Readability

John J. Pikulski, Ph.D.

Professor

University of Delaware

Readability: A definition.

Readability continues to be among the most discussed, misunderstood, and misused concepts in reading. It is all too commonly, but erroneously, thought to be a precise numerical score, obtained through the use of readability "formulas," that indicates the level of difficulty of a text. In such an oversimplified view of readability, the degree of difficulty resides completely in the text. In a very global sense this view has great intuitive appeal; some texts clearly seem inherently more difficult than others. For example, Dr. Seuss books certainly have a lower level of readability than Shakespearean plays. However, within narrower ranges of texts the readability or difficulty of texts is not as clear. Is a science text measured to be at fourth-grade readability really easier to read than a fifth-grade readability social studies text? Might one child find the science text easier, while another might be more successful with the social studies text?

A more reasonable definition of readability that is in keeping with more recent research and theory is *the level of ease or difficulty with which text material can be understood by a particular reader who is reading that text for a specific purpose.* Readability is dependent upon many characteristics of a text and many characteristics of readers.

Thus, one important characteristic of a useful, informed definition of readability is that it reflects the *interactive* nature of the construct. As the *Literacy Dictionary* (Harris & Hodges, 1995) points out, "Text and reader variables interact in determining the readability of any piece of material for any individual reader." (p. 203) Likewise, Jeanne Chall, coauthor of one of the most popular readability formulas, endorses an interactive view of readability. She concluded, "The purpose of readability assessment is to effect a 'best match' between intended readers and texts . . . thus, optimal difficulty comes from an interaction among the text, the reader, and his/her purpose for reading" (Chall & Dale, 1995, p. 45-46). Zakulek and Samuels (1958) are quite critical of readability measures, concluding they "are inadequate because they consider only one source of information—that contained on the printed page."

The same text materials may be very easy for one reader yet extraordinarily difficult to another. For example, prior knowledge will greatly influence how well a

reader can understand text dealing with a particular topic. While a coastal engineer might easily read a technical report on the topic of tidal erosion, most of us would find the same text incomprehensible.

When considering readability, it is important to consider the complexity of the text. As early as 1935 Gray and Leary listed 288 factors that influence how easy or how difficult text might be to read. It would indeed be very inappropriate to expect that a single number or index, such as those derived from formulae, could *precisely* measure the complex relationships that exist between texts and readers.

That readability is *complex* and not a simple, *precise* measure is reflected in a statement endorsed by the Board of Directors of the International Reading Association. The statement reads, "Many factors enter into determining the readability of materials, including the syntactic complexity of sentences, density of concepts, abstractness of ideas, text organization, coherence and sequence of ideas, page format, length of type line, length of paragraphs, intricacy of punctuation, and the use of illustrations and color. In addition, research has shown that student interest in the subject-matter plays a significant role in determining the readability of materials" (1985).

The International Reading Association points to two major clusters of factors that are not assessed by readability formulas. The first cluster emphasizes what Rubin (1985) has termed *conceptual* readability, which relates to factors such as "density of concepts, abstractness of ideas, text organization and coherence and sequence of ideas." The second cluster of factors including "page format, length of type line, length of

paragraphs, . . . and the use of illustrations and color" all refer to what Gray & Leary (1935) referred to as format or design factors in what is now regarded as a classic study of readability (Klare, 1984). They used the combined judgment of librarians, publishers, teachers, and directors of adult education to identify the major factors contributing to text difficulty. Of the 288 factors that were identified, format or design factors were found to be among the three most important clusters of contributors to making a text easier or more difficult to read.

Thus, in addition to traditional measures of semantic (vocabulary challenge) and syntactic (sentence complexity as estimated by sentence length) difficulty, which are the usual factors measured by readability formulas, thoughtful approaches to readability must also include, at minimum, both conceptual and design approaches to text difficulty.

Readability Formulas.

Most authorities (e.g., Klare, 1984) trace the beginnings of readability formulas to the publication of Thorndike's *The Teacher's Word Book* in 1921. By 1973 more than 200 language variables had been tried in almost 200 different readability formulas. The vast majority of these formulas are no longer in use.

