McLAUGHLIN, G. HARRY (1968)
"Proposals for British readability measures" in *The Third International Reading Symposium* edited by John Downing and Amy L. Brown. London: Cassell, 1968. pp. 186-205

Proposals for British Readability Measures

by G. Harry McLaughlin
York University, Toronto

Knowing that one has to present a paper confers the same benefit as the knowledge that one is about to be hanged—it clears the mind wonderfully. Thus until I was invited to participate in this Symposium I had only a fuzzy intention of one day extending my researches from adults' reading to children's. But now this ambition has grown into a megalomaniac scheme for enlisting the help of schoolteachers in producing what I believe would be the first British children's readability predictor.'

However, before revealing this grandiose design, I must explain what is meant by readability, how it can be measured, and how it can be predicted.

WHAT READABILITY MEANS

One of the least ambiguous published definitions of readability is that given by English and English in their *Dictionary of Psychological Terms.* 'Readability', they say, 'is the quality of a written or printed communication that makes it easy for any given class of persons to understand its meaning, or that induces them to continue reading.'

Here, right at the start, we find revealed that impreciseness which bedevils so many studies of readability. There is no *single* quality which makes a text compelling—that is to say, which induces a reader to continue reading it. Even if we amend the

[1] When I prepared this paper I was Lecturer in Communication at The City University, London. Having now retreated to Canada, I wish the best of British luck to anyone brave enough to carry out this scheme.

definition to say that readability is that combination of qualities which makes a text comprehensible or compelling, there remains the objection that we are in effect proposing a choice of three definitions: readability is what makes a text comprehensible; *or* readability is what makes a text compelling; *or* readability is what makes a text both comprehensible and compelling. In fact, many researchers have carried on as if all these three definitions were interchangeable. In any case, every one of these definitions is scientifically unsatisfactory, as becomes clear if one asks: How do I measure the qualities that make a text compelling (or whatever)? For example, how does one measure the quality of containing an optimal number of ideas? This quality is obviously needed for compellingness, because if a text contains nothing new to a reader he will not be very keen to read it. On the other hand, he will soon give up if it contains too many new ideas crammed too close together. But it is impractical to try to measure this quality, not only because it is difficult to measure the exact number of ideas there are in a given text, but also because one cannot decide whether this number is optimal or not until one has ascertained the extent to which its readers know about the subject matter of the text, the rate at which they can absorb new information, and so on.

As for measuring the qualities that make for comprehensibility, how can one determine the degree of ease with which a subject can understand any given text? We may judge that he has understood it in so far as he can summarize it in his own words and answer questions about it, but measures of these performances on different texts are not comparable. We can achieve comparability by Taylor's (1953) doze procedure, in which the subject has to replace words missing from mutilated versions of various texts; but success in this task does not necessarily indicate overall comprehension.

It can now be seen that it is not very useful to define the readability of a text in terms of its qualities, because we cannot satisfactorily measure some of the most crucial qualities.

On the other hand, it is entirely feasible to ascertain what proportion of a clearly specified class of people find a clearly

specified kind of reading material compelling to the extent that they read it by choice. Furthermore, one can presume that people generally choose to read only material that they can understand. It therefore seems preferable to define readability as the degree to which a given class of people find certain reading matter compelling and, necessarily, comprehensible. If the kinds of texts that people read from choice are classified by measures of style—such as the ratio of adjectives to active verbs—these measures can then serve, singly or in combination, as predictors of readability.

USES

There are two important uses for a readability predictor. One is that it enables anybody, without special knowledge or training, to determine fairly precisely what proportion of a given population would find the style of a given text acceptable; even an experienced teacher or librarian might usefully employ an objective predictor as an adjunct to her intuition in selecting books for children. The other use for a readability predictor is as a warning device: a writer hoping to reach a particular public can check whether his writing does in fact suit his intended readers—civil servants please note! However, this device must not be used as a guide to style; there is more to clear writing than can be measured by a practicable readability predictor.

VARIABLES AFFECTING READABILITY

Let us consider some of the many variables—typographical, motivational, and logical—which affect readability, although they are not generally used as predictors.

