed other formulas.

The primary purpose of measuring readability is to determine in part which individuals can read the text with understanding. Readability formulas are based on characteristics of text that can be measured objectively and dependably, such as sentence length and word difficulty. The results of a readability analysis are reported using a numeric scale of some kind that corresponds to a comparable scale of an individual's level of reading achievement. The scale that is most frequently used to report readability is a grade-level estimate of the difficulty of text.

Although the construct of readability seems relatively straightforward, it is important to consider what it means in practical terms. A book with a readability estimate of grade 4.5 is written in a way that is understandable to individuals who have reading comprehension skills typical of a student in the middle of grade 4. The readability measure does not reflect either the content of the book, which may or may not be appropriate for a fourth grader, or the background and interests of the reader.[2]

---

[1] Touchstone Applied Science Associates, Inc. (TASA), now called Questar Assessment, Inc., known internationally for its Degrees of Reading Power test and readability scale, was a key partner in the ATOS development process. Thus, the formal name for the improved readability formula that was developed is the Advantage/TASA Open Standard for Readability, but it is more commonly referred to as ATOS.

[2] In AR, each book is assigned one of four interest level codes that refer to the sophistication and maturity level of a book's content, ideas, and themes, and is based on publisher recommendations: LG for lower grades (K–3), MG for middle grades (4–8), MG+ for more mature middle-grade readers (6 and up), and UG for upper grades (9–12). Interest level and ATOS book level are intended to work together to inform book selection. For example, Alice Walker's *The Color Purple* and John Steinbeck's *The Grapes of Wrath* have ATOS readability scores of 4.0 and 4.9, respectively, indicating that average-performing students in fourth grade or higher would be likely to understand the text. However, because of the books' mature themes, both are coded with an interest level of UG, or upper grade.

As is true with other measures, there is a certain amount of error associated with readability analyses. We can count with precision the number of words in a sentence or calculate the average length of each word, but when these or other factors are used as part of a readability formula, especially with a book, the resulting value is an estimate of the understandability of the text. This estimate must be balanced with the subjective decision about the appropriateness of the content of the text for the reader. Also, there is error variance associated with the reader's estimated level of reading achievement.

The most common use of readability formulas is to match appropriate textbooks and trade books to school students to read independently with understanding. Devising a reliable and valid readability formula is important because students are likely to develop their reading skills most by reading books that are neither too hard nor too easy for them. Matching books to students' reading achievement levels helps students feel successful and keeps them motivated. It ensures that the books students read are challenging enough to provide useful practice and self-learning but not so hard as to be frustrating (Chall, Conard, & Harris-Sharples, 1991).

A correct match between books and students allows students to devote their intellectual resources to understanding what they are reading. Effective comprehension requires a balanced allocation of limited attentional resources. If components of the process that are normally automatic in good readers, such as decoding, begin to absorb too many attentional resources because the text is overly difficult, higher order processing is likely to suffer (LaBerge & Samuels, 1974; Van Dijk & Kintsch, 1983).

An often-overlooked benefit of an accurate readability measure is that it can help students self-select books. When students choose books that match their interests and level of reading achievement, they gain a sense of independence and commitment. They are more likely to complete, understand, and enjoy the book they are reading.

As mentioned earlier, readability analysis of a book is an incomplete measure. It does not take into account an individual student's purpose for reading, degree of prior knowledge, or interests. Some students are able to read difficult texts that interest them while they struggle with other texts that are nominally of the same readability level.

Furthermore, no readability measure can take into account less objective characteristics such as structural and organizational features of the text, textual coherence and sequence, concept density, degree of abstraction, atypical punctuation, illustrations, and so on. If the concepts presented in a book are relatively simple, students will comprehend the book even if the text is poorly organized. In contrast, if the content is too difficult, students will be less likely to understand it even when the text is well organized (Roller, 1990). Thus, a single readability formula might not be equally appropriate for fiction and nonfiction, trade books and textbooks, books for emergent readers, and so forth.

A readability estimate can be considered a probability statement rather than an exact scientific measure. Any estimate of text difficulty needs to be supplemented by the judgment of the student and significant adults. Despite these limitations, however, readability formulas are still useful tools in the management of successful reading practice. An accurate readability estimate will help teachers maximize student reading growth by guiding students to books more or less within their zone of proximal development, or ZPD (a concept inspired by Vygotsky, 1978). This is the range of difficulty that is neither too hard nor too easy—the level at which optimal learning takes place.[3]

[3] When using AR, students select books to read that fall within recommended ZPD ranges based on their scores from a norm-referenced, standardized measure of general reading achievement, such as Renaissance Learning's STAR Reading assessment. For more information, see Appendix F: Accelerated Reader Goal-Setting Chart, p. 31.

Previous Measures of Readability

There have been many attempts since the 1920s to devise readability formulas that are reliable, valid, and easy to apply. Some were more effective than others. Gray and Leary (1935) compiled a list of almost 300 factors from four categories: content, style of expression and presentation, format, and organization. They reduced this list to a more manageable number and found that five style factors produced a multiple-regression correlation coefficient of .65 with their criterion variable, the Adult Reading Test (Zakaluk & Samuels, 1988).

Subsequent researchers focused on semantic and syntactic elements rather than style factors (Bormuth, 1969; Chall & Dale, 1995; Dale & Chall, 1948; Flesch; 1948; Fry, 1968; McLaughlin, 1969; Spache, 1968; Vogel & Washburne, 1928). Their formulas included variables such as mean number of words (per page or other unit), mean sentence length, mean word length, number of syllables, number of polysyllabic words, number of words appearing in lists of high frequency (easy) words, and so on. The variables served as proxies for semantic and syntactic difficulty, which cannot be measured directly. Although the classic readability formulas provided educators with an acceptable level of guidance, they were based on data sources limited by then contemporary technology.

In addition to developing individual formulas, the researchers often recommended different sampling procedures. Initially, sampling was necessary to overcome the difficulty associated with analyzing large amounts of text by hand. Because there are meaningful differences in the characteristics of separate parts of a book, sampling typically contributed additional error variance.

A recent development in readability is the attempt to create compatible scales capable of reflecting both text readability and student reading competence. Two of these efforts, the Lexile Framework by MetaMetrics, Inc. (Stenner & Burdick, 1997), and Degrees of Reading Power (Koslin, Zeno, & Koslin, 1987), use a continuous metric rather than the more practical grade-level method featured in most approaches to readability and student achievement.

Any approach to readability must be interpreted thoughtfully and with professional judgment. For instance, the formula cannot take into account how much prior knowledge the reader brings to the text. Nor can a readability formula measure the context in which a book was written, the reader's interests, the difficulty of the concepts in a book, or even its organizational structure and coherence of text.

The following two passages demonstrate one of the shortcomings of any system of analysis based solely on text characteristics.

> *Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation or any nation so conceived and so dedicated can long endure.*
>
> *Endure long can dedicated so and conceived so nation any or nation that whether testing, war civil great a in engaged are we now. Equal created are men all that proposition the to dedicated and liberty in conceived, nation new a continent this upon forth brought fathers our ago years seven and score four.*

The passages are the first two sentences of Lincoln's Gettysburg Address, but the second passage is backward. All readability formulas would rate these two passages exactly equal, even though the second is gibberish. The simple truth is that no readability formula can determine the understandability or quality of text. They are imperfect tools that can provide useful information for trained educators to make appropriate instructional decisions about students.

Again, a readability formula can quantify some of the characteristics of text but not the appropriateness of content. Text readability levels matched to a student's independent reading range do not necessarily indicate that the content of the book is suitable in any other way for the student. In some books, readability is deliberately kept low, while the content or interest level is appropriate for students of higher grade levels. (This is why interest level must be used in conjunction with ATOS book level.) The teacher or parents might not believe that reading such books would be appropriate for younger students, just as not every movie is appropriate for young children, no matter how well made it is. The teacher, parent, or other significant adult must make their own judgments regarding the maturity level of the theme of the book, the nature of the language within it, or any other relevant characteristics. Moreover, there are likely to be local, cultural, and individual variations in these judgments.

The ATOS Development Process

During the ATOS development process, extensive research was done by Renaissance Learning in the area of assigning text readability and student reading competence to the same scale. A panel of noted educators and psychometricians was involved in the development of the initial formula and subsequent refinements. The data used included a large sample size and actual book-reading data. The Renaissance Learning approach linked students' recommended book readability ranges (ZPDs) to the students' standardized grade-equivalent scores on a nationally standardized test. This resulted in an "open standard" in which any reading test that returns grade-level scores can be used to match appropriate books to students.

While developing the formula that would serve as the basis for the book-matching process, weaknesses of previous formulas were considered, which led to the identification of several areas to improve upon:

- Updating and broadening the corpus of words used as the basis for the semantic aspect of a readability formula
- Expanding the kinds of texts for which the process would be appropriate
- Considering possible adjustments to account for the repetition of special words (e.g., proper nouns) within text
- Investigating ways to account for variance in the use of words across grade levels (e.g., words with overall low frequency but early emergence)
- Examining how book characteristics (e.g., nonfiction or high interest/low readability) impact the accuracy of readability formulas

Several other considerations also guided the ATOS development process. It was determined the formula should be implemented in a way that is accessible to educators at no or low cost, produces results quickly while being understandable and transparent to educators and parents, and is instructionally sound and plausible to those involved in the education process.

Collaborating with Renaissance Learning were several outside readability experts, including staff from Touchstone Applied Science Associates, Inc., or TASA (now Questar Assessment, Inc.), the makers of the Degrees of Reading Power (DRP) test and readability scale. This partnership combined the resources of

two leading organizations using readability formulas and merged their expertise in psychometrics and assessment. The goal was to develop an updated and improved readability formula to be known as the Advantage/TASA Open Standard for Readability, or ATOS.

It is important to keep in mind that the ATOS system comprises three separate but related approaches to reporting readability. Each approach is based on the same formula but has a specific purpose. The adaptations of the basic formula will be discussed later:

1. The **ATOS for Text Readability Formula** is a grade-level scale intended for use with short passages, such as those that appear in textbooks or on reading tests.
2. The **ATOS for Books Readability Formula** is the grade-level scale used with Accelerated Reader.
3. Two alternate versions of ATOS for Books convert the grade-level scale to the **ATOS 100-Point Scale** and the **ATOS 2000-Point Scale**, similar to those used by DRP and Lexile, respectively.

