Prologue

Once upon a time many years ago in the middle of Africa (Uganda, 1961) I found a need for readability. I was assigned by the university to improve the reading of a group of community college instructors who were in turn going to improve the reading skills of their students. The practical problem was that the students had difficulty reading manuals for autos and other machinery.

I knew from some years running a reading clinic in the United States that you can make more progress if you start the students out on relatively easy reading material so the students can read the materials with some comprehension and success. But how do you tell these African instructors how to select “relatively easy” reading materials in technical English? The answer - use a readability formula. There not being any easily accessible readability formulas lying around in Uganda - I decided to make one. It worked and with modification and much more supporting research it has served many thousands of teachers far remote from that little group of community college instructors.

Definition

Readability is defined as “the ease of comprehension because of style of writing” in *The Literacy Dictionary* (Harris and Hodges 1995).

This article is mostly about readability formulas which objectively assign difficulty levels to books and other written passages. Brenda Weaver in another IRA publication *Leveling Books K-6* (2000) defines leveling as “selecting books to match the competencies of a reader or writer”. Both are right.

Readability to most reading professionals usually means the application of a readability formulas. True readability does have a more general meaning found in popular dictionaries such as “easy or interesting to read - capable of being read” (The Random House Dictionary of the English Language, 1983). But in classrooms and publishing houses readability is often thought of as an objective numerical score obtained by applying a readability formula. Leveling on the other hand also yields a score of difficulty, but it is less objective and takes into account some subjective factors of judgment. Readability formulas frequently assign a grade level or some other numerical designation to a book.

However readability designations have been applied to almost every kind of prose including laws, newspaper articles, test passages, military manuals, and advertising.

It is not surprising that readability scores are most often found for textbooks and literature books used in the schools since most of the early research was done by university based educational psychologists and much of the later work by or for educational publishers.

To put it in broad prospective, readability is really part of the Scientific Movement in education which began in the 1920’s when schools started using standardized tests to measure students achievement and word counts to aid in
curriculum development. Before that and continuing to the present, many educators use more subjective judgment with statements such as, “I think this book would be about right for my 3rd graders”. Most readability formulas on the other hand are so objective that they can be done by computers - simply type in a passage or scan in a whole book and the computer will give you a readability formula score that it is at the 3rd grade level. However it is possible to use many readability formulas “by hand”, that is without a computer.

Most computer formulas are based on two inputs that have been verified by many research studies (see Klare 1988):

1. **A measure of syntactic difficulty** (grammatical complexity), usually measured by sentence length.
2. **A measure of semantic difficulty** (meaning or word meaning).

A common measure of this is word length measured in syllables or number of letters but sometimes semantic difficulty is judged by frequency, either an actual frequency count of the word or the fact that the word does or does not appear on a list of familiar words.

However these two basic factors have been supplemented, and occasionally substituted by a number other input factors such as book length and subjective judgment.

In fact any readability formula must be used along with subjective judgment because formulas do not take many important factors into account such as:

- **Motivation** - is the student really interested in the subject and/or are there other incentives.
- ** Appropriateness** - adult novels might not be appropriate for junior high though some of them have 8th grade readability scores.
- ** Readers background** - readers all have different cultural background and educative experiences. All are in some ethnic group, some social class, some geographical region, and have parents of different backgrounds in occupation and education.

**PURPOSE**

**What is the purpose of readability formulas?**

Certainly one of readability formulas most common purpose is to help students learn to read better. Teachers have long known that giving the student a book at the right level will both cause him or her (1) to really read it, (2) to comprehend it and (3) also to enjoy it. Give the student a book that has too high a readability level, and one or all of those three things is apt to be missing.

Aiding comprehensibility, or the transfer of information, is certainly one of the major purposes of readability. A readability formula attempts to predict the readers understanding of the written passage. This is certainly important in selecting text books. Much education is expected to take place by reading the text book in many subjects at every level from elementary school through college. Comprehensibility is also very important in many areas outside of school, it is no accident that General Motors, the Army, and the Navy have all done extensive work in readability as all are major publishing organizations.
whose output of manuals, correspondence, directives, and advertisements rival
many traditional publishing houses.