Visibility

Typographical determinants of readability include visibility and legibility. Reading is obviously impossible unless the letters are visible; one must be able to detect and identify them, and this performance may be affected by seeming trivialities. For example, visibility is usually reduced by glossy paper because

this reflects so much light that it produces glare which masks the printing. Visibility is also reduced by tinted paper; the contrast in brightness between paper and ink should be as high as possible for maximum visibility. Another requirement is that the letters should be fully distinctive. In traditional type-faces, the top halves of letters are much more differentiated than the bottom halves. There is greater distinctiveness in the Initial Teaching Alphabet (i.t.a.). There the type designer has dared to swing below the base-line in characters such as  ,  and  . It would be interesting to find out whether such typographical innovations would increase legibility if they were used with the traditional orthography.

However, the most important variable for visibility is the size of the type. Tinker (1965), especially, has done a lot of research on this variable and on suitable amounts of leading, that is to say the spacing between the lines. The findings may be summarized in a little rule-of-thumb which I have concocted: divide a child's age by two, and that gives the number of printed lines per inch with which his perceptual system can cope satisfactorily.

Legibility

A grapheme may be roughly defined as a letter or combination of letters representing a single sound. The efficiency with which graphemes and combinations of graphemes in a text can be recognized at speed is termed the text's legibility—for which visibility is, of course, a prerequisite. Burt (1959) alleges that certain founts of type are easier than others for children to read; but generally, so far as traditional type-faces are concerned, it seems that the most legible are simply those to which one is most accustomed. One factor determining legibility is the degree to which the letters forming a word cohere perceptually. Subtle adjustments for optical illusions, proportional spacing, morticing and serifs are all intended to assist the linking of letters into words. The Initial Teaching Alphabet carries this process a step further by joining together in ligatures the letters which form a single grapheme.

Legibility also depends upon the perceptual system being

given some help in sweeping back from the end of one line of type to the start of the next. Very wide lines may make this difficult but, in books, any setting between three and five inches wide is acceptable at any age, though it helps if the line length fits the rhythm of the text. What matters more is that there should be adequate leading between the lines.

Individual differences

Reading efficiency, as measured by tests of speed and comprehension, is also affected more or less adversely by glare or inadequate illumination, incorrect positioning of the reading matter, discomfort, noise and other distractions, faulty eye-movement habits, poor binocular fusion, optical defects, ill-health and fatigue.

But wait! There are more determinants of readability to reckon with. Consider such motivating variables as interestingness and aesthetic appeal. For instance, margins have been found to have no effect on legibility. It must be aesthetic preference which dictates that 4o per cent of a normal page shall consist of blank margin. And whether a child is motivated to select a certain book to read may depend more often than not upon the quality of its illustrations. Above all, the motivation of a reader to go on reading is determined by the interestingness of the text. But what interests a particular reader at a particular moment is decided by what he knows already, what he needs to know, how much time he can give to reading, how much he likes reading in general, and other factors which depend upon numerous events in his life-history.

Logical determinants

Furthermore, there are what may be called logical determinants of readability, in particular, orderliness of presentation and optimal idea-density. In principle, it would be possible to measure the logical organization of a text by translating the statements made in it into some system of symbolic logic, then counting the superfluous propositions and those which must be supplied to make valid any conclusions given in the text, and, finally,

combining these counts with a measure of the orderliness in which the propositions appear in the text. Idea-density could also be measured as the ratio of the number of propositions to the number of words in a text; however, as we have seen, such a measure would hardly be meaningful unless one knew the extent of the reader's background knowledge, so as to decide what proportion of propositions were redundant to him.

Thus we see that what makes a text comprehensible and compelling depends upon a whole host of interacting typographical, motivational and logical factors, most of which it would be impractical to measure. So, in practice, all these factors are ignored in readability prediction!

LINGUISTIC PREDICTORS

What basis for prediction, you may well ask, can possibly remain? The answer is that there are linguistic variables which make a text more or less readable. We will examine these linguistic variables a little later. Let me just say for the moment that linguistic difficulty can be measured by word length and sentence length. These crude indices of vocabulary and style suffice to predict with surprising accuracy the age, educational attainment and other characteristics needed for a child to understand a given book.

