Readability versus leveling

Both of these procedures can help teachers select books for readers at different stages.

Teachers of reading have long known that it is best to use introductory reading material that is not too difficult. The procedure for determining difficulty has been objectified by the use of readability formulas since 1923 (Lively & Pressey, 1923) and continues to the present day. Readability formulas usually give a numerical score to rank books or other reading matter in an order of difficulty. Sometimes this numerical score corresponds to a suggested approximate grade level. Leveling is more modern (Clay, 1991) and is similar to readability in that it employs various systems for determining the difficulty of books or reading material, but it is more subjective and is used mostly at the beginning reading levels.

This article shows some similarities and differences between readability formulas and leveling procedures and reports some current large-scale uses of readability formulas.

Dictionary definition of readability and leveling

If you use some standard definitions of readability and leveling you still might need more help in finding similarities and differences. In The Literacy Dictionary (Harris & Hodges, 1995) readability is defined as “the ease of comprehension because of style of writing” (p. 203). Weaver (2000) defined leveling as “selecting books to match the competencies of a reader or writer” (p. 57). Both are right. Both are incomplete.

Readability to most reading professionals usually means the application of 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.

History and background of readability and leveling

The first readability formula was published in 1923, but leveling has a much longer history. When the United States was being founded, reading instruction commonly involved learning the alphabet and a few phrases and then jumping straight into the Bible. William Holmes McGuffey developed the first widely used “leveled” set of readers in 1836. The famous McGuffey Readers did not use grade-level designations but a simple numerical ranking (book 4 was more advanced than book 3).

The McGuffey Readers sold over 130 million copies between the 1840s and the early 1900s. Considering that the total population of the United States was 23 million in 1850 and only 76 million in 1900, a very high percentage of the school population used McGuffey’s leveled readers. This is interesting proof that leveled reading textbooks had widespread early acceptance.

In the 1920s there were a large number of different reading series, and certainly by the 1930s these books began to be firmly “graded” (e.g., a reader with a 4 on it meant it was for fourth grade).

In the later part of the 20th century, educators lost their love affair with graded reading
series because of the restricted content written just for the reading textbook that often used a controlled vocabulary. This led to a rise in “literature-based” readers in which the stories were more interesting but not graded.

The use of leveling in modern schools is due in no small part to the work of the New Zealand department of education. It was popularized in the United States partly 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 required teachers to find books with closely spaced difficulty levels, particularly at the first- and second-grade levels. Most traditional readability formulas are not particularly sensitive at those levels. Traditional wide-range readability formulas such as the Dale-Chall (Chall & Dale, 1995), and Fry Graph (Fry, 1977; see Figure on next page) give only whole-grade designations at primary grades. Large company book readability formulas such as Lexiles, Degrees of Reading Power (DRP), and ATOS 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), give designations in 10ths of a year. However, the Spache and Gunning, like the other traditional formulas, still use only the two traditional inputs of sentence length and vocabulary.

In 2001, the ERIC database had over 2,500 citations under the search term readability. Klare (1984) found over 1,000 research-based articles on readability. These facts indicate that readability is a widely researched area in education. With its long history dating back to McGuffey and its current metamorphosis as leveling, the grading of materials is apt to be around for some time to come. These citation counts do not include leveling as a search term, but they soon will because the two terms are closely related.

There are several rather detailed histories of readability (Chall, 1958; Gilliland, 1972; Harris & Jacobson, 1979; Klare, 1984). Discussions of readability also occur in many major reading methods textbooks used in teacher training such as Harris and Sipay (1985), Manzo and Manzo (1995), Ruddell (1999), and Vacca et al. (2003).

**What goes into determining readability and leveling scores**

Let us take a look at what determines the score of both procedures. Most traditional readability formulas are based on two measures that have been verified by many research studies:

1. **Syntactic difficulty** (grammatical complexity), usually measured by sentence length.
2. **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.

Most readability formulas 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. However, it is possible to use many readability formulas “by hand,” that is, without a computer.

There is a good degree of objectivity in the formulas. They provide fairly accurate comparisons of books so that a match can be made between the book and a student’s demonstrated abilities on an informal or formal reading assessment test. The objectivity also means that two different people or computers using the same formula will get the same score for the same book.