Examples of readability formulas that have long-standing use are the Spache Formula, designed for use in children's materials in grades 1–3; the Dale-Chall Formula, designed for use with materials beyond third-grade level; and the Harris-Jacobsen Formula, which can be applied to materials from first through eighth-grade

levels. All three of these formulas consider only two major text factors: the average number of words per sentence and the number of words in the sample that are not found on a list of "familiar" (or "easy") words. These, like all commonly used formulas, certainly do not come to grips with the enormous complexity of readability; however, they are relatively easy to calculate and have some general ability to broadly predict text difficulty. Thus, they have some practicality, but they should not be regarded as precise, final measures.

Chall (1984), one of the foremost authorities on readability, points out that most readability researchers have always acknowledged the need to consider factors beyond those measured in formulas. Klare (1995) warns, "Do not rely on formulas alone in selecting materials; seek the opinion of experts or get reliable consensus opinions to examine characteristics that formulas cannot predict . . . ." (p. 205).

Thus, the results of available formulas must be looked at as very tentative and potentially misleading. The IRA statement referred to earlier pointed out that "Readability formulae are simply insufficient as a guide to matching students with books and other instructional material. It is not what readability formulae measure that concerns us; it is what they do NOT measure."

Computer Generated Readability Scores.

There may be a need to be particularly cautious about what Duffelmeyer (1985) has called the "aura of precision" created through the computer application of readability

formulas. He found that what appeared to be a promising computer formula seriously overestimated the difficulty level of materials; he concluded that "the computer merely does formulas faster and less tediously. . . . The marriage of experience and common sense is just as important as it ever was, particularly as far as the conceptual difficulty of passages is concerned."

Very seldom do different people using the same formula software and computers arrive at the same score for the same reading selection. In spite of the seeming simplicity and objectivity of computer-based or other readability formulas, numerous human judgments must be made when using them: Exactly what section or sections of text should be chosen for inclusion in the calculation? Should items in exercises following a section be used? Should brief stage directions in a play be counted? Clearly, the use of computer software in determining readability is much more complicated than just typing the text, choosing a formula, and pushing a button.

One of the most enlightening studies related to readability and the lack of precision in formulas is reported by Beck, et al (1984). In that study Beck and her associates rewrote two passages in order to make them more comprehensible and coherent. That they were successful with their revisions is reflected in the fact that when groups of students read the original and rewritten versions, they scored significantly higher on two separate measures of comprehension after reading the rewritten, as compared to the original version. However, when readability scores were calculated for the original and rewritten passages, in both cases the rewritten passages were rated by the readability measure as one

grade level *higher* than the original passages! Whereas actual comprehension measures showed that the rewriting made the passages easier, readability formulas indicated that they had become harder. The researchers concluded that, depending on the use to which they are put, ". . . readability formulas are at best useless, and at worst misleading . . . . "

Bruce, Rubin, and Star (1981) provide a succinct, critical evaluation of readability formulas and conclude: "Unfortunately, readability formulas just don't fulfill their promise. . . . From a theoretical point of view, they ignore or violate much of current knowledge about reading and the reading process. Second, their statistical bases are shaky, being at once poorly supported mathematically and difficult to generalize. Finally, as practical tools either for matching children and texts or for providing guidelines for writers, they are totally inappropriate." It seems necessary to conclude that, to date, no objective, accurate way of measuring the concept of readability has been devised.

The Lexile™ Framework.

A more recent application of traditional readability formulas is known as the Lexile™ Framework. This approach to readability attempts to address the interaction that exists between a reader and text in terms of a text's difficulty. Based on the Lexile analysis of many currently available standardized reading tests, a reader who takes one of those tests can have her or his performance translated into a Lexile score. Lexile scores usually are reported in units of 100, theoretically ranging from zero to above 2000 Lexiles, or a range from very

beginning reading through post graduate level reading.

Reading texts are also analyzed through Lexile procedures and each text is also assigned a Lexile score. In other words student performance on a standardized test and the readability analyses of texts are translated into the same units of measurement–Lexiles.

In theory, students can then be directed to texts that are purported to be at an appropriate level of challenge. For example, a student whose performance on a standardized reading measure resulted in a Lexile score of 500, according to Lexile proponents, would be expected to read texts assigned a similar score (500) with 75% comprehension, a level that is challenging in terms of vocabulary, content, and reading skills, but not likely to be frustrating to that reader; the student scoring at the 500 Lexile level would be expected to read texts with lower Lexile ratings with greater than 75% comprehension and those with higher ratings with lesser comprehension. Lists of thousands of texts with their Lexile ratings are available for purchase from MetaMetrics, the publisher that produces these ratings.