Usually the measures are combined in an equation called a readability formula. Here is a formula devised by Flesch (1948), as recalculated by Powers, Sumner and Kearl (1958):

$$4{\cdot}55\ w.l. + 0.0778\ s.l\ - 2.2029 = r.g.$$

where *w.l.* (word length) is the average length in syllables of the words in a passage, *s.l.* (sentence length) is the average length in words of its sentences, and *r.g.* (reading grade) is the average American school reading grade of a child who can correctly answer half the comprehension questions on the passage. The formula has a standard error of only 0.85—which means that, if the formula predicts that children need to reach a certain reading grade before they can answer correctly half the comprehension

questions on each of a number of passages, then 68 times out of 100 the predicted grade will be correct within 0.85 of a grade, and 95 times out of 100 it will be correct within 1.7 grades.

Such predictive power is high enough to be useful. Do not, however, let me mislead you into thinking that readability prediction is without severe limitations. Remember the multitude of factors which affect readability. Any prediction from linguistic variables must assume that the reader is interested in the subject-matter of a text, and knows enough to be able to understand this subject-matter, and that it is coherently organized, as well as being sufficiently legible and attractively produced.

MEASURES OF READABILITY

I have proposed that readability should be measured in terms of compellingness (and comprehensibility), but this is not the only possible criterion. Some investigators have used reading speed as a measure of readability; and there is a weak positive correlation between comprehension and rate of reading. However, poorly motivated readers tend to go faster, but their eye-movements are very regular, which generally indicates reading without comprehension. Thus a high speed of reading can indicate either good comprehension, or hardly any! Which disposes of that criterion.

Judgements of reading ease and interest, made either by acknowledged experts or typical readers, have also been used to validate readability predictors. In view of the general discrepancy between subjective guesses about likely behaviour and actual scores, such a criterion is best ignored: of course, if people could make accurate judgements of comprehensibility and so forth, there would be no need for readability predictors anyway!

A few studies have assessed readability by measuring the amount of text that a reader examines before he skips on to something else, but this criterion is only suitable for measuring the interestingness of magazine articles. The amount of a text that a reader can recall, either immediately or after some time, has been used to assess readability; unfortunately it is difficult to measure amounts of text consistently and meaningfully recalled.

In fact all the readability formulas in general use have been calculated, directly or indirectly, from the comprehension scores of American children reported by McCall and Crabbs in their *Standard Test Lessons in Reading.* More specifically, in the 1950 revision of the test lessons there are 383 short prose passages, for each of which there is given the average reading grade of pupils who can correctly answer just 50, 75, and 90 per cent, etc., of the comprehension questions on the passage. Quite apart from the difficulty of making comprehension questions comparable, to my mind this is not adequate as a measure of readability, for what is comprehensible is not necessarily compelling.

Criterion for adult reading

On the other hand, what people repeatedly choose to read must be comprehensible. I am taking advantage of this fact to measure the readability of adult periodicals. Every day millions of Britons participate in a self-selection test by buying newspapers and magazines; and throughout the year 16,000 people, carefully selected to give a cross-section of the entire adult population, are interviewed on behalf of advertising agencies to ascertain what periodicals they read. Among other details, they are asked the age at which their formal education ceased. Of course, this Terminal Educational Age—T.E.A. for short—is not the only factor which determines whether someone will read one periodical rather than another. But nobody will regularly read a periodical which is too silly or too serious for his taste, and surveys show that his taste will largely correspond to the extent of his schooling. Thus I can measure the readability of a periodical in terms of the proportion of its readers who fall within each of certain terminal educational age ranges. For example, 43 per cent of readers of *The Times* have T.E.A.s of 19 years or more, and 22 per cent have T.E.A.s of 15 or less, whereas only 1 per cent of the *Daily Mirror's* readers have T.E.A.s of 19 or more, and 95 per cent have T.E.A.s of 15 or less (Abrams, 1963).

Now let us recapitulate the main points made so far: first, readability is best defined as the degree to which a text is compre-

hensible and compelling to a given class of people; second, the readability of a text can best be measured in terms of the distribution by educational level, or some similar characteristic, of the people who find that text most compelling; third, despite the influence of numerous other variables, the readability of a text can be predicted efficiently for children from measures of word and sentence lengths.

