Fresh ideas on readability and a "quick and easy" formula are offered for readers' reactions.

SMOG Grading — a New Readability Formula

Go To SMOG Calculator

Go To Home Page

G. HARRY MC LAUGHLIN

IN A RECENT ISSUE of this journal, Fry *(4)* described "a readability formula that saves time." The panel immediately below gives a method of assessing readability which is even quicker—and so simple that one statistician who glanced at an earlier version of this paper thought that it was a "put on," a fake. Although this system, SMOG Grading, is laughably simple, it is in fact more valid than previous readability formulas. The rest of this paper is devoted to substantiating that claim.

SMOG Grading

1. *Count 10 consecutive sentences near the beginning of the text to be assessed, 10 in the middle and 10 near the end.* Count as a sentence any string of words ending with a period, question mark or exclamation point.
2. *In the 30 selected sentences count every word of three or more syllables.* Any string of letters or numerals beginning and ending with a space or punctuation mark should be counted if you can distinguish at least three syllables when you read it aloud in context. If a polysyllabic word is repeated, count each repetition.
3. *Estimate the square root of the number of polysyllabic words counted.* This is done by taking the square root of the nearest perfect square. For example, if the count is 95, the nearest perfect square is 100, which yields a square root of 10. If the count lies roughly between two perfect squares, choose the lower number. For instance, if the count is 110, take the square root of 100 rather than that of 121.
4. *Add 3 to the approximate square root. This gives the SMOG Grade, which is the reading grade that a person must have reached if he is to understand fully the text assessed.*

A readability formula is simply a mathematical equation derived by regression analysis. This procedure finds the equation which best expresses the relationship between two variables, which in this case are a measure of the difficulty experienced by people reading a given text, and a measure of the linguistic characteristics of that text. This formula can then be used to predict reading difficulty from the linguistic characteristics of other texts.

The linguistic measures which have been found to have greatest predictive power are word and sentence length. I have shown *(9)* that these measures are, respectively, indicators of semantic and syntactic sources of reading difficulty. In English, word length is associated with precise vocabulary, so a reader must usually make extra effort in order to identify the full meaning of a long word, simply because it is precise. Long sentences nearly always have complex grammatical structure, which is a strain on the reader's immediate memory because he has to retain several parts of each sentence before he can combine them into a meaningful whole.

Regression analysis can find the best formula only if the investigator happens to have chosen the best general form for the equation. What previous investigators have generally overlooked is the fact that semantic and syntactic difficulty interact. A slight difference in word or sentence length between two passages does not indicate the same degree of difference in difficulty for hard passages, as *it* does for easy passages. Therefore, a readability formula should not be of the usual form,

Readability = $a + b$ (Word Length)+ c (Sentence Length),
but should be of the form

Readability = $a + b$ (Word Length x Sentence Length) where a, b and c are constants.

By a stroke of good fortune, the more valid type of formula is easier to calculate, not merely because it has one constant less than the traditional type, but because, with a bit of ingenuity, one can eliminate the chore of multiplication completely! Obviously, you must measure word length and sentence length separately if you are going to add the two measures together. But you achieve the equivalent of multiplying the two measures if you simply count out a fixed arbitrary number of sentences and then count,

DR. G. HARRY MC LAUGHLIN *is an Associate Professor of Communications in the School of Journalism, Syracuse University. His Ph.D. in psycholinguistics was obtained at University College, London. He taught for six years at The City University, London, and one year at York University, Toronto, before going to Syracuse.*

say, the number of syllables within those sentences. For any given average number of syllables per word, the count will increase if the sentence length is increased; likewise, for any given average number of words per sentence, the count will be greater if the word length is increased.

Fortunately there is no need to follow Flesch's (3) system of counting every syllable in a passage in order to obtain a valid measure of its semantic difficulty. I have found a law relating the number of syllables in a passage to the percentage of polysyllabic words, defined as words of three or more syllables. For practical purposes, the total number of syllables per 100 words may be calculated by this rule of thumb: multiply the number of polysyllabic word by 3 and add 112.

It was Gunning *(5)* who first had the idea of counting polysyllabic words to obtain a measure of semantic difficulty. I call my system of readability prediction SMOG Grading in tribute to Gunning's Fog Index (The term also refers to my birthplace, smog having first appeared in London, though, like so many other things, it has since been improved upon in several American cities).

So far we have seen how to eliminate the multiplication of the word and sentence length measures. What about the constant multiplier *b?* We can virtually get rid of that, too, by making it equal to unity through the simple device of picking a suitable arbitrary number of sentences to be counted!

By a process of trial and error I have found that 30 sentences is a suitable number for the criterion of readability used here. Other readability prediction systems invite one to use samples of only 100 words. Such a sample is so small that it may be quite uncharacteristic of the text being assessed. To get a more reliable prediction, you have to calculate the reading difficulty of several samples and then average them. By sampling 30 sentences, which typically cover 600 words, you get a reliable prediction straight away, particularly if the 30 sentences are divided into three groups of ten consecutive sentences, each group being in a different part of the text.