Resources used for the development of ATOS

Renaissance Learning has more than two decades of experience using readability measurements to help teachers improve students' learning opportunities. Readability has always been important to AR users. Determining the readability level of books serves two vital purposes in AR. First, it helps both students and teachers select books that will provide the most beneficial reading practice. Second, readability level and book length determine the point value of books, which is an accurate, objective measure of the amount of reading practice achieved. Previously, AR employed the popular Flesch-Kincaid formula for estimating readability. Although Flesch-Kincaid was an acceptable formula, the panel of experts believed a better formula could be developed. The search for a more accurate readability measurement that improved on Flesch-Kincaid and other readability formulas led to two years of study and several more years of refinement.

This project is believed to be the largest and most comprehensive study of readability ever undertaken. Conducting the study required the creative use of resources and tools uniquely available to the partners. The criterion variable was a large set of more than 250 authentic reading comprehension items. They were drawn from TASA's norm-referenced DRP test, which is widely used and has extensive evidence of reliability and validity.

The semantic variable included in the formula was based on the world's largest word-frequency corpus, approximately 474 million words representing all the text of 28,000 fiction and nonfiction K–12 books in the AR quiz library.[4] These were authentic books read by real students. Because many were published recently, this corpus was more relevant than those used in prior readability research.

The corpus was supplemented with an improved and expanded graded vocabulary list of almost 24,000 words. The starting point was TASA's analysis of trade books and textbooks, which shows the levels where words are most often found as well as their distribution across other grades. This analysis, combined with an examination of several standardized tests, helped to determine an appropriate grade-level assignment for each word.

The AR Reading Practice Database, which reflected the AR performance of more than 3 million students, was a critical source of data used to verify the validity of the formula. For many of the students, the database also contained reading achievement scores from Renaissance Learning's STAR Reading computer-adaptive assessment. Thus, for the first time, substantial data on student book reading was incorporated into the design of a readability formula.

[4] As of this writing, the Renaissance Learning word frequency corpus is much larger, consisting of more than 145,000 books and more than 2 billion words.

Technical development of ATOS

The first step in developing the ATOS formula was to identify the objective characteristics of text that could be used as the basis for a readability formula. Most readability formulas seek to measure text difficulty through both semantic (word-level) and syntactic (sentence-level) characteristics (Zakaluk & Samuels, 1988). This was followed by regression analysis to determine which variables, alone or in combination, related most closely to text difficulty as reflected by student performance on the DRP comprehension test.

The characteristics identified included commonly used attributes of text such as sentence length, word length, and word frequency, as well as less common attributes, such as the number of dependent and independent clauses, verb forms, pronoun use, the length of clauses, paragraph length, ratio of unique to total words, and other attributes that contribute to the understandability of text.

Table 1 shows the relationship to text difficulty of several semantic and syntactic variables. The R^2 numbers indicate how much of the variation in text difficulty could be explained by each variable independently. Higher R^2 values suggest stronger relationships between the predictor and criterion variables. Three variables were ultimately selected as the key components of the ATOS formula: words per sentence, average grade level of words, and average characters per word.

Table 1: Correlation of Semantic and Syntactic Variables to Text Difficulty

Variable	Variance Explained (R^2)
Words per sentence*	.897
Average grade level of words*	.891
% familiar words	.867
Syllables per word	.839
Characters per word*	.839
Word frequency	.769

** Variables used in final ATOS formula*

What was surprising to the research team was that average grade level of words in the text proved to be a better predictor of text difficulty than percent of familiar words and word frequency, two indices used by other readability formulas. One possible explanation appeared to be that many words are common at a certain age or grade level but are less common in later grades. An example is the word “kitten,” which is frequent at lower grades but far less frequent at higher grades. In cases like this, infrequency at higher grade levels does not make these words difficult. In addition, raw-word frequencies are not corrected for derivations such as plurals and past tense forms. The word list used for the ATOS formula uses not only root words but also their close derivatives of comparable familiarity, such as dog-dogs and play-played.

Building the graded vocabulary list

The ATOS formula’s graded vocabulary list is perhaps the most extensive and accurate resource of its kind. A number of data sources were used to develop the list:

- The Dale (1931) long list of familiar words
- TASA’s familiar word list and *Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995)
- More than 28,000 works of popular children’s literature in the AR quiz library
- Analysis of DRP/STAR Reading passages

- A list of words from several nationally normed standardized tests (see next page)
- The Dolch (1936) Basic Sight Vocabulary List
- The EDL Core Vocabularies (Taylor, Frackenpohl, & White, 1979)
- The Harris-Jacobson (1972) basic vocabularies
- *The Living Word* (Dale & O'Rourke, 1976)

A preliminary list of words was developed by pooling the sources, excluding duplicates, adding common derivatives when necessary, and removing

- Individual letters except the word a
- Proper names (i.e., people, place names)
- Numerals (including Roman Numerals)
- Acronyms (e.g., DDT)
- All abbreviations but the most common (e.g., Mr., Mrs., Dr.)
- Hyphenated words

Assigning grade levels to words involved using information from existing word lists as well as word frequency studies, vocabulary test results, and expert judgment. For many words, there was sufficient concordance among the sources to place a word with confidence. When there were discrepancies, the more current sources were given greater weight.

For derivative forms of words, assigning grade levels involved using a modified version of Harris and Jacobson's (1972) rules because the original rules were not considered sufficiently comprehensive. It was thought this could have led to misplacing words at levels much lower than they should have been placed.

Derivatives of words can vary greatly in frequency from the base word. In assigning grade levels, derivatives with parallel meanings to the base word were included in the same grade level as the base since derivatives are taught to students from first grade onward. If the derivative form of a word assumed a different meaning than the root, like "hand" and "handily," the derivative forms were examined to make grade assignments. Some examples of the various grade assignments are in Table 2.

Table 2: Examples of Grade Assignments for Derivative Forms of Base Words

Word	Grade	Word	Grade	Word	Grade
hear	1	hold	1	wick	3
hears	1	holder	4	wicked	4
hearings	6	holding	1	wicker	6
		holdings	6		

In the lower grades, additional placement decisions were made based on factors such as decodability, introduction of words in basal series, and common usage. Special categories of words were also reviewed and graded systematically, including:

- Familial relations (e.g., mother, father)
- Pronouns
- Numbers
- Days of week and months of year

In order to study the validity of the grade assignments, a sample of the words on the improved graded vocabulary list (including all derivatives) were compared to words used at various grade levels on five major standardized tests:

- California Achievement Test, Form A (CTB/McGraw-Hill, 1992)
- Comprehensive Test of Basic Skills, named the Terra Nova, Form A in its latest form (CTB/McGraw-Hill, 1997)
- Stanford Achievement Test, Form S (Harcourt Educational Measurement, 1995)
- Metropolitan Achievement Test, Form S (Harcourt Educational Measurement, 1993)
- Iowa Test of Basic Skills, Form K (Riverside Publishing Company, 1993)

More than five percent of the words on the list appeared on at least one of the standardized tests. The majority of the words were within one grade level of the list assignment. The outliers, words that were several grades off the list assignment, were typically distracters rather than target words or multi-meaning words with both common and uncommon meanings, such as “bear” the animal and “bear” meaning to carry. These results suggested that the grade-level placements on the improved graded vocabulary list were consistent with the standardized tests. One would not expect a perfect match because the tests include words with different p-values (percent correct) to test a broad range of students. In other words, an eighth-grade word may appear as a target word on a fifth-grade standardized test, but its p-value would be very low. P-values were not used to compare grade-level placements, just the grade at which a word appeared.

Best predictor analysis

A total of 25 semantic and syntactic variables were evaluated in the best predictor analysis. The criterion variable was the Rasch difficulty value of 256 DRP criterion-referenced test items. The project team first examined the individual correlations between the predictor and criterion variables. This examination was followed by a multiple-regression analysis. All the predictor variables had a positive relationship with the criterion variable.

In conducting the analysis, some of the predictor values were transformed as suggested by Mathews (1992). Although he mentioned 11 different transformations, a number of those suggested were not appropriate for this particular analysis. Table 3 summarizes the six transformations deployed in the analysis where y is the Rasch difficulty value of the items and x is the semantic or syntactic variable. Numerical type labels (in the first column) are those suggested by Mathews.

Table 3: Types of Transformations

Type of Transformation From Mathews	Transformation Formula From Mathews	Semantic or Syntactic Variable	Rasch Difficulty
1	y = Ax + B	x	y
2	y = A/x + B	1/x	y
6	y = ALn(x) + B	Ln(x)	y
7	y = Cexp(Ax)	x	Ln(y)
8	y = Cx^A	Ln(x)	Ln(y)
10	y = Cxexp(-Dx)	x	Ln(y/x)

Table 4 shows the best predictors emerging from the initial regression analysis and the percentage of variance accounted for. The ***bold italic*** cells indicate the best transformation type for each variable. The left column shows the best predictor variables (which are defined below the table). The top row shows the type of transformation from Table 3. Each cell represents the variance accounted for by the predictor in its transformed value. The dependent variable is the Rasch difficulty of the passage items.

Table 4: Squared Correlations of the Best Predictors With Each of the Selected Six Transformations

Predictor Abbreviation	Type 1	Type 2	Type 6	Type 7	Type 8	Type 10
AvgChar	0.839	***0.873***	0.861	0.810	0.840	0.844
SDChar	0.870	0.869	***0.884***	0.849	0.880	***0.884***
AvgWords	***0.847***	0.741	0.818	0.843	0.835	0.748
FamWords	0.867	0.787	0.834	0.847	0.804	***0.885***
AvgGrade	0.848	0.865	***0.880***	0.820	0.871	0.872
AvgGrad100	0.868	0.856	***0.885***	0.848	0.884	0.859
SDGrade	0.849	0.831	***0.861***	0.832	0.861	0.827
AvgSyll	0.839	***0.884***	0.868	0.809	0.848	0.862
SDSyll	0.855	0.813	0.863	0.837	***0.867***	0.814
Mono	0.868	0.804	0.845	0.855	0.819	***0.877***
Poly			0.818		***0.829***	

Predictor abbreviation definitions:

AvgChar: Average word length, calculated as average number of characters per word for words found on improved graded vocabulary list (sentence letters/tokens on word list)

SDChar: Standard deviation for number of characters per word in complete sentences, or average word length (sentence letters/sentence tokens)

AvgWords: Average sentence length or average number of words found in complete sentences (sentence tokens/complete sentences)

FamWords: Relative frequency of familiar words to total words found in complete sentences. Familiar words are easy words or proper nouns drawn from TASA's database of words (number of familiar words/number of total words)

AvgGrade: Average grade level for words found on graded vocabulary list

AvgGrad100: Average grade level for words found on graded vocabulary list excluding top 100 most frequent words in ALS/TASA corpora combined

SDGrade: Standard deviation for average grade level for words found on graded vocabulary list

AvgSyll: Average number of syllables per word based on tokens in complete sentences and referenced in a syllable dictionary of 69,794 words (sentence syllables/sentence tokens)

SDSyll: Standard deviation for number of syllables per word based on a syllable dictionary

Mono: Number of monosyllabic words divided by total number of words

Poly: Number of polysyllabic words divided by total number of words. Polysyllabic words have 3 or more syllables.