Leveling, on the other hand, is less objective. It cannot be done by a computer. Often leveling takes a number of “text support” factors into consideration such as the following:

- **Content**—Is it appropriate or familiar to that age group?
- **Illustrations**—Do pictures tell the story or explain vocabulary?
- **Length**—Are there two words on a page? How many pages in the book?
- **Curriculum**—How are levels related to teaching methods or framework?
- **Language structure**—Does language include repetitious words or phrases, flow?
Graph for estimating readability—extended
by Edward Fry

Directions: Randomly select 3 one hundred word passages from a book or an article. Plot average number of syllables and average number of sentences per 100 words on graph to determine the grade level of the material. Choose more passages per book if great variability is observed and conclude that the book has uneven readability. Few books will fall in gray area but when they do grade level scores are invalid.

Count proper nouns, numerals, and initialization as words. Count a syllable for each symbol. For example, “1945” is 1 word and 4 syllables and “IRA” is 1 word and 3 syllables.

Example:

<table>
<thead>
<tr>
<th></th>
<th>Syllables</th>
<th>Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>First hundred words</td>
<td>124</td>
<td>6.6</td>
</tr>
<tr>
<td>Second hundred words</td>
<td>141</td>
<td>5.5</td>
</tr>
<tr>
<td>Third hundred words</td>
<td>158</td>
<td>6.8</td>
</tr>
<tr>
<td>Average</td>
<td>141</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Readability seventh grade (see dot plotted on graph)
Judgment—Are the readers' background and experience appropriate to understand the text?

Format—How will the type size, spacing, and page layout affect readers' understanding?

These text support factors are absent from most readability formulas.

Modern leveling procedures used by Clay (1991), Fountas and Pinnell (1999), and Weaver (2000) include a number of judgment factors such as format, content, length, illustrations, repetition of words, and so forth. The addition of these factors has wide appeal to teachers, particularly those at the primary and remedial levels.

However, the inclusion of subjective or text support factors does not mean that these factors cannot be used along with traditional formulas. In fact, Weaver (2000) recommended the use of the Dale-Chall, Fry Graph, or DRP as part of her leveling procedure. Gunning (1998) and Chall and Dale (1995) recommended the use of some text support factors along with their readability formulas.

Large commercial companies have added a level of sophistication to readability by using computers. Lexile units by Metametrics (Zakaluk & Samuels, 1995), ATOS grade levels by Advantage Learning Systems (2000), and Degrees of Reading Power units by Touchstone Applied Science Associates (1999) are all traditional readability formulas based on sentence length and vocabulary difficulty. Because of computers they use much larger samples or even the whole content of books, and their units have finer gradations than whole-grade levels. Their scores, including Lexiles and DRP units, can be translated into grade-level scores like the other readability formulas. The companies also produce their own reading achievement tests that correlate with their readability units. However, more generally used reading achievement tests can be correlated with any readability units.

Readability formula makers have long known that formulas have limitations and do not include such important factors as motivation or appropriateness. What high school teacher has not seen a student with reading test scores near fourth-grade level sit for hours and read a drivers' license manual that has a readability of ninth grade? This same student unfortunately has a great disinclination to read the social studies book, which also has a 9th-grade readability level. Readability formulas do not take into account factors inside the reader's head and tend to be "bottom up" or text based in theoretical terms. Leveling attempts to compensate for part of this limitation.

Do readability and leveling cover the same range of difficulty?

There is a decided difference in the range or levels of difficulty between readability and leveling. Readability has a much wider range; most readability formulas have a grade-level range from 1 through 12 or 1 through 17. Leveling tends to grade books between kindergarten and grade 6 with some going only to third or fourth grade (see Table). Leveling is not used outside of the elementary classroom, while readability formulas are used both inside and outside the classroom. Readability outside the classroom is used for a wide range of material including such things as military training manuals, plain-language laws for insurance policies or loan contracts, and newspaper articles. Leveling is used most