An interesting use of readability is in the law and government field. A
number of states have Plain Language Laws which state that various documents
such as loan contracts or insurance policies must be “readable”. At least 27
states required “plain language” in insurance policies sold in their state. In
practice this often means written at about the 8th grade level. President Carter
signed a directive to federal government agencies that they should produce
“readable” documents but there is some doubt if this well intentioned order has
had much effect on such agencies as the IRS or the Department of Justice.

Readability formulas have fared rather well when tested in court. I was an
expert witness in a class action suit brought against the Medicare system for
informing users of their appeal rights in a difficult to read letter (Readability Grade
Level 16). Judge Weinstein in the New York federal appellate court case found
readability formulas appropriate and ordered the U.S. Dept. of Health and

Some law professors have tried to get their students and the legal
profession to write more clearly and drop “legalese”. But considering how many
law suits pivot on the understanding of a written document, readability has barely
scratched the legal surface.

VALIDITY
How do we know if readability formulas are any good?

Readability is one of the most widely researched areas in education.
Klare (1988) stated that there are over 1000 published articles on readability. A
more recent look (2006) in the ERIC system under the entry term “readability”
yielded 2692 documents. The overwhelming majority say readability works in
fields as diverse as adult basic education and horticulture.

The proof of success is most frequently a correlation with a
comprehension test. The student understands less as the readability score
increases. The proof of student comprehension is often a multiple choice
comprehension test but sometimes it is a written or oral response.

A somewhat unique comprehension proof is the use of a cloze test in
which the student is asked to fill in blanks where words were omitted. One
advantage of a cloze score is that it is very objective, for example, every 5th word
can be omitted and only exact word replacement is counted. Cloze thus
eliminates the subjectiveness of writing or selecting comprehension test items.

Another proof of readability formulas is the correlation with oral reading
errors. As readability score increases so do oral reading errors. This
interestingly enough substantiates the three levels, Frustration, Instructional, and
Basic, suggested by Emmett Betts in 1946. He proposed that a student read a
passage and that one oral reading error in every 20 words would place that
reading material on the student’s Instructional Level, more errors would place
the material on the student’s Frustrational Level, and less errors on the
Independent Level. But teachers have long known that too difficult material
shouldn’t be used. Readability formulas simply refine and objectify this concept.
“Readership” is a concept borrowed from library science and journalism. Basically it is concerned with how many people are reading their materials.  

Studies have shown that by lowering the readability score, readership increases.  

Readability can also be shown by eye-voice span. Turn off the light in the middle of a student reading aloud and see how many words he continues to say. The more words he continues the easier the readability of the passage. In other words, the greater the distance between where the eye is fixating and where the voice is saying them, the easier the reading material.  

A similar objective proof is sub-vocalization. Sub vocalization is often thought of as the little voice you hear inside your head while reading silently. More objectively it can be measured by electrodes placed on the lips to monitor tiny muscle movements. The greater the subtle muscle movement the more difficult the passage.  

Subjective judgment has also been used to validate formulas. A group simple reads a number of passages then then ranks them. This ranking is greatly facilitated (made more reliable) if judgment is aided by comparisons with a set passages of known standard difficulty. This is the basis of the Singer (1975) readability procedure. Singer assigned Readability grade levels by having the teacher or editor compare graded paragraphs with the text sample. Subjective judgment is part of Reading Recovery book leveling procedure.  

Finally, most formulas agree fairly well on ranking a set of books. We could say they have concurrent validity or good correlation. However, there is less agreement on giving the level. For example, one formula might place a book at 5th grade and another place the same book at 6th grade. Grade level scores are usually within a year difference. Lately readability measures have used a variety of types of scores such as Level K or Lexile 370. See Table 1 for some comparisons of different types of scores.  

CRITICISM  
Are readability formulas criticized?  