Various combinations of variables were used in a multiple regression analysis. The final form of the ATOS for Text Readability Formula proved to be the best predictor of the criterion variable.

ATOS for Text Readability Formula

The first stage of the development process resulted in the ATOS for Text Readability Formula, which is based on the number of words per sentence, average grade level of words, and average characters per word. Although it correlates highly to text difficulty, the formula was developed using nationally normed reading passages and thus does not take into account the chief difference between short passages of text and full-length books.

To tailor ATOS for use with trade books, the unique aspects of book reading were investigated and several adjustments were developed. Both the effectiveness of these adjustments and the impact of ATOS on different categories of books were examined using the AR Reading Practice Database. Some of these investigations are highlighted below.

An initial study was conducted to place the difficulty values of DRP test items and STAR Reading test items on the same scale. Both tests use item response theory (IRT) to scale item difficulty, and both scales are based on a method developed by Rasch (Rasch, 1980; see also Fischer & Molenaar, 1995). The objective of the study was to "link" the two scales to convert the STAR Reading Rasch difficulty parameters to the DRP scale, thus ensuring the two assessments were measuring the same underlying construct.

The main study then sought to explore the predictive value of 25 semantic and syntactic variables to item difficulty using mean Rasch difficulty values as the criterion for an initial data set of 256 DRP criterion-referenced test items and 294 STAR norm-referenced test items. The final data set consisted of 225 DRP criterion-referenced test items. Validation was carried out using 294 STAR items, 36 aggregated STAR items, an additional 138 DRP test items, 52 samples of authentic text used to validate the Dale-Chall readability formula revised in 1996 (Chall & Dale, 1995), and three popular language arts basal series. In addition, regressions to measure the predictive value of variables believed to be unique to emergent-reader books were conducted. In the end, the following formula for reading difficulty, expressed on the Rasch logit scale, was developed:

ATOS Rasch Difficulty Formula = -8.54 + 1.95 * Ln(AvgWords) + .46 * AvgGrad100 + 1.74 * Ln(AvgChar)

Grade-level conversion

Once the Rasch Difficulty formula was developed, a separate study was conducted to convert the predicted Rasch difficulty parameters produced by the ATOS formula onto a grade-level scale. The primary data source for accomplishing this task was the AR Reading Practice Database which then contained more than 950,000 Accelerated Reader records for more than 30,000 students reading and testing on actual literature books. To ensure that each student carefully read a sufficient number of books at an appropriate level, only students who had scored between 60% and 100% on the quizzes for at least 10 books were included in the analysis. Reading records for nearly 30,000 students in grades K–12 who read more than 9,000 different books were analyzed.

To develop a grade-level conversion formula, the average Rasch difficulty and average AR readability level of each student's reading list were calculated. The average predicted Rasch difficulty values were plotted against the average readability levels, and a quadratic function was fitted to the data points, resulting in a transformation from Rasch difficulty to grade-level equivalents.

Impact of book length

Another study qualified the influence of book length. This variable was not previously used in readability formulas, but it was suspected that longer books would prove generally harder to read than shorter books. Many teachers have intuitively felt this to be true and have taken length into consideration when

recommending books to students. To test this assumption, samples of books were selected and grouped into categories with about the same difficulty level as indicated by their Flesch-Kincaid score. Within each difficulty category, books were split into two groups by their length (number of words). AR quiz scores for the books were averaged by group, and it was confirmed that length was negatively correlated with comprehension scores. Therefore, the formula ultimately used to determine ATOS readability for books takes length into account. (See "Adjustment for book length" below and Appendix G, p. 32, for more information.)

More recently, a sample of 3 million AR quiz records captured during fall 2008 was selected to reexamine the relationship between length and difficulty. It was confirmed that very short books tended to have the highest comprehension scores, and generally, those scores tended to decrease as the number of words increased. Students scored an average of 87% correct on comprehension quizzes covering books with up to 500 words, and 84% correct on books with 501–5,000 words. For very long books (250,001 to 500,000 words), average percent correct was 65%.

Fiction/nonfiction and high interest/low readability books

Data from the AR Reading Practice Database also allowed the examination of how well the formula predicted the relative difficulty of fiction, nonfiction, and high/low books (those of high interest with a low readability level), with difficulty indicated by performance on AR comprehension quizzes. It was found that previous formulas tended to understate the difficulty of nonfiction, possibly because it often contains specialized vocabulary, which is not properly analyzed by just looking at word length. On average, nonfiction books in the AR Reading Practice Database were found to be 0.42 grade levels more difficult than fiction books with comparable characteristics. In the case of high/low books, the analysis showed that older formulas tended to understate the difficulty of these books, presumably because of their shorter length.

Emergent-reader books

Books written for emergent readers have always presented a problem for readability formulas, typically by understating text difficulty. Reading Recovery Council of North America's Reading Recovery scale (Fountas & Pinnell, 1999) was used to identify key variables for leveling such books. Book length was found to be highly correlated to Reading Recovery levels, and as a result, is more heavily weighted in the ATOS calculation for these books. The correlation between ATOS and Reading Recovery levels was also investigated. ATOS was the first formula to provide a guideline for teachers to convert between Reading Recovery and grade-level measurements (see Appendix H, p. 33, for this conversion).

ATOS for Books Readability Formula

Adjustments were made to the ATOS for Text Readability Formula to account for two critical factors that affect the understandability of books, length and variations in internal structure, which resulted in the ATOS for Books Readability Formula.

Adjustment for book length

Teachers have consistently reported that shorter books are easier for students to read than longer books. Likewise, students are often less motivated to read longer books than shorter books. In addition, an analysis of the AR Reading Practice Database showed that when students with similar skills read books with comparable difficulty, they tended to score lower on the AR quizzes for longer books. This finding supports teachers' observations about the relationship between book length and difficulty.

In order to adapt the ATOS for Text Readability Formula to account for book length, a new factor was developed called book-length grade level (BLGL). The following formula describes the relationship between book length and grade level for books longer than 500 words. (Books with fewer than 500 words are treated differently.)

BLGL for Books With More Than 500 Words = 0.68 * Ln(Book Length) – 1.87

This formula was developed by examining more than 10,000 books in the AR Reading Practice Database. (For more information, see Appendix G, p. 32.)

An ATOS for Books readability level is found by combining ATOS for Text and BLGL, weighted as shown in Table 5. As an example, a book of 50,000 words would have a grade level that reflects 80% of ATOS for Text plus 20% of book-length grade level. Before the weights in Table 5 are applied, ATOS for Text is converted from the Rasch metric to a grade-level equivalent by means of a quadratic formula (see "Grade-level conversion," p. 10). The ATOS for Books Readability Formula is shown below.

ATOS for Books Readability Formula = ATOS Wght * ATOS for Text + BLGL Wght * BLGL

Table 5: Weights of ATOS for Text and Book-Length Grade Level

Number of Words	ATOS for Text Weight	Book-Length Grade Level Weight
500	.50	.50
5,000	.60	.40
50,000	.80	.20
100,000	.85	.15
250,000	.90	.10

The 500-word cutoff was chosen because most Reading Recovery books and books for emergent readers have fewer than 500 words. The second level of 5,000 words was chosen because more than half of AR books with readability levels less than grade 5 had between 500 and 5,000 words. Similarly, more than half of AR books with readability levels of grades 5 to 8 had between 5,000 and 50,000 words. The final two cutoffs were chosen because books with more than 50,000 words become more common at readability levels above grade 8 (e.g., *My Antonia* by Willa Cather), and books with more than 100,000 words become more common at readability levels above grade 10 (e.g., *War and Peace* by Leo Tolstoy).

A book-length adjustment was also developed for books with fewer than 500 words. Early investigations into a separate formula for emergent-reader books, using Reading Recovery levels as the criterion, revealed book length as a strong predictor of student comprehension. Rather than develop a separate formula for emergent-reader books, the ATOS for Books Readability Formula was modified slightly for books with fewer than 500 words by increasing the weight of book-length grade level to 60% and reducing the ATOS weight to 40%. (For more information on BLGL, see Appendix G, p. 32.)

Adjustment for extreme sentence length

Because books and assessment passages may differ with respect to syntactic and semantic characteristics, an analysis of sentence length and vocabulary makeup was undertaken. Sentence length was found to be considerably more variable in books than in test passages. The semantic variables, however, were comparable.

To compensate for the syntactic differences between books and test passages, upper and lower sentence-length limits were defined. These limits were established by examining the semantic and syntactic characteristics of the assessment passages that served as the criterion variable. The examination resulted

in an expected range of sentence length for text at a given grade level. When the average sentence length of a book exceeds the upper or lower limits, the value of that book's sentence length is adjusted appropriately. Details about the sentence-length analysis can be found in Appendix I, p. 34.

Alternate Scales for ATOS for Books

The ATOS formula returns a grade-level estimate of text difficulty. Because some educators may be familiar with readability systems that use other scales, a procedure was developed to convert ATOS grade levels to 100-point and 2000-point scales similar to those used by DRP and Lexile, respectively.

The ATOS 100-Point Scale is based on the DRP values and ATOS levels of 4,427 books. Smoothed curves were fit through the data points resulting in the conversion formulas below. A correlation coefficient of .87 between the original DRP values and the ATOS 100-Point Scale suggests the two measures are highly related.

For ATOS <= 4.0: DRP = 10.5 * Ln(ATOS) + 34.5

For ATOS > 4 and <= 7: DRP = 3.1 * ATOS + 36.8

For ATOS > 7: DRP = 15.1 * Ln(ATOS) + 30.2

The initial, pre-2005 conversion to a 2000-point scale similar to the Lexile Framework is based on the published Lexile values for 6,254 books and their corresponding ATOS levels. Two curves were fit through these data points resulting in the conversion equations below. A correlation coefficient of .89 between the original Lexile values and the ATOS 2000-Point Scale suggested that the two measures were highly related.