<table>
<thead>
<tr>
<th>Comparing readability formula and leveling ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
</tr>
<tr>
<td>New Dale-Chall (F)</td>
</tr>
<tr>
<td>Lexile (F)</td>
</tr>
<tr>
<td>Grade 1</td>
</tr>
<tr>
<td>Gunning (F &amp; L)</td>
</tr>
<tr>
<td>ATOS (F)</td>
</tr>
<tr>
<td>Degrees of Reading Power (F)</td>
</tr>
<tr>
<td>Grade primer</td>
</tr>
<tr>
<td>Fry Graph (F)</td>
</tr>
<tr>
<td>Fountas &amp; Pinnell (L)</td>
</tr>
<tr>
<td>Grade 1</td>
</tr>
<tr>
<td>Reading Recovery (L)</td>
</tr>
<tr>
<td>Grade 1</td>
</tr>
<tr>
<td>Cloze (C)</td>
</tr>
<tr>
<td>Grade 1</td>
</tr>
<tr>
<td>Weaver (L)</td>
</tr>
</tbody>
</table>

Note. F indicates readability formula; L indicates leveling procedure; C indicates cloze, a procedure using reader responses.

Readability versus leveling 289
that it is really part of what became known as the scientific movement in education. This era began in the 1920s when standardized tests entered schools. These included not only achievement tests like nationally standardized reading tests, but also achievement predictors like IQ tests (Venezky, 1984). Curriculum, too, was affected by the scientific movement; a number of word-frequency counts were done on reading materials and children’s writing (Thornodyke, 1921). These word-frequency counts heavily influenced reading series in the 1930s and in several decades following. In fact some reading series actually tried to tabulate the number of times a word appeared in successive stories. This tight word control was a serious attempt to level reading textbooks. The problem was that it led to some rather contrived stories, and the reaction of many educators was to move toward a language experience approach in which children and teachers wrote their own stories to be read, and toward the use of a wide variety of children’s literature books as in the whole-language movement.

But the problem in using children’s literature books is that they often were not graded. As readability researchers have shown, when you give a beginning reader a book at too difficult a level it makes instruction more difficult. Not only is reading instruction more difficult but also there is a tendency for the child not to read a book voluntarily for pleasure when it is too difficult. It is no accident that many best-selling adult novels are written at the eighth-grade level. It is also no accident that the Associated Press hired readability formula maker Rudolf Flesch as a consultant to help lower the readability of news articles. Book editors and newspaper editors want their output read widely. Nobody is saying that adults or children should not read difficult material, but rather that if you want to increase readership and comprehension on a large scale, or with a struggling reader, then you had better pay some attention to readability.

The need for finding reading materials at the right level for instruction makes common sense to most teachers. It has been made more specific by classic reading educators like Emmett Betts (1946), who told teachers how to separate materials into independent, instructional, and frustration reading levels through the analysis of children’s oral reading errors. On an international and more theoretical level, Vygotsky (1978) extensively at the primary levels in conjunction with teaching reading.

Discussion: Finding the right level
Looking back at the beginnings of readability formulas (Lively & Pressey, 1923) we can see...
developed the concept of the Zone of Proximal Development (ZPD), which urged educators not to use materials beyond learners’ “language ceiling.”

Readability has the strength of a large research base with many formal validity studies showing high correlations with reading comprehension, oral reading errors, readership (number of readers and amount of reading), and even physical observations such as eye movements and subvocalization (Fry, 1977). Both readability and leveling have the strength of fostering success in classroom and remedial teaching.

Readability has the strength of objectivity and consistency that any person or computer will get the same score. Leveling has the strength of taking more factors into account than traditional readability formulas and usually provides finer gradations at the primary levels.

Large companies with computerized readability formulas can do massive amounts of work. For example, the ATOS (Advantage Learning Systems, 2000) system has graded 25,000 books using every word in each book. The DRP (Touchstone Applied Science Associates, 1999) system has a list of 15,000 books, and the Lexile (Zakaluk & Samuels, 1995) system has a list of 26,000 graded trade books. Leveling deals in smaller but still considerable numbers; Fountas and Pinnell (1999) have 7,500 leveled books, Weaver (2000) over 2,000, and Gunning (1998) has over 1,000. However, classroom teachers can also grade their own books using a traditional readability formula or leveling procedure. Different formulas have fairly good agreement in ranking a set of books but less agreement on assigning grade levels.

Both those in the field of readability and those in its newer (or older) sister of leveling are trying to do one important thing—help classroom teachers to teach reading better by helping them select books for their students. Selecting books by a readability formula or leveling procedure is only one of the many teaching techniques that a reading teacher should use, but it is one that will help many students have a successful learning experience.

References