Yes, they certainly are. Some of the major criticisms are partly philosophical, namely, there are still many educators who oppose the Scientific Method. To them classifying literature prose into a numerical score of difficulty is abhorrent. These same people frequently oppose standardized test scores. They prefer subjective judgment to objective scores. Flying under such banners as Whole Language, Literature Based, or earlier Progressive Education, they point out the flaws of readability formulas. Here are some of them:  

Formulas and tests are sometimes wrong. Correct, there is a Standard Error of Measurement in all prediction or assessment scores. There is no Standard Error of Measurement in subjective judgment because there is no way of quantifying it. To take it out of measurement terms, readability formulas are sometimes wrong, but subjective judgment is sometimes wronger. It is difficult to have a Standard Error if you don’t use objective measures or understand or don’t wish to use the applications of probability inherent in the normal distribution curve.
Readability formulas are also criticized for causing bad writing sometimes called “formula writing” or “dumbing down textbooks”. The criticism goes something like this “The reason many elementary textbooks and some children’s trade books are so bad is because the writer had to write them to a certain grade level”. While there is some truth in this, the formula didn’t cause the bad writing, the formula is meant to estimate (predict) readability after a piece is already written. It is not the formula’s fault if the publisher has a bad editor and a bad writer. For some comments on how to simplify and clarify writing see Fry’s article on Writeability (in Zakaluk & Samuels 1988 or ERIC ED 220 799) In brief readability formulas are not writers guides, there is much more to good writing that two simple inputs such as word and sentence length.

The criteria are correct in pointing out that readability formulas do not take into account all of the more cognitive factors that make reading difficult. The formula makers simply point out that the formulas are useful in doing what they are supposed to do, and that is in predicting comprehension and reading ease by a number of traditional and experimental measures.

Readability is not the same thing as “legibility”. Legibility has to do with type size, spacing and the quality of letter formation (in handwriting for example). Publishers sometimes take an old children’s story and set it in large type. This might increase sales to parents but it really does not change the true readability much and the readability score not at all.

HISTORY
How old is the readability field?

The first readability formula was published in 1923 by Lively and Pressey. However before that there have been numerous discussions of literary criticism, rhetoric, and writing style and they date back to Aristotle and continue to the present. William Holmes McGuffey is credited with writing the first series of graded readers in the mid 1800’s. The McGuffey Readers sold over 130 million copies between the 1850’s and the early 1900’s. Considering that the total population of the United States was 23 million in 1850 and only 76 million in 1900, that means a very high percentage of the school population used McGuffey’s graded readers. This is an interesting proof that graded reading textbooks had wide acceptance.

There are several rather detailed histories of readability (Chall 1958), (Klare 1963 and 1984), (Gilliland 1972), and (Harris & Jacobson 1979). Discussions of readability also occur in many major reading methods textbooks used in teacher training such as Harris and Sipay (1985), Ruddell(1999), Manzo and Manzo (1995), Vacca, Vacca, and Gove (1995).

The most widely used formula in schools and with educational publishers in the 1950’s 1960’s and 1970’s was the Dale Chall formula (1948). Businesses tended to use the Flesch formula (1948) or it’s slight modification known as the Flesch Kincaid,

The original Dale Chall formula graded books from grade 4 through 12 so for grades 1 to 4 in 1960’s, the Spache formula (1953) became popular for
grading books below 4th grade. Both used a list of familiar words, sentence length and required a numerical calculation.

In the 1980’s popularity shifted as Klare (1988) writes: “Another significant move toward the ease of usage is the Readability Graph developed by Fry (1963). The Graph permits a direct estimate of reading grade level on entering with syllable length and number of sentences per 100 word sample, thus providing another way of avoiding the manual use of a formula. It seems safe to say that in its most recent version (1977), Fry’s Graph is the most widely used of all readability methods.”

By the year 2000, readability has become less dependent on one or a few formulas and much more diverse. The readability grading of books is now largely done by publishers and large companies often using formulas which require their computers. For example, Advantage Learning Systems has a graded list of 25,000 books which contains most of the trade books used in schools. They use their own formula known as ATOS (Advantage TASA Open Standard) that uses the inputs of sentence length, average grade level of words and length of book. This formula uses a computer and the entire content of the book, not just samples of text. The formula yields grade level scores that are coordinated with their own STAR reading achievement test (Standardized Test for Achievement of Reading).

Another company that has analyzed 15,000 books is TASA (Touchstone Applied Science Associates) using their own computerized readability formula that yields DRP units (Degrees of Reading Power). Their inputs are sentence length, word length and proportion of common words. These readability levels coordinate with their own reading test which yields DRP units.