For ATOS < 2.0: Lexile = 199.4 * ATOS – 29.9

For ATOS >= 2.0: Lexile = 138.1 * ATOS + 93.5

In 2005, the conversion was revised using a different procedure and a larger sample of books. Because the relationship between the ATOS and Lexile scales is curvilinear, an equipercentile equating procedure was used to obtain more accurate equivalent scores. The data analyzed were the ATOS book levels and Lexile measures for 19,760 books for which both measures were available. A software equating application called LEGS, developed by Robert Brennan (2004) at the University of Iowa, was used for the analysis. LEGS produces several alternative tables of equivalence, including linear, raw equipercentile, and smoothed and rounded equipercentiles. A smoothed, rounded equivalence table was chosen. It provides an integer-valued Lexile equivalent for every ATOS book level from 0 through 16.9, in increments of one-tenth of a grade level.

Table 6 (next page) shows sample ATOS book levels and corresponding ATOS 100-Point Scale and ATOS 2000-Point Scale values. The full conversion table, in ATOS increments of one-tenth of a grade level, is presented in Appendix J, p. 35.[5]

[5] Also available in Renaissance Learning. (2006). *Matching books to students: How to use readability formulas and continuous monitoring to ensure reading success.* Wisconsin Rapids, WI: Author. Available online from http://doc.renlearn.com/KMNet/R003544312GE0BA6.pdf

Table 6: Sample of Conversion of ATOS for Books to 100-Point and 2000-Point Scales

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
0.5	27	31
1.0	34	51
2.0	41	241
3.0	46	440
4.0	49	619
5.0	52	781
6.0	55	921
7.0	58	1029
8.0	61	1121
9.0	63	1201
10.0	65	1293
11.0	66	1364
12.0	67	1434

Conclusion

Matching books to students successfully is one of the most essential steps in promoting student achievement and helping students enjoy what they read. To accomplish this match, a dependable readability formula is necessary. It is for this reason that Renaissance Learning was willing to dedicate so many resources to the development of the ATOS readability formula.

The science that underlies ATOS is substantial. A broader range of variables was considered in developing the formula than in any previous effort, and the criterion variable was based on standardized comprehension assessments with a history of successful use. Moreover, other readability formulas were included in the development process in order to capitalize upon the legacy of prior researchers. It can be said with considerable confidence that the ATOS formula is more sensitive to differences in text structure than any other approach to measuring readability.

Unlike other readability formulas, ATOS has been adapted to reflect the characteristics of different texts. The result is that ATOS has two versions, one for short passages of text such as those appearing in magazines or on tests, and one for books. Having two versions of the formula increases the likelihood that the measured readability level is an accurate reflection of the reading achievement level and effort needed to understand the text.

During its development and initial use, the ATOS readability formula was validated using data from thousands of students who completed comprehension assessments for books they read. The performance of the students corresponded closely to what was anticipated based on the ATOS formula. Given this correspondence, it is safe to say that ATOS accurately measures the characteristics that contribute to the understandability of text and can be used with confidence to match texts and books to students for a variety of purposes.

Overview of Appendices

Appendix A: Proper Nouns Study (p. 16)

Appendix B: Fiction Versus Nonfiction Study (p. 18)

Appendix C: High Dialog Study (p. 21)

Appendix D: High Interest/Low Readability Study (p. 25)
A number of text characteristics were discussed during the development of ATOS because it was thought that they might affect the readability and comprehension of books students read. Due to the availability of student comprehension scores from the Accelerated Reader Reading Practice Database, it was possible to investigate the effects of these characteristics. The first four appendices present the results of these studies. (Please note: The data source and method were the same for the first three studies, so this information is provided once in Appendix A.)

Appendix E: Full-Text Sampling Study (p. 27): This study was conducted to compare the error associated with calculating readability based on a sample of text instead of the full text. It showed that readability levels vary throughout a book and result in unpredictable ranges of variability in readability levels when only a portion of the book is used; therefore, it was decided that ATOS would be based on full-text samples.

Appendix F: Accelerated Reader Goal-Setting Chart (p. 31): This chart provides recommended ZPD ranges by student reading achievement level, as measured by the STAR Reading assessment or any other norm-referenced reading test that produces a grade-equivalent score. Educators and students use this chart to set individualized reading-practice goals.

Appendix G: Adjustment for Book Length (p. 32): Because longer books tend to be more difficult to read, the ATOS readability formula was adjusted for book length. This appendix explains the development of the formulas used to adjust readability based on book length.

Appendix H: Conversion: ATOS for Books to Reading Recovery, Guided Reading (p. 33): Books written for emergent readers have always presented a problem for readability formulas, typically by understating text difficulty. ATOS was the first formula to provide a guideline for teachers to convert between the Reading Recovery scale (Fountas & Pinnell, 1999) and grade-level measurements. This appendix presents the conversion.

Appendix I: Definition of Sentence-Length Limits (p. 34): This appendix outlines the analysis conducted to establish an expected range of sentence length in ATOS for text of a given grade level. When the average sentence length of a book exceeds the upper or lower limits, the value of that book's sentence length is adjusted appropriately.

Appendix J: Conversion: ATOS for Books to 100-Point, 2000-Point Scales (p. 35): The ATOS formula returns a grade-level estimate of text difficulty. Because some educators may be familiar with readability systems that use other scales, a procedure was developed to convert ATOS grade levels to 100-point and 2000-point scales similar to those used by DRP and Lexile, respectively. The conversions are shown here.

Appendix A: Proper Nouns Study

Little research has been done on the relationship between the proportion of proper nouns in a book and a student's understanding of the book. Proper nouns typically do not appear in graded vocabulary lists, so they have an unknown effect on the measured readability of books. This study sought to explore the extent to which proper nouns affect students' reading comprehension by examining comparable students' AR quiz scores both for books with high numbers of proper nouns and for books with low numbers of proper nouns.

Data source

The data for the studies in Appendices A–C came from the AR Reading Practice Database. The analysis included 28,808 students in grades 3 through 7 for whom STAR Reading test data and AR quiz data for the 1997–1998 school year were available. The students read 9,285 different books for a total of 1,462,895 student-book pairs. Student-book pairs were the unit of analysis for this study. Books with quiz scores less than 60% were excluded from the analysis because they may have been too difficult for the students who read them, and the results may reflect guessing rather than purposeful effort. Table A1 shows the number of students and books at each grade level.

Table A1: Students and Books at Each Grade Level (Proper Nouns Study)

Student Grade	Number of Students	Number of Different Books	Total Student-Book Pairs
3	7,799	6,234	582,444
4	8,579	7,155	448,984
5	6,939	7,355	283,303
6	4,128	6,922	128,256
7	1,363	3,681	19,908
Total	28,808	9,285	1,462,895

Method

An estimate of each student's level of reading achievement at the time of each AR quiz was determined using STAR Reading scores. To compute the estimates, students' measured grade-equivalent (GE) scores were adjusted assuming an average growth of 0.1 of a grade per month. For students with two measured STAR Reading GEs, the average of the two tests served as the basis for the adjustment.

The student-book pairs were divided into three bands—high, average, and low—based on the estimated STAR Reading GE scores of the students when they took the AR quizzes. Within each band, the ratio of book readability level and student GE was computed and student-book pairs with similar ratios were organized into 10 subgroups. The ratio of book readability to student GE gives a measure of the relative difficulty of a particular book for a particular student. The ratio is less than 1 when the readability is lower than a student's nominal grade and more than 1 when the readability is greater than a student's nominal grade. In other words, when the ratio is higher, the book is relatively harder for a given student.

Within each subgroup, the AR quiz scores on books with proportionally many, a moderate amount, and few proper nouns were compared. The proper-noun category cutoffs were determined by calculating the ratio of proper nouns to the total number of words in the book and dividing these into the lower quartile, middle two quartiles, and upper quartile.

Results

The ratio of proper nouns in a book had no consistent influence on AR quiz scores, suggesting that comprehension is not affected systematically by the ratio of proper nouns to common nouns in a book. For most of the subgroups, there was no difference in quiz scores among books with high, moderate, and low proportions of proper nouns. In a few subgroups, the results suggested that students scored higher on books with more proper nouns. This result appeared most often among students in the upper quartile of reading achievement. The implication is that proper nouns have little effect on the comprehension of poor or average readers, but they make text somewhat easier for good readers. The data are not sufficiently robust to warrant an adjustment of a book's readability level based on the proportion of proper nouns it contains.

Appendix B: Fiction Versus Nonfiction Study

Educators and researchers have believed for years that fiction is easier for most students to understand than nonfiction. In this study, student comprehension of fiction and nonfiction was examined by comparing the comprehension scores of students with similar levels of reading achievement who read fiction and nonfiction books of comparable difficulty. The purpose of the study was to establish an empirical rationale for the former Renaissance Learning recommendation to adjust a student's Zone of Proximal Development (ZPD) down by one-half to one grade level for nonfiction books. The ZPD is the ideal book-level range to promote student reading progress.

The **Data source** and **Method** were the same as in the Proper Nouns Study in Appendix A, p. 16. The comparison of interest involved AR quiz scores for fiction and nonfiction books.

Results

The analysis of the data clearly showed that when students read books of comparable difficulty, they scored lower on AR quizzes for nonfiction books than on quizzes for fiction books. Tables B1 (below) and B2 (p. 20), respectively, show a subset of the results for students in the middle two quartiles of grades 3 and 6.

This finding suggests that students have a more difficult time comprehending or recalling the information in nonfiction books when compared to fiction books. In addition, as books get more difficult relative to the nominal grade of the reader, comprehension scores decrease faster for nonfiction than for fiction books. This pattern holds for nearly all subgroups.

Although the results of this study support the notion of implementing an adjustment for nonfiction books, the adjustment proved to be unnecessary. The ATOS for Books Readability Formula takes into account book length and vocabulary level. These two factors contribute in a meaningful way to the difficulty associated with nonfiction books, so a further adjustment for nonfiction books was not needed.