We have now seen that the relative reading difficulty of a passage may be assessed simply by counting the polysyllabic words in 30 sentences. The polysyllable count thus obtained must next be converted into some more meaningful number.

For this purpose I used the 390 passages included in the 1961 edition of the *McCall-Crabbs Standard Test Lessons in Reading (8).* A number of questions are given for each passage. Following each set of questions there is a table which shows the average reading grade of subjects who could answer correctly none, some, or all of these questions. I took as the indicator of the reading difficulty of each lesson the grade of subjects showing complete comprehension, because this is a more meaningful standard than that used by previous investigators—namely the ability to answer 50 or 75 percent of the questions. Complete comprehension means precisely that, whereas the ability to answer a certain proportion of questions will depend much more on the nature of the questions.

To find out just how standardized the *Standard Test Lessons are,* I obtained the following information from Dr. McCall (7). The mean grade score given for the subjects correctly answering a certain number of questions on each Lesson was obtained by averaging their reading grades as ascertained from the Thorndike-McCall Reading Test, which was administered at the same time as the questions on the Lessons published in 1925. For Lessons added to later editions the Gates Reading Test was used to determine the reading grades. Actually the published grades were not obtained directly, but were derived from smoothed curves. About ten lessons were given to each child, and all told, there were "thousands" of subjects.

The degree of approximation inherent in this method of standardization led Dale and Chall *(2)* to observe that a McCall-Crabbs criterion has "serious deficiencies." Though I have proposed a procedure for obtaining a more valid criterion (10), the *Test Lessons* still provide the best criterion we have.

I therefore set out to find a regression equation relating the polysyllable count of each Lesson to the mean grade score of students who could correctly answer all questions on that Lesson. For this purpose the TSAR statistical package was used on an IBM 360/50 computer. I found that 30 sentence samples gave vastly more accurate predictions than smaller samples. The samples were obtained by arranging the Lessons in order of difficulty according to grade score, which was then averaged over each set of three consecutive Lessons. The polysyllable count was obtained from ten sentences in each Lesson. If a Lesson contained fewer than ten sentences, the average number of polysyllabic words per sentences was determined and multiplied by ten.

By trial and error, four powerful regression equations were derived. They are set out in Table 1. Equation (a) yields predictions which correlate 0.71 with the criterion—equal to the highest correlation ever obtained before using the McCall-Crabbs criterion *(6,* p. 114). Unfortunately, the equation cannot predict readability below sixth grade, and it involves a multiplication which is difficult to do in one's head. I therefore tried taking the square root of the polysyllable count and obtained equations (b) and (c). The trouble with (b) is that there is again an inconvenient multiplier; in (c) two additions are involved, but the multiplier, being almost unity, is virtually eliminated. I therefore set up equation (d) which is a compromise between (b) and (c). Because (d)

TABLE 1

Regression equations relating p (number of polysyllabic words in samples of 30 sentences) to g (reading grade of subjects able to answer correctly all questions on material sampled).

Regression Equation	*Correlation*	*Standard Error*
(a) $g = 6.2380 + 0.0785\, p$	0.713	1.4661
(b) $g = 4.1952 + 0.8475 \sqrt{p}$	0.709	1.4751
(c) $g = 2.8795 + 0.9986 \sqrt{p + 5}$	0.729	1.4446
(d) $g = 1.0430\,(3 + \sqrt{p})$	0.985	1.5159
$= 3.1291 + 1.0430 \sqrt{p}$		

involved forcing the regression line through the origin, the correlation coefficient is spuriously high. The real indicator of the validity of (d) is that its standard error of the estimate is only trivially higher than that of the other formulas.

For practical purposes equation (d) can be approximated by *SMOG grade = 3 + square root of polysyllable count.* It is this simple formula which is expressed in the instructions given in the panel at the start of this paper.

The standard error of the predictions given by the simplified SMOG Grade formula is only about 1.5 grades. In other words, this formula will predict the grade of a passage correctly within one and a half grades in 68 percent of cases. This is apparently less accurate than in the predictions given by other regression formulas *(6,* p. 114). But it must be noted that other formulas have had to be supplemented by arbitrary "corrections." For example, Dale and Chall found a regression equation which predicts that even the hardest passage is suitable for a student in grade ten; they therefore had to introduce a table which shows that a grade yielded by the formula should be converted into a prediction with a range of two or three grades. Thus a readability grade of 10 is converted to 16 +. Other formulas rely upon equally arbitrary conversion tables or charts.

Furthermore, it takes only about nine minutes to derive a SMOG Grade based on a sample of about 600 words, whereas it takes the same time to derive the Dale-Chall prediction using a sample of only 100 words or a Flesch Reading Ease score based on but two 100—word samples.