A third company that is doing computerized readability on large amounts of materials (26,000 trade books) is Meta Metrics. This program output is in units called Lexiles. Lexiles can be translated into grade levels. A student can take their achievement test that yields Lexiles, or any well known standardized test like the Stanford 9 Achievement Test.

Traditional readability formulas like the Dale Chall and Fry Graph are also available in computer format for individual use. One company providing these is Micro Power & Light Co. Lexiles, DRP and ATOS are not available for individual computers.

LEVELING
What is “leveling”?

Leveling refers to various systems of grading books for difficulty using a larger number of subjective and objective factors than most readability formulas. Leveling sometimes incorporates more traditional readability formulas or the inputs of traditional formulas. It is used much more at the primary levels than upper levels. The goal of leveling is often fluency, rather than strictly comprehension. Some of the factors taken into account by some of the leveling procedures are:

- Content - is it appropriated or familiar to that age group?
- Illustrations - do pictures tell the story or explain vocabulary?
Length- are there 2 words on a page?, how many pages in the book?
Levels- (scores) are not necessarily grade levels, often finer grading.
Curriculum- levels related to teaching methods and/or framework.
Language structure - includes repetitious words or phrases, flow.
Experience- levels can be adjusted using subjective judgment from
teaching experience.
Format - type size, spacing, page lay out.

The modern use of “leveling” is due in no small part to the work of the New Zealand department of education. It was partly popularized in the United States by Marie Clay (1991) and her Reading Recovery system which used early intervention of reading tutoring for children who had a high probability of failure. The Reading Recovery system found a need to find books with closely spaced difficulty level, particularly at the 1st and 2nd grade levels. Most traditional readability formulas are not particularly sensitive at those levels. Traditional wide range readability formulas such as the Dale Chall and Fry Graph only give whole grade designations at Grades 1 and 2. Large company book readability formulas such as Lexiles, DRP and ATOS do have finer unit designations but usually lack the more subjective text support factors.

Readability formulas aimed at the primary level, Grades 1 to 4, like the Spache (1953) and Gunning (1998) do give tenths of a year designations. However the Spache and Gunning like the other traditional formulas still use only the two traditional inputs of sentence length and vocabulary.

Several leveling systems have book lists of a number of leveled books. Fountas and Pinnell (1999) have over 7000 books, Weaver (2000) has over 2000 books and Gunning (1998) over 1000 books. Gunning’s leveling incorporates his own primary readability formula, and Weaver’s leveling incorporates the Fry Graph, Dale Chall formula, or DRP. Book leveling can be done by classroom teachers.

Book leveling is also a major part of Reading Recovery system of reading teaching (Clay 1991). She discusses “text support” which includes text features that are predictive, repetitive, and close to a student’s natural language.

FOREIGN LANGUAGES
Are readability formulas available for other languages?
Klare (1988) reported that there are readability formulas for 14 languages ranging from Afrikaans to Vietnamese. Most of the languages are alphabetic and use the two major inputs of syntax and semantics which in practice is sentence length and vocabulary. An interesting problem occurs in non alphabetic languages like Chinese which is written in ideographs (characters). The inputs for Chinese are vocabulary (proportion of words on a 5,600 word list) and brush strokes per character. Other languages have unique problems, for example, Spanish words has many longer polysyllabic words than English. Gilliam, Pena, & Mountain found that in using the Fry Graph for the first 3 grades you need to subtract 67 from the average syllable count.
A readability formula developed in Sweden and that is used in Europe is the "lix" which is short for "lasbarhetsindex" which translated means readability index, and is simply sentence length plus word length. A later modification is called "Rix" and it is number of long words divided by sentence length.

FORMULAS
What do readability formulas actually look like?

Since there are over 100 readability formulas it would be difficult to show them all but here are a few to give you a more concrete idea of what they look like.

The original readability formula developed by Livey and Pressey (1923) used 5 inputs and 6 numerical constants in their formula to yield an average comprehension score:

\[
x_1 = .01029x_2 + .009012x_5 - .02094x_6 - .03313x_7 - .01485x_8 + 3.77
\]

- \(x_1\) = average comprehension score
- \(x_2\) = number of hard words not on Dale List of 769
- \(x_5\) = number of personal pronouns
- \(x_6\) = average number of words per sentence
- \(x_7\) = percentage of different words
- \(x_8\) = number of prepositional phrases

Readability formulas have since been greatly simplified. The New Dale Chall Readability Formula (Chall 1995) has only 2 inputs to get grade level but is necessary to have the Manual which includes a 3000 word vocabulary of familiar words and tables to yield grade level or cloze scores.