As with traditional readability formulae, Lexile analyses of standardized reading tests and for text materials are based on only two factors: the frequency with which the words in the test or text are commonly used in English and by sentence length . . . two factors which are easy to calculate, but which fall short of addressing the multifaceted nature of conceptual and text design readability. The Lexile framework does not include provision for addressing factors such as motivation, interest, graphic supports, text structure or other factors that broadly affect text difficulty and student reading performance.

Leveling of Texts as an Alternative to Readability.

The concept of leveling texts is probably earliest and most prominently associated with the early intervention program *Reading Recovery*. Peterson (1991), whose chapter dealing with the leveling of texts is widely used and cited, indicates that the concept of leveling texts, as a way of describing gradients of text difficulty, originated with Marie Clay's pioneering work with *Reading Recovery* in New Zealand. Peterson notes that, "The hierarchy of levels was developed in New Zealand based on teachers' observations of their students' progression through the *Ready to Read* books, a graded series of little books read by all beginning readers designed to reflect the experiences of children growing up in New Zealand. Thus, these books came to serve as benchmarks for comparison in assessing levels of new books." (p. 123).

Since *Reading Recovery* is exclusively a first-grade intervention program, the *Reading Recovery* levels apply only to first-grade materials. They range from very simple, easily memorized level 1 books to level 20 books, which are described as "roughly corresponding to those found in typical basal reading programs for children at the end of first grade year." (p. 123).

As suggested by Peterson's description, the level of a text's difficulty is a judgment made by a teacher working with a student (*Reading Recovery* is a one-on-one tutoring program). It is emphasized that levels are approximations and not absolute designations. Judgments about levels of text difficulty go beyond the statistical nature of readability formulas that tend to rely on measures of semantic (usually word length or word

frequency) and syntactic (usually sentence length) features. Leveling criteria represent an attempt to address more of the complex features which make texts easier or harder to read such as the repetition and predictability of words and language patterns, the degree to which illustrations support the text, familiarity of the story or its content and placement, physical design of the text, density of the text on a page, etc.

Peterson's (1991) description of the twenty *Reading Recovery* levels is widely cited and used. She describes the levels in groups of 4. For example, in outline form, books that would be designated as falling in levels 1–4 would have the following characteristics:

- consistent placement of print
- repetition of 1-2 sentence patterns (1-2 word changes)
- oral language structures
- familiar objects and actions
- illustrations provide high support

In contrast the texts that would be designated as mid first grade (levels 9-12) would have the following characteristics:

- repetition of 3 or more sentence patterns
- or, varied sentence patterns (repeated phrases or refrains)
- blend of oral and written language structures
- or, fantastic happenings in framework of familiar experiences
- illustrations provide moderate support

The concept of "leveling" of books was modified and extended by Fountas & Pinnell (1996). They, like Clay & Peterson, placed a great deal of emphasis on teacher judgment of a text's difficulty and on multiple characteristics that affected difficulty such as

length of text, size and layout of print, vocabulary and concepts, language structure, text structure and genre, predictability and pattern of language, and illustration support.

While the Clay & Peterson criteria were designed to be applied to texts that were to be used in individual, first-grade tutoring sessions, the Fountas & Pinnell criteria were designed to be applied to texts being used for guided reading with small groups of children at several grade levels. Fountas & Pinnell concluded that while the fine gradient of twenty levels was appropriate for teachers working closely with individual children, such fine distinctions were neither practical nor useful when working with groups of children. The original Fountas & Pinnell system included sixteen levels encompassing kindergarten through third grade with nine levels from kindergarten through first grade (Level A-I), four levels for second grade (J-M), and three levels for third grade. More recently, Fountas & Pinnell have extended this system through grade 6.

Cooper, Boschken, McWilliams & Pistochini (1998) have also extended the concept of leveling of texts for use in grades 3–8. These criteria take into account "the amount of print on the page, picture support, complexity of storylines, and/or complexity of facts presented." (p. 3).

While leveling of texts lacks the objectivity and simplicity of readability formula, it does attempt to deal with the multiple, complex dimensions that make texts easier or more difficult to read.