Table B1: Grade 3 Middle Quartiles

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Fiction Versus Nonfiction

Ratio of Flesch-Kincaid to Grade Equivalent	Fiction or Nonfiction	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
<.3371	F	0.92	28,337	0.12
	N	0.89	780	0.14
	Total	0.92	29,117	0.12
.3371–.5771	F	0.92	27,666	0.12
	N	0.87	1,463	0.14
	Total	0.92	29,129	0.12
.5771–.7189	F	0.91	27,277	0.12
	N	0.86	1,834	0.14
	Total	0.91	29,111	0.12
.7189–.8368	F	0.90	27,367	0.12
	N	0.85	1,744	0.14
	Total	0.90	29,111	0.12

Table B1: Grade 3 Middle Quartiles (continued)

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Fiction Versus Nonfiction

Ratio of Flesch-Kincaid to Grade Equivalent	Fiction or Nonfiction	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
.8368–.9529	F	0.90	26,811	0.13
	N	0.85	2,317	0.14
	Total	0.89	29,128	0.13
.9529–1.0720	F	0.90	27,118	0.13
	N	0.83	2,001	0.14
	Total	0.89	29,119	0.13
1.0720–1.2071	F	0.89	26,936	0.13
	N	0.82	2,189	0.14
	Total	0.89	29,125	0.13
1.2071–1.3656	F	0.89	26,900	0.13
	N	0.81	2,215	0.15
	Total	0.88	29,115	0.13
1.3656–1.6059	F	0.88	25,971	0.13
	N	0.81	3,140	0.15
	Total	0.88	29,111	0.14
>1.6059	F	0.88	24,050	0.14
	N	0.81	5,071	0.15
	Total	0.87	29,121	0.14
Total	F	0.90	268,433	0.13
	N	0.83	22,754	0.15
	Total	0.90	291,187	0.13

Table B2: Grade 6 Middle Quartiles

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Fiction Versus Nonfiction

Ratio of Flesch-Kincaid to Grade Equivalent	Fiction or Nonfiction	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
<.4160	F	0.93	6,163	0.11
	N	0.89	245	0.14
	Total	0.93	6,408	0.11
<.4160–.5274	F	0.91	6,077	0.12
	N	0.86	341	0.13
	Total	0.91	6,418	0.12
.5274–.5996	F	0.90	6,094	0.12
	N	0.86	321	0.14
	Total	0.90	6,415	0.12
.5996–.6604	F	0.90	6,056	0.12
	N	0.85	352	0.14
	Total	0.90	6,408	0.12
.6604–.7219	F	0.89	6,064	0.12
	N	0.84	347	0.14
	Total	0.89	6,411	0.12
.7219–.7885	F	0.89	5,953	0.12
	N	0.85	461	0.14
	Total	0.89	6,414	0.12
.7885–.8714	F	0.89	5,761	0.13
	N	0.84	658	0.14
	Total	0.88	6,419	0.13
.8714–.9831	F	0.89	5,260	0.13
	N	0.84	1,152	0.14
	Total	0.88	6,412	0.13
.9831–1.1619	F	0.88	4,438	0.13
	N	0.83	1,974	0.14
	Total	0.86	6,412	0.14
>1.1619	F	0.87	3,152	0.14
	N	0.81	3,258	0.14
	Total	0.84	6,410	0.14
Total	F	0.90	55,018	0.12
	N	0.83	9,109	0.14
	Total	0.89	64,127	0.13

Appendix C: High Dialog Study

The amount of dialog in a book may affect its understandability. To explore this possibility, students' AR quiz scores were compared for books with little dialog versus those with more dialog.

The **Data source** and **Method** were the same as the Proper Nouns Study in Appendix A, p. 16.

Results

The analysis of the data suggests that for many readers, more dialog makes a book easier to comprehend and recall. In all grades, readers in the upper quartile scored higher on assessments based on books with more dialog. Readers in the middle two quartiles also scored higher in books with more dialog, although the difference was often smaller. Students in the lowest quartile had comparable scores no matter what the level of dialog.

Given the differential performance of students on books with varying amounts of dialog, it would be difficult to derive a generally applicable readability adjustment based specifically on dialog. Moreover, books with high amounts of dialog typically have shorter sentences and easier vocabulary. Because these two factors are considered in the ATOS formula, there is no need for a separate adjustment.

Tables C1 (below) and C2 (p. 23), respectively, show the results for students in the lower quartile of grade 4 and the middle two quartiles of grade 5.

Table C1: Grade 4 Lower Quartile

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Amount of Dialog

Ratio of Flesch-Kincaid to Grade Equivalent	Amount of Dialog (1 = low, 3 = high)	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
<.04502	1	0.87	7,783	0.15
	2	0.90	2,134	0.13
	3	0.93	1,006	0.12
	Total	0.88	10,923	0.15
.04502–.4508	1	0.88	2,943	0.14
	2	0.91	3,500	0.12
	3	0.90	4,227	0.13
	Total	0.90	10,670	0.13
.4508–.6870	1	0.89	2,209	0.14
	2	0.88	3,940	0.14
	3	0.90	4,419	0.13
	Total	0.89	10,568	0.13
.6870–.8518	1	0.88	1,925	0.14
	2	0.89	4,622	0.14
	3	0.89	3,968	0.13
	Total	0.89	10,515	0.14

Table C1: Grade 4 Lower Quartile (continued)

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Amount of Dialog

Ratio of Flesch-Kincaid to Grade Equivalent	Amount of Dialog (1 = low, 3 = high)	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
.8518–1.0072	1	0.88	1,877	0.14
	2	0.88	4,507	0.14
	3	0.88	4,221	0.13
	Total	0.88	10,605	0.14
1.0072–1.1713	1	0.86	1,726	0.15
	2	0.87	4,395	0.14
	3	0.87	4,440	0.14
	Total	0.87	10,561	0.14
1.1713–1.3504	1	0.87	1,690	0.15
	2	0.87	4,650	0.14
	3	0.87	4,281	0.14
	Total	0.87	10,621	0.14
1.3504–1.5793	1	0.86	1,824	0.15
	2	0.86	4,839	0.14
	3	0.86	3,987	0.14
	Total	0.86	10,650	0.14
1.5793–1.9409	1	0.85	2,276	0.15
	2	0.85	4,969	0.15
	3	0.87	3,603	0.14
	Total	0.86	10,848	0.14
>1.9409	1	0.86	3,026	0.16
	2	0.85	5,313	0.15
	3	0.86	2,612	0.15
	Total	0.86	10,951	0.15
Total	1	0.87	27,279	0.15
	2	0.87	42,869	0.14
	3	0.88	36,764	0.14
	Total	0.87	106,912	0.14

Table C2: Grade 5 Middle Quartiles

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Amount of Dialog

Ratio of Flesch-Kincaid to Grade Equivalent	Amount of Dialog (1 = low, 3 = high)	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
<.4022	1	0.89	3,923	0.14
	2	0.93	5,049	0.11
	3	0.93	5,061	0.11
	Total	0.92	14,033	0.12
.4022–.5431	1	0.89	1,598	0.13
	2	0.91	6,679	0.12
	3	0.92	5,712	0.11
	Total	0.91	13,989	0.12
.5431–.6298	1	0.89	1,254	0.13
	2	0.90	6,611	0.12
	3	0.91	6,183	0.12
	Total	0.90	14,048	0.12
.6298–.7019	1	0.89	1,073	0.14
	2	0.89	6,395	0.12
	3	0.90	6,572	0.12
	Total	0.90	14,040	0.12
.7019–.7694	1	0.88	1,100	0.14
	2	0.89	6,283	0.12
	3	0.90	6,790	0.12
	Total	0.89	14,173	0.12
.7694–.8433	1	0.86	1,262	0.14
	2	0.89	6,278	0.13
	3	0.90	6,654	0.12
	Total	0.89	14,194	0.12
.8433–.9289	1	0.86	1,606	0.14
	2	0.88	6,553	0.13
	3	0.89	6,058	0.12
	Total	0.89	14,217	0.13
.9289–1.0436	1	0.86	2,267	0.14
	2	0.88	6,846	0.13
	3	0.89	5,080	0.13
	Total	0.88	14,193	0.13
1.0436–1.2289	1	0.84	3,088	0.15
	2	0.87	7,194	0.13
	3	0.88	3,813	0.13
	Total	0.86	14,095	0.13

Table C2: Grade 5 Middle Quartiles (continued)

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and Amount of Dialog

Ratio of Flesch-Kincaid to Grade Equivalent	Amount of Dialog (1 = low, 3 = high)	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
>1.2289	1	0.84	4,658	0.15
	2	0.85	7,726	0.14
	3	0.86	1,672	0.13
	Total	0.85	14,056	0.14
Total	1	0.86	21,829	0.14
	2	0.89	65,614	0.13
	3	0.90	53,595	0.12
	Total	0.89	141,038	0.13

Appendix D: High Interest/Low Readability Study

Many publishers produce books designed for older, reluctant readers. These books have high-interest content but a low readability level. Balancing these opposing requirements is a complex task that may affect the semantic and syntactic characteristics of the text. This study sought to explore the extent to which standard readability formulas accurately identify the level of high-interest/low-readability books compared to other books by comparing comprehension scores on for high/low books and standard books of similar difficulty.

Data source

The data for this study came from the AR Reading Practice Database. The analysis included 1,363 students in grade 7 for whom STAR Reading test data and AR quiz data for the 1997–1998 school year were available. Out of all the seventh-grade students, 355 read high/low books. These students read 189 different high/low books for a total of 1,304 student-book pairs. Student-book pairs were the unit of analysis for this study. Books with quiz scores less than 60% were excluded from the analysis because they may have been too difficult for the students who read them, and the results may reflect guessing rather than purposeful effort. Table D1 shows the number of students and books at each grade level.

Table D1: Students and Books at Each Grade Level (High Interest/Low Readability Study)

Student Grade	Number of Students	Number of Different High/Low Books	Total Student-High/Low Book Pairs
7	355	189	1,304

Method

The method for this study was comparable to those found in Appendices A, B, and C. Within each subgroup, the AR quiz scores on high/low books were compared to scores on all other books.

Results

Students tended to score higher on standard books compared to high/low books. This suggests that traditional measures of readability for high/low books may be too low compared to standard books. This tendency is more apparent among readers in the bottom quartile and middle two quartiles. This finding is tentative, however, because the differences in percent correct are small and there are few student-book pairs in some subgroups. A subset of the results is shown in Table D2 (next page).

One possible explanation for the lower comprehension scores on high/low books is that they are frequently nonfiction, which is somewhat more difficult for students to understand than fiction. Among these seventh graders, most were reading high/low fiction books. Splitting the data into fiction and nonfiction still shows a slight tendency for students to score lower on high/low books regardless of whether they are fiction or nonfiction. The results are not sufficiently compelling to draw firm conclusions.