I want to spend the rest of my time trying to answer three questions : why are word and sentence lengths so effective in predicting readability? what is wrong with present readability formulas? and, how could we produce an effective British children's readability predictor?

WORD AND SENTENCE LENGTHS

Researchers have counted at least a hundred different linguistic variables in the vague hope that they might be useful as predictors of readability. Incidentally, I have always had a soft spot in my heart for the genius who predicted readability from the percentages of words beginning *w*, *h* or *b* (which he considered easy) and of words beginning i or *e* (considered hard).

Luckily, most linguistic variables vary together. Carroll (196o) showed this in an analysis of 150 varied 500-word passages. Each was measured on 39 linguistic variables; it was also rated on 29 subjective scales by English literature scholars. All the measurements were then correlated and subjected to factor analysis. This is a process of mathematical magic that reveals which variables vary together, or to put it differently, factor analysis shows which measurements are largely measuring the same factor, that is, the same aspect of the passages. Thus there was found to be a factor which was measured by subjective scales such as goodness of style, pleasantness and interestingness : anyone who distrusts statistical counts will be gratified to know that not one objective measure was associated with this factor of stylistic evaluation. The only other factors of importance were: one associated with judged vagueness and a high proportion of noun clauses; a factor associated with judged impersonality and a high proportion of

long words; and a factor associated with judged complexity and a high proportion of long sentences. Because people who can identify noun clauses with certainty are rather rare, this measure would not be very useful in practice as a predictor of readability. So we are left with word and sentence length as predictors. They are quite powerful predictors: mean word length correlates — 0.6 with children's comprehension scores, and mean sentence length correlates — 0.5.

A model for word comprehension

Long words are not much more difficult to utter than short ones; even small children often find them supercallifragilistic-expialidocious. Why, then, is length an indicator of difficulty? Well, by an historical accident, the everyday words in our language have Anglo-Saxon roots and these tend to be short, whereas learned words come from Classical languages which rejoiced in polysyllables. Quite apart from the fact that in English there are a hundred-odd function words which provide the grammatical glue that holds sentences together—so that they account for half the occurrences of words in any text—in everyday life one encounters everyday words and thus becomes familiar with their meanings. In contrast, learned words are so rarely used in speech, which provides most of our verbal experience, that one may be uncertain of the meanings of many of them. Obviously this is particularly true of children. So children, especially, are likely to find long words difficult simply because they do not know their meaning.

However, this does not seem to be the whole explanation, because even an adult having a large recognition vocabulary will find a book difficult if it is replete with sesquipedalian vocables. Evidently, long words are not difficult simply because they are rare but, rather, rare words present some kind of difficulty even if one does know their meaning. The nature of the difficulty presented by rare words becomes apparent from consideration of the oddity that common words are very, very, common and rare words are very, very rare. This statement is expressed mathematically in an equation called the Zipf–Mandlebrot Law.

In some work to be published elsewhere I have proposed an emendation to the law so that it will fit word distributions that actually occur, rather than the distributions that Zipf and Mandlebrot felt ought to occur.

More importantly, I have shown how the amended law can be derived. First I assume, in common with most other psychologists, that if you are making sense of the stimuli that impinge upon you at every moment, then you must be mentally labelling each stimulus according to the categories to which you think it belongs. Only by categorizing a stimulus do you recognize what it is. Now words act as stimuli. So to understand a word you must allocate it to one or more categories. But, if you are to perceive the word consistently, the number of your categories must be almost constant. And, given a fixed number of categories, the number of words that can be classified by only one category is quite small; the number that can be classified by a combination of two categories is larger; the number classifiable by three categories is much larger still, and so on. Word frequencies are distributed in a precisely similar pattern : the number of very, very frequent words is quite small; the number of very frequent words is larger; the number of merely frequent words is much larger still, and so on.

All that remains to be explained is why one-category words should be very, very common; why two-category words should be very common; why three-category words should be merely common, and so forth. That is easy. If we have to categorize words in order to understand them, it must be assumed that we have to search among our categories in order to produce a word. Both the process of searching the categories in order to utter a word and the process of classifying in order to understand it must take time. The amount of time available to utter or to understand a word, especially in speech, must be strictly limited. Therefore people will generally use one- or two-category words in preference to multi-category words.