Readability In Houghton Mifflin Reading.

In the preparation of *Houghton Mifflin Reading,* authors and editors operated with a complex, interactive, conceptual, and design definition of readability.

In the first four themes of grade 1, materials were developed that carefully, yet flexibly, control the level of challenge of the texts. Recognized authors of children's literature were contracted to write engaging stories that allow children to apply the phonics/word identification skills and the high-frequency words they were learning. In those first four themes, then, there is a very controlled, steady progression from selection to selection in terms of level of text challenge. These carefully prepared beginning reading materials are designed to help students quickly develop a set of phonics and word identification skills that will allow them to read successfully, both independently or with teacher guidance, texts that are less controlled than these very beginning reading texts. Beginning with theme 5 of grade 1, through grade 6, all anthology selections are authentic literature. The selections were still chosen with an eye to creating a gradient of difficulty at each grade level and from one grade level to the next, but they were also chosen for their literary merit and were grouped according to thematic content.

In the vast majority of the themes at second grade and beyond, selections move from being less to more challenging so that there is a repeated scaffolding effect of having students read an easier theme-related text before moving to a more difficult one, and then starting the next theme at a somewhat easier level. The ordering also took into

consideration the interaction of the theme selections to the development and integrity of the content of the theme and the need to order selections so that they included variety that would stimulate student interest. For example, at each grade level there is a careful balance of fiction and informative, nonfiction texts.

Informative texts are characteristically more challenging both in terms of readability formula scores and conceptual readability, but it would obviously be inappropriate to cluster all nonfiction pieces at the end of a grade level. However, even in instances where there is formula and conceptual readability variation that interrupts a precise gradient of difficulty, the editorial design and format features of the selection, were used to mitigate the text's difficulty. In addition, modest variation in text difficulty is viewed as an instructional strength since in real world reading students will sometimes meet a more challenging text followed by an easier text.

To determine the text readability of selections used in *Houghton Mifflin Reading*, a multifaceted approach was taken, involving the use of: expert judgment; field testing; affecting readability through instructional design; affecting readability through instructional support; and use of readability formulas and leveling.

Expert Judgment.

Houghton Mifflin authors and editors, who have strong backgrounds in children's literature, were responsible for initially making judgments about selections that might become part of *Houghton Mifflin Reading* and about the levels at which these selections should be placed.

In addition, Houghton Mifflin periodically conducts a national survey of children's librarians to determine which literary works children most enjoy and the developmental level at which the selections are most appropriately read. All selections that have literary merit are then critically read by children's literature experts, authors, editors, and perhaps most important, by classroom teachers. Rubin (1985) cites evidence that teachers are good judges of readability; therefore, their opinions are weighed heavily in the selection and placement of texts in the program. The multiple steps of this initial review of literary works based on expert opinion provide a wealth of information about the developmental level at which selections should be placed.

Field Testing.

Periodic surveys of children's library book choices and librarians' judgments about the newest literature offer a combination of expert judgment and field testing information that is used in choosing and placing reading selections in *Houghton Mifflin Reading*. Selections that are initially judged suitable for placement at a particular level are further field tested by authors, editors, educational consultants, and classroom teachers at that grade level to ensure that selections are very likely to be understood and enjoyed by children at a given developmental level. At Houghton Mifflin we take seriously Rubin's (1985) advice that "Publishers must choose texts that are appropriate for children with a wide variety of backgrounds. . . . To do the best job of text selection, publishers must participate in the kind of process teachers experience: watching individuals and groups of children reading and

discussing candidate texts. Analyzing readers' descriptions of their successes and difficulties can provide the most useful, detailed information to guide text selection and revision." (p. 74).

Readability Formulas and Leveling.

For selections under consideration for inclusion in *Houghton Mifflin Reading*, the editors applied the Harris-Jacobson formula to materials for grades 1-3 and both the Dale-Chall and Harris-Jacobson formulas to materials beyond grade 3. The texts were also leveled using the Fountas & Pinnell criteria and the Cooper, et al, criteria for grades three through six. Those results were factored in with the conclusions drawn from judgments of experts and field testing results, along with the planned instructional design and instructional support for a given selection. All of these factors were considered in determining the ultimate placement for selections.