Table D2: Grade 7 Middle Quartiles

Percent Correct on AR Quiz by Flesch-Kincaid/Grade Equivalent and High/Low Versus Other Books

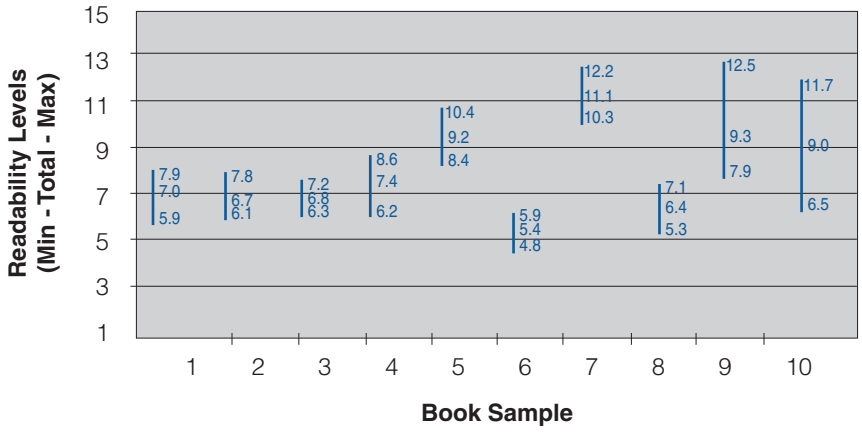
Ratio of Flesch-Kincaid to Grade Equivalent	High/Low?	Mean Percent Correct	Number of Student-Book Pairs	Standard Deviation of Percent Correct
<.4185	N	0.91	677	0.12
	Y	0.90	282	0.11
	Total	0.90	959	0.12
.4185–.4920	N	0.90	870	0.11
	Y	0.88	117	0.12
	Total	0.90	987	0.12
.4920–.5392	N	0.88	929	0.12
	Y	0.85	80	0.13
	Total	0.88	1,009	0.13
.5392–.5808	N	0.90	957	0.12
	Y	0.87	57	0.13
	Total	0.90	1,014	0.18
.5808–.6230	N	0.89	991	0.12
	Y	0.86	50	0.13
	Total	0.89	1,041	0.12
.6230–.6707	N	0.89	984	0.13
	Y	0.83	36	0.14
	Total	0.89	1,020	0.13
.6707–.7276	N	0.88	1,005	0.13
	Y	0.83	37	0.15
	Total	0.88	1,042	0.13
.7276–.7998	N	0.88	987	0.12
	Y	0.87	35	0.14
	Total	0.88	1,022	0.13
.7998–.9225	N	0.87	974	0.13
	Y	0.72	25	0.13
	Total	0.87	999	0.13
>.9225	N	0.84	937	0.14
	Y	0.76	52	0.12
	Total	0.83	989	0.14
Total	N	0.88	9,311	0.13
	Y	0.86	771	0.13
	Total	0.88	10,082	0.13

Appendix E: Full-Text Sampling Study

In the past, readability analyses generally used only samples of text rather than whole books. This procedure may have introduced sampling error due to intratext variance in the readability of books. This study sought to explore to what extent analyzing a sample of text rather than the full text might be a source of error in readability analysis.

The initial aim of the study was to analyze intratext variance of different sample sizes from books. The study showed that readability levels vary throughout a book and result in unpredictable ranges of variability in readability levels when only a portion of the book is used in the reading-level calculations (see Figure E1).

Figure E1: Analysis of Within-Text Readability of 10 Books Based on 1/8 Book Samples



Readability information on 10 books is represented in Figure E1. Book 1, *20,000 Leagues Under the Sea*, has a readability of grade 7.0 when the entire text is considered. When the book is divided into eights and each section analyzed, the range of readability for the sections goes from a low of grade 5.9 to a high of 7.9. As will be discussed later, the range of readability increases even more when smaller sections of text are analyzed.

The within-text differences in readability are not simply an artifact of measurement, but reflect meaningful differences in the text. Passages in a book that contain a high proportion of description, action, introspection, and the like, typically have a higher readability level because the sentences are longer and the words may be less frequent. The following portion of text from *20,000 Leagues Under the Sea* has a readability level of grade 12.0. Note the length of the sentences and the number of long, relatively unfamiliar words.

> *For several days our work hours were spent in all sorts of experiments, on the degree of salinity in waters of different depths, or on their electric properties, coloration, and transparency, and in every instance Captain Nemo displayed an ingenuity equaled only by his graciousness toward me. Then I saw no more of him for some days and again lived on board in seclusion.*

Passages that are principally dialogue have a lower readability level because the sentences are shorter and contain more common words. This portion of text from the same book has a readability level of grade 4.9. The text comprises short sentences, and most of the words are very high frequency and thus are familiar to most readers.

> *"The hotel will keep them for us."*
>
> *"What about master's live babirusa?"*
>
> *"They'll feed it during our absence. Anyhow, we'll leave instructions to ship the whole menagerie to France."*
>
> *"Then we aren't returning to Paris?" Conseil asked.*

Data source and method

Ten popular books of more than 100 pages each were selected for this study. Text samples were taken from each book representing various fractions of the book as shown in Table E1. Flesch-Kincaid readability levels were calculated for each of the samples. Then, all of the samples from each book were combined and a Flesch-Kincaid readability level was calculated for the total book.

Results

Within the 10 books, the Flesch-Kincaid readability varied from 0.7 to 5.4 grades, with the average difference being 3.5 readability levels (and an average standard deviation of .853). Thus, smaller samples taken throughout a book can vary greatly from the reported overall grade level. All of these statistics are higher than those found using the Fry Index with three 100-word samples taken from first, middle, and final thirds of a book. Therefore, it is likely that readability level variability is more a factor of sampling error than of the method used.

The within-text differences in readability of books become even more evident when the size of the analyzed text segment is reduced and the number of samples is increased. In Table E1, the minimum and maximum readability levels of different segments of text are reported. The first column represents a readability analysis of all the text in a book. The second column shows a readability analysis of the two halves of a book, and the two values correspond to the readability of each half. The subsequent columns represent increasingly smaller portions of text, and the numbers show the readability of the easiest (minimum) and hardest (maximum) portions.

As a rule, the greater the number of segments analyzed, the greater the difference between the minimum and maximum readability measures. Perhaps the most understandable example of this is Book 6, *Black Beauty*, a book that is familiar to many students and adults. When one reads the book, it seems to flow smoothly, and the text seems relatively consistent. If the text is dividend into 128ths and a readability analysis is conducted, however, the segments vary greatly with respect to their semantic and syntactic characteristics. The differences are for the most part unnoticed by the reader, and they reflect typical aspects of telling a story through a novel.

So, if the readability analysis of a text is based on segments, there is a good chance that anomalous segments might be chosen. These segments would not be representative of the book as a whole, and even though the analysis of the text might be conducted accurately, the results would not be representative of the true readability level of the book.

Table E1: Analysis of Within-Text Readability of 10 Books Based on Samples of Varying Length

		Whole	Halves	Quarters	8ths	16ths	32nds	64ths	128ths
(1) *20,000 Leagues Under the Sea*	Max FK	7.0	7.4	7.5	7.9	9.6	11.5	13.5	14.9
	Min FK	7.0	6.7	6.3	5.9	5.6	5.0	4.8	4.2
	FK Range	0	0.7	1.2	2.0	4.0	6.5	8.7	10.7
	Stdevp	0	0.35	0.47	0.06	0.09	1.23	1.46	1.78
	Words	108,219	54,110	27,055	13,527	6,764	3,382	1,691	845

Table E1: Analysis of Within-Text Readability of 10 Books Based on Samples of Varying Length (continued)

		Whole	Halves	Quarters	8ths	16ths	32nds	64ths	128ths
(2) *Aladdin*	Max FK	6.7	6.9	7.1	7.8	8.3	0	0	0
	Min FK	6.7	6.6	6.3	6.1	4.8	0	0	0
	FK Range	0	0.03	0.08	1.7	3.5	0	0	0
	Stdevp	0	0.15	0.29	0.06	0.91	0	0	0
	Words	5,271	2,636	1,318	659	329	0	0	0
(3) *Anna Karenina*	Max FK	6.8	6.8	7.2	7.2	8.0	8.9	9.8	12
	Min FK	6.8	6.8	6.5	6.3	6.0	4.9	4.7	4.4
	FK Range	0	0	0.07	0.09	2.0	4.0	5.1	7.6
	Stdevp	0	0	0.25	0.03	0.58	0.84	1.05	1.30
	Words	351,434	175,717	87,859	43,929	21,965	10,982	5,491	2,746
(4) *Around the Word in 80 Days*	Max FK	7.4	7.6	8.2	8.6	9.7	10.9	12.2	13.9
	Min FK	7.4	7.2	6.4	6.2	5.9	5.3	5.1	4.2
	FK Range	0	0.4	1.8	2.4	3.8	5.6	7.1	9.7
	Stdevp	0	0.20	0.67	0.75	1.15	1.47	1.71	2.09
	Words	62,878	31,439	15,720	7,860	3,930	1,965	982	491
(5) *Robinson Crusoe*	Max FK	9.2	9.2	9.5	10.4	11.2	12.2	13.7	15.6
	Min FK	9.2	9.1	8.8	8.4	7.2	6.7	5.8	4.8
	FK Range	0	0.1	0.7	2.0	4.0	5.5	7.9	10.8
	Stdevp	0	0.05	0.30	0.67	1.07	1.45	1.70	2.09
	Words	121,449	60,725	30,362	15,187	7,591	3,795	1,898	949
(6) *Black Beauty*	Max FK	5.4	5.6	5.7	5.9	6.3	7.1	7.9	9.5
	Min FK	5.4	5.2	5.0	4.8	4.1	3.7	2.8	2.2
	FK Range	0	0.4	0.7	1.1	2.2	3.4	5.1	7.3
	Stdevp	0	0.20	0.25	0.40	0.54	0.75	1.04	1.43
	Words	59,779	29,890	14,945	7,472	3,736	1,868	934	467
(7) *Gulliver's Travels*	Max FK	11.1	11.3	11.7	12.2	12.3	13.8	14.1	14.8
	Min FK	11.1	10.9	10.8	10.3	9.8	9.2	7.7	8.0
	FK Range	0	0.4	0.9	1.9	2.5	4.6	6.4	6.8
	Stdevp	0	0.20	0.35	0.52	0.76	1.26	1.43	1.65
	Words	103,680	51,840	25,920	12,960	6,480	3,240	1,620	810