Consider next the nature of a word characterized by a large number of categories. Clearly it would be a rather specialized word: its numerous categories would give it a very precise

meaning, and, being so specialized, it would not be a word that could be used in very many contexts, so it would have very few dictionary meanings. And, indeed, it turns out that the fewer meanings a dictionary-maker can find for a word, the rarer that word is likely to be.

There is also some evidence for the hypothesis that the probability of a particular word being produced is limited by time. A writer can take as much time as he likes to search for whatever words most nearly express the precise meanings he wishes to convey. A speaker is much more pressed: if he pauses too long, his listeners—even the speaker himself—may forget how his utterance began, so that he will fail to communicate effectively. The speaker will therefore tend to use the first more or less suitable word that occurs to him. So rare words should occur even less frequently in speech than in writing. And, indeed, that is the case.

To put it in a nutshell: in English, long words tend to be those which are rare and therefore less familiar, so that their meanings are less likely to be known; but, even if one does know the meanings of a rare word, it takes more time and effort to locate it in one's mental classification system.

A model for sentence comprehension

Except to the very young, there is nothing intrinsically difficult about long sentences, while some short sentences baffle almost everybody. To adapt an example from Yngve (1960) it takes more than a moment to unravel 'What what what he wanted cost in London would buy in Southampton was amazing', though that sentence is merely equivalent to 'It was amazing what could be bought in Southampton for the cost in London of what he wanted.' The difficulty of the first sentence is obviously due to its grammatical complexity. Generally, however, complex sentences tend to be long. So the question 'Why are long sentences difficult to understand?' can be rephrased as 'Why are grammatically complex sentences difficult to understand?'

Before I can answer that one, I must define grammatical complexity, but before one can define grammatical complexity one

has to decide which grammar to use . . . English grammar, of course! But the point is that there are many quite different grammars of English. A grammar is a systematic description of the patterns of the language, and just as you can describe the best route from London to Southampton in a hundred different ways, all equally correct, so the patterns of English can be differently described. Yet not all descriptions of the route will be equally useful: I would prefer to know when to turn left or right, rather than east or west, because I have not got a compass. Similarly some grammars of English will be useful for one purpose, some for another.

I wanted a grammar which appeared to have some psychological meaning, so I chose string analysis. According to this grammar, devised by Zellig Harris (1962), every sentence contains a so-called centre string of words, none of which can be deleted without seriously injuring the sense: to this string as a whole, or to any constituent word of this string, there may be adjoined a further string of one or more words, called an adjunct; to each adjunct, or its constituents, further adjuncts may be adjoined, and so on.

To analyse a sentence, all you have to do is delete the largest string of contiguous words that you can, while still leaving a meaningful sentence; from this remainder delete the next largest string; and continue the process until only the centre string is left. Each of the strings you have excised is an adjunct together with its attendant adjuncts, which can be stripped off by a similar deletion process.

The word or string to which an adjunct is adjoined is called its head. Adjuncts nearly always appear immediately before or after their heads, but this means that more often than not they interrupt the string in which their head occurs. For example, in Meredith's line 'She whom I love is hard to catch', 'whom I love', being an adjunct to 'she', interrupts the centre string 'she is hard to catch'. There being only one short interruption, that sentence is not particularly difficult to understand. But you have only got to make a few adjunctions to adjunctions and you will produce a sentence which, while perfectly 'correct', is very far from easy, such as : 'The delights of reading Henry James are,

except to those so baffled by the length of some sentence the author felt he had to employ in order to express a nuance that he intended to convey to every reader capable of appreciating the *mot juste* they lose the thread, exquisite.'

What I hope is clear is that grammatical complexity can be measured in terms of the number of adjoined words separating the segments of a string. This measure I call separation. To show that separation does reduce comprehension, though mere sentence length does not, I took two 500-word passages and rewrote them into shorter sentences having the same total separation; in a further version I also reduced the separation. When I tested students' comprehension of the various versions, using the doze procedure, I found there was no significant difference in comprehension between the original passages and the versions having merely shorter sentences. But the differences in comprehension between these versions and the third version were so consistent that the odds against the differences being due to chance are a million to one.