Affecting Readability Through Instruction Design.

Student anthologies were carefully designed to promote the development of reading skills and strategies, to promote comprehension of the selections, and to develop interest and literary appreciation. Factors such as careful placement of text on pages, choice of font, size of print, the absence of overprinting, use of illustration and other graphic supports were all very carefully planned and employed.

The student anthologies, at all grade levels, also contain special features designed to address text readability and enhance comprehension of the selections. A two-page "Get Set" spread precedes each major selection. These two pages, which are part of the student anthology, use a combination of illustrations and text to build background, concepts, and vocabulary needed for understanding the selection.

The student anthology also contains a strategy focus preceding each major selection. Important syntheses of reading research (e.g., National Reading Panel 2000) make it clear that the application of reading strategies enhances text understanding.

The student anthologies at all grade levels also include a "Responding" page which consists of questions and suggestions for stimulating thinking and discussion. There is unequivocal evidence (e.g., Gambrell, 1996) that active discussion moves comprehension and understanding to a higher level.

Affecting Readability Through Instruction Support.

Houghton Mifflin Reading provides teachers with carefully planned instructional activities that prepare students for reading the text material that is part of the program. The ability of readers to deal with that text is significantly enhanced by what the teacher does in *Houghton Mifflin Reading* before the selection is read.

Prior knowledge (schema) very seriously affects a reader's comprehension. Each selection in *Houghton Mifflin Reading* is carefully analyzed by authors and editors to identify its major concepts. Concepts that are judged likely to create comprehension challenges are indicated for the teacher; in addition to the "Get Set" materials in the

pupil text, lessons or activities are suggested in the Teacher's Edition to prepare students to embark upon the reading with the background necessary for comprehension. Lessons or activities are also outlined for the teacher that help to ensure that the background knowledge that students bring with them is activated and brought to bear on their reading.

Another strong instructional support feature to enhance the comprehensibility of selections in *Houghton Mifflin Reading* is the use of graphic organizers throughout the program. Following teaching suggestions for focusing students on the application of a key reading strategy, students are consistently introduced to a graphic organizer to support their reading and understanding. Graphic organizers included in the program include K-W-L charts, story maps, predictions, details, sequence of events, etc. The National Reading Panel Report (2000) concludes that use of such graphic organizers enhances comprehension.

An Example of Selection Readability from *Houghton Mifflin Reading.*

It might be helpful to look at an example of how the concept of readability is addressed in a specific selection from *Houghton Mifflin Reading.* This selection, *Lou Gehrig: The Luckiest Man*, by the award winning author, David Adler, is from the fifth theme, Heroes, from the fourth-grade anthology.

In addition to the general review by program authors and editors, the selection and its placement was very carefully, critically reviewed by seventeen reading specialists/supervisors and thirteen classroom teachers, representing a very wide geographical range of the United States. Many of these reviewers field tested this selection with some or all of their students.

David Adler, the author of this selection, has impressive credentials, having written over one hundred nonfiction texts.

Lou Gehrig is the third biography in the "Heroes" Theme. Therefore, students will have had two recent, previous experiences with this particular genre, which should facilitate their comprehension of this third biographical piece.

Perhaps the most blatant limitation of a readability formula measure that does not address the interactive nature of readability as it applies to this selection lies in the different levels of information students may already have about baseball and the interest students have in the subject.

The concepts dealt with in the selection–what it takes to be a sports hero and to show good sportsmanship–are likely to be fairly familiar to most fourth graders, as is the baseball terminology such as "shortstop", "first baseman" and "fielding". However, the "Get Set" piece deals with these central concepts, and the vocabulary, for students who may have recently come to the United States, and for those who have limited interest in sports. The Teacher's Edition also offers additional background and motivation-building suggestions, and alerts the teacher to a CD-ROM resource that provides even more background building and vocabulary support, along with summaries of the selection in both Spanish and English.

Additional preparation for reading the selection comes from the strategy focus and through the availability of a graphic organizer which supports the recognition of facts and opinions.

The readability of the selection will also be affected by suggestions in the Teacher's Edition for previewing and reading the selection in sections, with discussion to enhance comprehension after each section. The Teacher's Edition strongly encourages teachers to adjust the level of support to the degree that the students reading the selection need it in order for the selection to be readable.