SMOG Grading implicitly makes two claims; that counting polysyllabic words in a fixed number of sentences gives an accurate index of the relative difficulty of various texts; and that the formula for converting polysyllable counts into grades gives acceptable results. Both claims had to be tested.

TABLE 2

Comparison of 64 subjects' mean reading efficiency scores with polysyllable counts of 8 passages.

Title of Article	*Compre-hension /Time*	*Poly-syllable Count*	*SMOG Grade*
"Music Education: Aesthetic or Anesthetic?" *Music Educator Journal,* Oct., 1968, pp. 52-53.	1.03	244	18
"What Organ Transplants Mean to Your Life," *Vogue,* October 15, 1968, pp. 100, 101, 147-151.	1.10	169	16
"Impact: A Program and a Process for Potential Dropouts," *Child Welfare,* April, 1968, pp. 229-232.	1.12	150	15
"Regeneration of Organs," *Science Journal,* September, 1968, pp. 69-74.	1.18	147	15
"Superego Manifestations in the Treatment of a Four-year-old Boy," *The Psychoanalytic Review,* Summer, 1968, pp. 278-288.	1.20	145	15
"How to Choose a Music Teacher," *McCalls,* October, 1968, pp. 64-68.	1.33	112	13
"Can You Be Sure Your Child Won't Be a School Dropout?" *Children's House,* Spring, 1968, pp. 10-14.	1.37	72	11
"Say No to Your Children," *Reader's Digest,* October, 1968, pp. 145-147.	1.51	39	9

The predictive power of polysyllable counts was revealed when 64 university students were each asked to read eight 1,000-word passages from various periodicals. To counteract order effects, the articles were presented in a predetermined random order which differed for each subject. Before the experiment, three specialists in literacy training made content analyses and identified the ten most important ideas in each passage. After reading a passage each reader was asked to recall its entire content as fully as possible. This method of unaided recall was used in order to avoid the prompts which are inevitably given to a subject when he is questioned directly. The recalls were tape recorded and later transcribed verbatim. Each transcript was then compared with the list of ten main ideas, so that every passage was rated for comprehension by each reader on a scale of 0 to 10.

Speed of reading was not controlled, so there was a tendency for a subject who read a passage with slow care to recall more points from it than from a passage he read rapidly. Therefore the average comprehension score on each passage was divided by the average time in minutes which subjects took to read it. This gives a measure of reading efficiency of the kind called for by Braam (1). The measure was always very close to unity because the average number of minutes of reading time roughly matched the average number of points recalled.

As can be seen from Table 2, there is a perfect negative rank correlation between polysyllable counts and the measures of reading efficiency.

The method of counting thus seems to be vindicated. As for the grade levels, Table 2 shows that they tend to be rather high. This is not unreasonable, considering that the grades are supposed to be those which a reader needs to ensure complete comprehension. (It is to be understood that SMOG Grades 13-16 indicate the need for college education; 17-18 the need for graduate training; and 19 and above, the need for a higher professional qualification.) Comparisons show that SMOG Grades are generally two grades higher than the corrected Dale-Chall levels, which purport to indicate "the grade at which a book or article can be read with understanding"—a less severe criterion than the one used here.

Acknowledgments

Several graduate students carried out validation studies; I should particularly like to thank Misses Nora Swindler, Cordelia Gobuyan, and Diana Likely, the Rev. Theodore Leidenfrost, and Mr. Willard Scoffield. I am also most grateful to Mr. Michael McGill for running the numerous computer programs.

REFERENCES

1. Braam, Leonard S. "Developing and Measuring Flexibility in Reading," *Reading Teacher, 16,* 1963, 247-251.

2. Dale, Edgar and Jeanne S. Chall. "A Formula for Predicting Readability," *Educational Research Bulletin, 27,* 1948, 11-20, 37-54.

3. Flesch, Rudolf F. "A New Readability Yardstick," *Journal of Applied Psychology, 32,* 1948, 221-233.

4. Fry, Edward. "A Readability Formula That Saves Time," *Journal of Reading, 11,* 1968, 513-516, 575-578.

5. Gunning, Robert. *The Technique of Clear Writing.* McGraw-Hill, 1952.

6. Klare, George R. *The Measurement of Readability.* Iowa State University, 1963.

7. McCall, William A. Telephone conversation, February 28, 1969.

8. McCall, William A. and Lelah M. Crabbs. *Standard Test Lessons in Reading.* New York: Columbia University Teachers College, rev. ed. 1961.

9. McLaughlin, G. Harry. "What Makes Prose Understandable." Ph.D. Thesis. University College, London, 1966.

10. McLaughlin, G. Harry. "Proposals for British Readability Measures," in John Downing and Amy L. Brown (eds.) *The Third International Reading Symposium.* London: Cassell, 1968.

☆ ☆ ☆ ☆ ☆