Table E1: Analysis of Within-Text Readability of 10 Books Based on Samples of Varying Length (concluded)

		Whole	Halves	Quarters	8ths	16ths	32nds	64ths	128ths
(8) Tale of Two Cities	Max FK	6.4	6.6	7.0	7.1	7.4	9.3	12.3	14.1
	Min FK	6.4	6.3	5.7	5.3	5.1	5.1	3.3	3.5
	FK Range	0	0.3	1.3	1.8	2.3	4.2	9.0	10.6
	Stdevp	0	0.15	0.50	0.56	0.68	0.92	1.33	1.72
	Words	136,237	68,119	34,059	17,030	8,515	4,257	2,129	1,064
(9) *House of the Seven Gables*	Max FK	9.3	10.2	10.6	12.5	13.3	13.6	14.9	15.5
	Min FK	9.3	8.6	8.1	7.9	7.7	7.2	6.4	5.2
	FK Range	0	1.6	2.5	4.6	5.6	6.4	8.5	10.3
	Stdevp	0	0.80	0.92	1.43	1.55	1.71	1.93	2.41
	Words	103,258	51,629	25,815	12,907	6,454	3,227	1,613	807
(10) *The Secret Garden*	Max FK	9.0	10.2	11.3	11.7	12.1	12.3	13.4	15.1
	Min FK	9.0	8.0	7.6	6.5	5.7	5.3	5.5	4.3
	FK Range	0	2.2	3.7	5.2	6.4	7.0	7.9	10.8
	Stdevp	0	1.10	1.35	1.52	1.68	1.89	2.13	2.36
	Words	84,861	42,431	21,215	10,608	5,304	2,652	1,326	663

Appendix F: Accelerated Reader Goal-Setting Chart

If you have Accelerated Reader and STAR Reading Real Time, the software will automatically recommend a ZPD and goals for each student. Otherwise, use Table F1 and the guidelines below, or the online **AR Goal Calculator**: http://argoals.renlearn.com/, to help set reading practice goals for your students, based on each student's reading level and the amount of daily reading practice that you provide.

1. **Identify ZPDs.** Identify each student's grade-equivalent (GE) score with a standardized assessment, such as STAR Reading, or estimate a GE based on the student's past performance. The corresponding ZPD is a recommended book-level range for the student. If books in that range seem too hard or easy for a student, choose a new range or create a wider one that better matches the student's abilities.

2. **Set Average-Percent-Correct Goals.** The most important goal for all students is to average at least 85 percent or higher on AR Reading Practice Quizzes. Meeting this goal has significant impact on reading growth. Averages of 90 percent and higher are associated with even greater gains. If a student struggles to maintain the minimum average, talk to the student and find out why. Then decide on a strategy that will lead to success.

3. **Set Point Goals.** The chart shows the number of points students are expected to earn based on GE and time spent reading. These are estimates. Set goals that are realistic for individual students.

Table F1: Accelerated Reader Goal-Setting Chart

Grade-Equivalent Score	Suggested ZPD	60 Min. Daily Practice			35 Min. Daily Practice			30 Min. Daily Practice			20 Min. Daily Practice		
		Points per Week	Points per 6 Weeks	Points per 9 Weeks	Points per Week	Points per 6 Weeks	Points per 9 Weeks	Points per Week	Points per 6 Weeks	Points per 9 Weeks	Points per Week	Points per 6 Weeks	Points per 9 Weeks
Emergent Reader		1.7	10	15	1.0	5.8	8.8	0.9	5.0	7.5	0.6	3.3	5.0
1.0	1.0-2.0	1.7	10	15	1.0	5.8	8.8	0.9	5.0	7.5	0.6	3.3	5.0
1.5	1.5-2.5	1.9	11	17	1.1	6.4	9.9	1.0	5.5	8.5	0.6	3.7	5.7
2.0	2.0-3.0	2.1	13	19	1.2	7.6	11.1	1.1	6.5	9.5	0.7	4.3	6.3
2.5	2.3-3.3	2.3	14	21	1.3	8.2	12.3	1.2	7.0	10.5	0.8	4.7	7.0
3.0	2.6-3.6	2.5	15	23	1.5	8.8	13.4	1.3	7.5	11.5	0.8	5.0	7.7
3.5	2.8-4.0	2.7	16	24	1.6	9.3	14.0	1.4	8.0	12.0	0.9	5.3	8.0
4.0	3.0-4.5	2.8	17	25	1.6	9.9	14.6	1.4	8.5	12.5	0.9	5.7	8.3
4.5	3.2-5.0	3.2	19	29	1.9	11.1	16.9	1.6	9.5	14.5	1.0	6.3	9.7
5.0	3.4-5.4	3.5	21	32	2.0	12.3	18.7	1.8	10.5	16.0	1.2	7.0	10.7
5.5	3.7-5.7	3.9	23	35	2.3	13.4	20.4	2.0	11.5	17.5	1.3	7.7	11.7
6.0	4.0-6.1	4.2	25	39	2.5	14.6	22.8	2.1	12.5	19.5	1.4	8.3	13.0
6.5	4.2-6.5	4.6	28	41	2.7	16.3	23.9	2.3	14.0	20.5	1.5	9.3	13.7
7.0	4.3-7.0	4.9	29	44	2.9	16.9	25.7	2.5	14.5	22.0	1.6	9.7	14.7
7.5	4.4-7.5	5.3	32	48	3.1	18.7	28.0	2.7	16.0	24.0	1.8	10.7	16.0
8.0	4.5-8.0	5.6	34	50	3.3	19.8	29.2	2.8	17.0	25.0	1.9	11.3	16.7
9.0	4.6-9.0	6.3	38	57	3.7	22.2	33.3	3.2	19.0	28.5	2.1	12.7	19.0
10.0	4.7-10.0	6.9	41	62	4.0	23.9	36.2	3.5	20.5	31.0	2.3	13.7	20.7
11.0	4.8-11.0	7.6	46	68	4.4	26.8	39.7	3.8	23.0	34.0	2.5	15.3	22.7
12.0	4.9-12.0	8.3	50	75	4.8	29.2	43.8	4.2	25.0	37.5	2.8	16.7	25.0

Appendix G: Adjustment for Book Length

The adjustment for book length is a weighted average of the ATOS for Text Readability Level and a grade level based on book length. Book-length grade level (BLGL) was determined by examining the relationship between book length and AR book level for 13,200 fiction books with more than 500 words. The following formula describes the relationship between book length and grade level for books with more than 500 words:

BLGL for Books With More Than 500 Words = 0.68 * Ln(Book Length) – 1.87

The relative weights of ATOS for Text and book-length grade level vary based on the length of the book. Table G1 shows the relative weights for different length books.

Table G1: Weights of ATOS for Text and Book-Length Grade Level

Number of Words	ATOS for Text Weight (ATOS Wght)	Book-Length Grade Level Weight (BLGL Wght)
500	.50	.50
5,000	.60	.40
50,000	.80	.20
100,000	.85	.15
250,000	.90	.10

Books with more than 500 words

The 500-word cutoff was chosen because most Reading Recovery books have fewer than 500 words. The next level, 5,000 words, was chosen as the second cutoff because more than half of AR books with readability levels less than grade 5 using the Flesch-Kincaid Readability Formula had between 500 and 5,000 words. Similarly, more than half of AR books with readability levels between grades 5 and 8 had between 5,000 and 50,000 words. The final two cutoffs were chosen because books with more than 50,000 words become more common at levels above grade 8 (e.g., *My Antonia* by Willa Cather) and books with more than 100,000 words become more common at levels above grade 10 (e.g., *War and Peace* by Leo Tolstoy).

A smooth curve was fit through this data so that the weights could be calculated with the following formulas:

ATOS Wght = .0648 * Ln(Book length) + .0894 BLGL Wght = 1 – ATOS Wght

The weighted average of ATOS for Text and book-length grade level produces the ATOS for Books Readability Formula shown here:

ATOS for Books Readability Formula = ATOS Wght * ATOS for Text + BLGL Wght * BLGL

Books with fewer than 500 words

Above, Table G1 shows the weights applied for books with 500 or more words. A book-length adjustment was also developed for books with fewer than 500 words. Early investigations into a separate formula for emergent reader books, using Reading Recovery levels as the criterion, revealed book length as a strong predictor of student comprehension. Rather than develop a separate formula for emergent reader books, the ATOS for Books Readability Formula was modified slightly for books with fewer than 500 words. Because book length was such a strong predictor of Reading Recovery levels, the proportion of ATOS for Books that comes from book-length grade level is 60% while the proportion from ATOS for Text is 40%. The value of book-length grade level ranges from .5 for the shortest books to 2.4 for books with 500 words. The relationship between book length and grade level for books with fewer than 500 words is given by the following formula:

BLGL for Books With Fewer Than 500 Words = .004 * Book Length + 0.4

Appendix H: Conversion: ATOS for Books to Reading Recovery, Guided Reading

Table H1: Conversion of ATOS for Books to Reading Recovery and Guided Reading Scales

ATOS Book Level	Reading Recovery Level	Guided Reading Level*	Guided Reading Grade Level*
.2–.4	1	A	K
.2–.4	2	B	K
.5–.6	3	C	K/1
.5–.6	4	C	K/1
.5–.6	5	D	1
.7–.9	6	D	1
.7–.9	7	E	1
.7–.9	8	E	1
.7–.9	9	F	1
1.0–1.2	10	F	1
1.0–1.2	11	G	1
1.3–1.5	12	G	1
1.3–1.5	13	H	1
1.6–1.9	14	H	1
1.6–1.9	15	I	1
2.0–2.4	16	I	1
2.0–2.4	17	J	2
2.5–2.9	18	J	2
2.5–2.9	19	K	2
2.5–2.9	20	K	2
2.5–2.9	21	L	2
3.0–3.4	22	M3	2
3.4–3.9		N	2/3
3.4–3.9		O	3/4
4.0–4.4		P	3/4
4.0–4.4		Q	4/5
4.5–4.9		R	4/5
4.5–4.9		S	5
5.0–5.4		T	5
5.0–5.4		U	5
5.5–5.9		V	6
6.0–6.9		W,X,Y,Z	6

** From the Fountas & Pinnell Guided Reading Leveling System*

Appendix I: Definition of Sentence-Length Limits

Comparisons of the characteristics of books to the characteristics of the DRP passages upon which ATOS was developed showed that sentence length is considerably more variable in books than in the test passages. The semantic variables, however, tend to be more similar between books and the DRP test passages. Thus, it may be problematic to use a formula developed from the DRP passages on books with much higher or lower sentence lengths than those in the test passages.