The Fry Graph used widely in schools and by educational publishers requires 2 inputs, sentence length and word length in number of syllables. These are entered into a graph to yield grade level. See illustration.

To get Lexiles, DRP units or ATOS Grade level you pretty well need a large computer which in practice means that it is usually done by the company which developed those formulas.

The Flesch Kincaid Reading Ease Formula is used widely in industry and is fairly easy to calculate:

\[
\text{Grade Level} = .4(\text{words/sentence}) + 12 (\text{syllables/word}) - 16
\]

Leveling requires specific directions according to which leveling system you wish to use. As was mentioned earlier leveling often requires a number of subjective judgment factors such as content appropriateness, format, language structure, and illustration use.

CONCLUSION

The concept of readability, and more specifically readability formulas, has had an important influence on American education and the selection of school reading materials. To a lesser extent readability has influenced written
communication in the armed services, industry, government, and law. Some important practical uses of readability outside of textbooks are:

Newspapers - Rudolf Flesh was hired by the Associated Press to bring down the readability of front page news stories - he did (from Grade 16 to Grade 12).

Public Health - Schools of Nursing found the need for materials on illness prevention and correction so they include readability formulas in some of their textbooks.

Insurance - A number of state insurance Commissioners demand that policies issued in their state be “readable”. In practice that means about Grade 8.

Banks - Some states such as New York Plain Language laws state that consumer loan documents be “readable”. The banks resisted saying that the legal language was necessary because it had been tested in the courts. Yet when the banks rewrote the loan documents in plain language they had less law suits - maybe because the customers understood what they were signing.

Readability has also has some impact in languages other than English in countries ranging from Sweden to China.

The fundamental purpose of readability is to improve reading comprehension. This is particularly important in selecting text books and trade books for school use, but it is important to consider readability in any type of written communication.

Other important uses of readability are in selecting materials for successful reading instruction and for increasing the readership of library books and periodicals.

Leveling is a variation of readability more often used along with instruction methods at the primary levels. It uses a number of subjective factors and has the related goals of fluency and teachability.

Research has shown that most readability formula are based on the two factors of syntax (often sentence length) and semantics (word difficulty). Furthermore most formulas will rank a set of materials in the same difficulty order but there is less agreement on obtaining the same grade level for any one piece of writing. There is even less agreement on the way the readability score is reported. A grade level score is widely used, but difficulty is also reported in a variety of scores such as alphabet (A, B, C, etc.), Lexiles, and DRP units.

Readability is an active field under continuous development. The over 2500 references to readability in the ERIC system testify to it's widespread uses and interest.

References


David v. Heckler 1984, Appellate Court Case No. 79C2813, U.S. District Court, E.D. New York


Lively, B.A. & Sidney L. Pressey 1923 “A Method of Measuring Vocabulary Burden of Textbooks” Educational Administration and Supervision 9, 389-398


Ruddell, Robert B. 1999. Teaching Children to Read and Write. Boston: Allyn and Bacon
Table 1
Readability Formula Score Correlation Chart

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Reading*</th>
<th>DRP</th>
<th>Lexile Fountas*</th>
<th>Wright*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dale Chall, Recovery</td>
<td>Pinnell</td>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fry, ATOS</td>
<td>.5</td>
<td>3 - 5</td>
<td>27</td>
<td>70</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>10 - 11</td>
<td>35</td>
<td>170</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>16 - 17</td>
<td>42</td>
<td>370</td>
<td>K</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>46</td>
<td>508</td>
<td>O</td>
</tr>
</tbody>
</table>

ATOS = Advantage TASA Open Standard (Similar to Flesch Kincaid)
DRP = Degrees of Reading Power (TASA 1999)
Reading Recovery (Clay 1991)
Lexile (Meta Metrics 1995)
Fountas & Pinnell (1999)
Wright Group (publishers Bothel, WA)
*Indicates scores from “leveling”. The others are readability formulas.

(Illustration to be inserted - Fry Graph for Estimating Readability - attached hard copy only)