Readability formulas also fail to factor in the enormous amount of graphic support that is provided by the interesting and informative illustrations that accompany this selection. Nor would a formula consider that the selection is made more accessible by having a carefully designed, reasonable amount of text on most pages, with occasional lighter loads of text to refresh the reader's stamina.

Because the Lou Gehrig selection is expertly written and illustrated and tells an engaging story about a very courageous person, it is almost certain to be very interesting to fourth-grade students, and it is well established that interest significantly affects the comprehensibility and readability of the text.

This example, using just one selection, vibrantly illustrates the complex, interactive nature of an instructionally useful concept of readability. To cite a single readability formula score would be a serious oversimplification and possibly a distortion of the real challenge of this selection.

Conclusions.

Readability formulas, unquestionably, have some utility. They have reasonable utility and predictability as a *starting point* for determining the level of challenge of a text. However, it is imperative to consider expert judgment, field testing, knowledge of students, and instructional and text design characteristics in determining readability. Therefore, readability scores should not be viewed as a summary statement about a selection outside the context of those other critically important dimensions.

Houghton Mifflin Reading, while making use of readability formulas, clearly takes a conceptual and design approach to readability and addresses an interaction between students and reading text materials that goes well beyond the application of oversimplified, two-factor readability formulas.

References

Bruce, B., Rubin, A., & Starr, K. (1981). *Why Readability Formulas Fail,* Reading Education Report No. 28. Champaign, IL: Center for the Study of Reading.

Chall, J. S. (1984). "Readability and Prose Comprehension: Continuities and Discontinuities," in James Flood (Ed.) *Understanding Reading Comprehension.* Newark, DE: International Reading Association.

Chall, J. S., & Dale, E. (1995). *Readability Revisited: The New Dale-Chall Readability Formula.* Cambridge, MA: Bookline Books.

Cooper, J. D., Boschken, I., McWilliams, J. & Pistochini, L. (1998). *Stopping Reading Failure: Reading Intervention for Intermediate Grade Students.* Boston, MA: Houghton Mifflin Co.

Duffelmeyer, F. (1985). "Estimating Readability with a Computer," in *The Reading Teacher, 38* (4), 392-395.

Fountas, I. C. & Pinnell, G. S. (1996). *Guided Reading: Good First Teaching for All Children.* Portsmouth, NH: Heinemann.

Fountas, I. C. & Pinnell, G. S. (2001) *Guiding Readers and Writers: Grades 3–6.* Portsmouth, NH: Heinemann.

Gambrell, L. B. (1996). "What Research Reveals About Discussion," in L. B. Gambrell & J. F. Almasi (Eds.) *Lively Discussion! Fostering Engaged Reading.* Newark, DE: The International Reading Association.

Gray, W. S., & Leary, B. E. (1935). *What Makes a Book Readable.* Chicago, IL: University of Chicago Press.

Harris, T. L., & Hodges, R. E. (Eds.) (1995). *The Literacy Dictionary: The Vocabulary of Reading and Writing.* Newark, DE: International Reading Association.

"IRA, NCTE take stand on Readability Formulae," in *Reading Today, 2* (3), 1 (Dec. 1984/Jan. 1985).

National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.* Washington, D.C.: The National Institute of Health.

Peterson, B. (1991). "Selecting Books for Beginning Readers," in D. E. DeFord, C. A. Lyons, & G. S. Pinnell (Eds.) *Bridges to Literacy: Learning from Reading Recovery.* Portsmouth, NH: Heinemann, pp. 119-147.

John J. Pikulski

John J. Pikulski is Professor of Education at the University of Delaware, where he has served as Director of the Reading Center and as Chair of the Department of Educational Development. Active in the International Reading Association, he has served on its Board of Directors and was president of the Association from 1997–98. He has served on the editorial advisory boards of *The Journal of Reading Behavior*. He is a member of the Planning Committee of the National Assessment of Educational Progress (NAEP) in reading.

Dr. Pikulski is coauthor of *The Diagnosis, Correction and Prevention of Reading Disabilities* and *Informal Reading Inventories*, and coeditor of *Approaches to the Informal Evaluation of Reading* and *The Acquisition of Reading*.

He is a Senior Author of *Houghton Mifflin Reading*.

HOUGHTON MIFFLIN