To compensate for this difficulty, upper and lower sentence-length limits were defined. When the average sentence length of a book exceeds the upper or lower limits, the value of that book's sentence length is limited to the boundary value plus 25% of the difference between the actual value and the boundary.

To define the sentence-length limits, the relationship between the sentence length portion and the semantic (word length and vocabulary level) portion of ATOS for Text in 263 STAR Reading items was examined. STAR Reading items were used as the basis for this adjustment because their range and variability of sentence length falls between that of the DRP items and that of books. To define the limits of extreme sentence length, the value of the natural log of average sentence length (LnAvgWrd) was regressed on the value of the semantic portion of ATOS for Text for the STAR Reading items. The resulting formula indicates the expected value of sentence length given a book's semantic characteristics:

Predicted LnAvgWrd = .4456 * (semantic part of ATOS) + .8294

The bounds that produced the greatest improvement in the correlation between percent correct and book level relative to reading skill were calculated. When the average sentence length of a book was much longer or shorter than the sentences in the criterion passages, an adjustment was made to the basic formula. The value of the natural log of average sentence length used in the calculation of ATOS becomes:

If LnAvgWrd is greater than upper bound
    then New LnAvgWrd = Upper bound + 25%(LnAvgWrd-Upper bound)
If LnAvgWrd is less than the lower bound
    then New LnAvgWrd = Lower bound + 25%(LnAvgWrd-Lower bound)

Appendix J: Conversion: ATOS for Books to 100-Point, 2000-Point Scales

Table J1: Conversion of ATOS for Books to 100-Point and 2000-Point Scales

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
0.1	10	15
0.2	17	19
0.3	21	23
0.4	24	27
0.5	27	31
0.6	29	35
0.7	30	39
0.8	32	43
0.9	33	47
1.0	34	51
1.1	35	71
1.2	36	92
1.3	37	120
1.4	38	140
1.5	38	157
1.6	39	170
1.7	40	190
1.8	40	202
1.9	41	221
2.0	41	241
2.1	42	261
2.2	42	289
2.3	43	301
2.4	43	329
2.5	44	348
2.6	44	361
2.7	44	381
2.8	45	401
2.9	45	420
3.0	46	440
3.1	46	459
3.2	46	479
3.3	47	491
3.4	47	511
3.5	47	530
3.6	47	549

Table J1: Conversion of ATOS for Books to 100-Point and 2000-Point Scales (continued)

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
3.7	48	561
3.8	48	580
3.9	48	600
4.0	49	619
4.1	49	631
4.2	49	650
4.3	50	669
4.4	50	681
4.5	50	701
4.6	51	720
4.7	51	731
4.8	51	750
4.9	52	769
5.0	52	781
5.1	52	800
5.2	52	811
5.3	53	829
5.4	53	841
5.5	53	860
5.6	54	870
5.7	54	888
5.8	54	899
5.9	55	910
6.0	55	921
6.1	55	931
6.2	56	941
6.3	56	951
6.4	56	969
6.5	57	979
6.6	57	989
6.7	57	999
6.8	57	1009
6.9	58	1019
7.0	58	1029
7.1	59	1040
7.2	60	1052
7.3	60	1061
7.4	60	1070
7.5	60	1080

Table J1: Conversion of ATOS for Books to 100-Point and 2000-Point Scales (continued)

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
7.6	60	1088
7.7	61	1098
7.8	61	1109
7.9	61	1117
8.0	61	1121
8.1	61	1130
8.2	62	1140
8.3	62	1149
8.4	62	1152
8.5	62	1161
8.6	62	1170
8.7	62	1178
8.8	63	1188
8.9	63	1192
9.0	63	1201
9.1	63	1211
9.2	63	1222
9.3	63	1237
9.4	64	1242
9.5	64	1258
9.6	64	1265
9.7	64	1272
9.8	64	1279
9.9	64	1286
10.0	65	1293
10.1	65	1300
10.2	65	1307
10.3	65	1314
10.4	65	1321
10.5	65	1328
10.6	65	1335
10.7	66	1342
10.8	66	1349
10.9	66	1356
11.0	66	1364
11.1	66	1371
11.2	66	1378
11.3	66	1385
11.4	67	1392

Table J1: Conversion of ATOS for Books to 100-Point and 2000-Point Scales (continued)

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
11.5	67	1399
11.6	67	1406
11.7	67	1413
11.8	67	1420
11.9	67	1427
12.0	67	1434
12.1	67	1441
12.2	68	1448
12.3	68	1455
12.4	68	1462
12.5	68	1469
12.6	68	1476
12.7	68	1483
12.8	68	1490
12.9	68	1497
13.0	69	1504
13.1		1511
13.2		1518
13.3		1525
13.4		1532
13.5		1539
13.6		1547
13.7		1554
13.8		1561
13.9		1568
14.0		1575
14.1		1582
14.2		1589
14.3		1596
14.4		1603
14.5		1610
14.6		1617
14.7		1624
14.8		1631
14.9		1638
15.0		1645
15.1		1652
15.2		1659
15.3		1666

Table J1: Conversion of ATOS for Books to 100-Point and 2000-Point Scales (concluded)

ATOS Book Levels	ATOS 100-Point Scale Values	ATOS 2000-Point Scale Values
15.4		1673
15.5		1680
15.6		1687
15.7		1694
15.8		1701
15.9		1708
16.0		1715
16.1		1722
16.2		1730
16.3		1737
16.4		1744
16.5		1751
16.6		1758
16.7		1765
16.8		1772
16.9		1779

References

Bormuth, J. R. (1969). *Development of readability analyses*. (Final Report, Project No. 7-0052, Contract No. OEC-3-7-070052-0326). Washington, DC: USOE, Bureau of Research, HEW.

Brennan, R. L. (2004). LEGS [Computer program]. Iowa City: University of Iowa, Center for Advanced Studies in Measurement and Assessment (CASMA).

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

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

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

CTB/McGraw-Hill. (1992). *California Achievement Test, Form A*. Monterey, CA: Author.

CTB/McGraw-Hill. (1997). *Comprehensive Test of Basic Skills (Terra Nova), Form A*. Monterey, CA: Author.

Dale, E. (1931). A comparison of two word lists. *Education Research Bulletin, 10*, 484–489.

Dale, E., & Chall, J. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*(1), 11–20, 28; *27*(2), 11–28.

Dale, E., & O'Rourke, J. (1976). *The living word vocabulary: The words we know*. Elgin, IL: Field Enterprises Educational Corporation.

Dolch, E. W. (1936). A basic sight vocabulary. *Elementary School Journal, 36*, 456–460.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag.

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

Fountas, I. C., & Pinnell, G. S. (1999). *Matching books to readers: Using leveled books in guided reading*. Portsmouth, NH: Heinemann.

Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11*, 513–516, 575–578.

Gray, W. S., & Leary, B. E. (1935). *What makes a book readable?* Chicago: University of Chicago Press.

Harcourt Educational Measurement. (1993). *Metropolitan Achievement Test, Form S*. San Antonio, TX: Author.

Harcourt Educational Measurement. (1995). *Stanford Achievement Test, Form S*. San Antonio, TX: Author.

Harris, A. J., & Jacobson, M. D. (1972). *Basic elementary reading vocabularies*. New York: Macmillan.

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

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Lively, B. A., & Pressey, S. L. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision, 9*, 389–398.

Mathews, John H. (1992). *Numerical methods for mathematics, science, and engineering* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

McLaughlin, G. H. (1969). SMOG grading—A new readability formula. *Journal of Reading, 22*, 639–646.

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2011). *Measures of text difficulty: Testing their predictive value for grade levels and student performance.* Technical Report to the Gates Foundation (also to be submitted for publication). Retrieved from http://www.ccsso.org/Documents/2012/Measures ofText Difficulty_final.2012.pdf

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. (Original work published 1960.)

Renaissance Learning. (2012a). *Text complexity: Accurate estimates and educational recommendations*. Wisconsin Rapids, WI: Author. Available online from http://doc.renlearn.com/KMNet/R00548821C95879F.pdf

Renaissance Learning. (2012b). *What kids are reading: The book-reading habits of students in American schools*. Wisconsin Rapids, WI: Author. Available online from http://www.renlearn.com/whatkidsarereading/

Riverside Publishing Company. (1993). *Iowa Test of Basic Skills*, *Form K*. Itasca, IL: Author.

Roller, C. M. (1990). The interaction of knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly, 25*, 78–89.

Spache, G. D. (1968). *Good reading for poor readers* (2nd ed.). Champaign, IL: Garrard Publishing Company.

Stenner, A. J., & Burdick, D. S. (1997). *The objective measurement of reading comprehension: In response to technical questions raised by the California Department of Education Technical Study Group*. Durham, NC: MetaMetrics.

Taylor, S. E., Frackenpohl, H., & White, Catherine E. (1979). *A revised core vocabulary*. New York: EDL/McGraw-Hill.

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

Vogel, M., & Washburne, C. (1928). An objective method of determining the grade placement of children's reading material. *Elementary School Journal, 28*, 373–381.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Zakaluk, B. L., & Samuels, S. J. (1988). *Readability: Its past, present, and future*. Newark, DE: International Reading Association.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

About the Author

Michael Milone, Ph.D., is a research psychologist and award-winning educational writer and consultant to publishers and school districts. He earned a Ph.D. in 1978 from The Ohio State University and has served in an adjunct capacity at Ohio State, the University of Arizona, Gallaudet University, and New Mexico State University. He has taught in regular and special education programs at all levels, holds a Master of Arts degree from Gallaudet University, and is fluent in American Sign Language. Milone served on the board of directors of the Association of Educational Publishers and was a member of the Literacy Assessment Committee and a past chair of the Technology and Literacy Committee of the International Reading Association. He has contributed to both readingonline.org and *Technology & Learning* magazine on a regular basis. Over the past 30 years, he has been involved in a broad range of publishing projects, including the SRA reading series, assessments developed for Academic Therapy Publications, and software published by The Learning Company and LeapFrog. He has completed 34 marathons and 2 Ironman races.

Renaissance™ Learning
Advanced Technology for Data-Driven Schools

L2580.0512.RN.1M
R42508