My theory to account for the psychological difficulty induced by separation is this: assume that the process of perceiving—that is, categorizing—a segment of a sentence sets up some kind of pattern of activity in the brain: assume also that this pattern of activity decays rather rapidly, but that comprehension depends upon one combining patterns evoked by grammatically related elements, so that they have to be retained simultaneously. It follows that the greater the separation between two related segments, the greater is the probability that the pattern evoked by the first segment will have decayed beyond recall before the second is perceived, so that the probability of complete comprehension of the sentence in which the segments occur is reduced.

This model can be extended to account for the way in which a written sentence can convey the same sort of information as a direct experience. I suggest that, when the reader mentally combines and re-combines the segments of a sentence into larger and larger strings, the patterns of activity in the brain corresponding to each segment are being combined and re-combined, and

therefore modified, until finally a pattern is produced which strongly resembles the complex pattern which would be evoked by a corresponding direct experience.

DEFECTS OF READABILITY FORMULAS

Let us now get back to readability formulas. One of the most used children's formulas, devised by Washburne and Morphett (1938), demands that one should count the number of different words in a sample of ,000. It is absolutely essential to keep to the recommended sample size: people sometimes ignore this because they do not realize that the logarithm of the number of different words in a sample varies as the logarithm of the length of the sample. Not only is the Washburne-Morphett formula liable to misuse, it is also extremely tedious to calculate.

Probably the best formula so far, for both children and adults, is the one devised by Dale and Chall (1948). It has the same form as the one set out above, but for the measure of word difficulty it uses the percentage of words outside Dale's list of 3,000 words familiar to American fourth-grade children. For adults, at least, Bongers's (1947) list of the 3,000 words most widely used in the written language might be more suitable. In any case, looking up words in a list is again extremely tedious. So, if you will countenance an increase of standard error from 0.77 to 0.85 of a grade, you can follow Flesch in using average length as the index of word difficulty. If you will put up with a trivial increase in standard error to 0.9 of a grade you can economize on effort still more by adopting Gunning's (1952) device of merely calculating the percentage of words of three or more syllables.

A serious defect of nearly all published formulas is that they give no indication as to how their accuracy is affected by variations in sample size. Even worse, they make two assumptions which are almost certainly false.

One assumption is that the relationship of word and sentence difficulty to readability is linear. Let me illustrate what this means. Figure 1 is a graph of the age at which a child finds a book just readable plotted against units of word and sentence difficulty

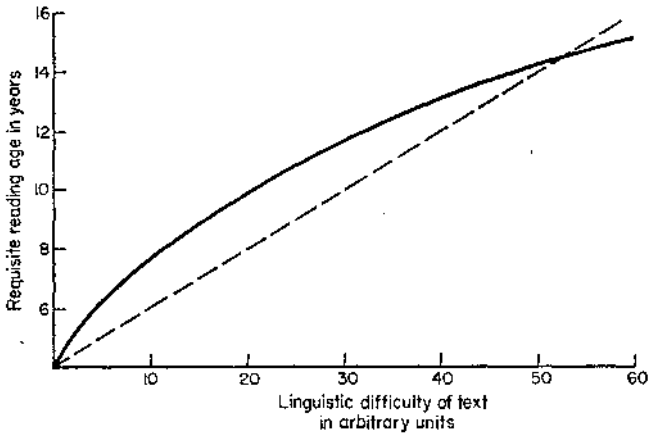

Figure 1.
The solid line shows the type of relationship which holds between the linguistic difficulty of a text and the average reading age needed for it to be acceptable. The dashed line shows the relationship assumed by a typical readability formula.

added together. What is assumed in every formula is that the relationship of readability to linguistic difficulty can be represented on such a graph by a straight line, although a little research would probably show that the relationship should be represented by a curve such as I have drawn.

The second assumption was shown to be false by Gray and Leary back in 1935, but their findings have been consistently ignored by every formula-maker since. This assumption is that, whatever the skill of a reader, the degree to which he is deterred by word difficulty compared with sentence difficulty remains constant. Again we can represent this assumption by a graph. If word length is plotted against sentence length, as in Figure 2, then it is assumed that contours of equal difficulty, measured by the age at which a child finds material of a given average word and sentence length just readable, can be represented by a series of parallel straight lines (b). In fact Gray and Leary showed that

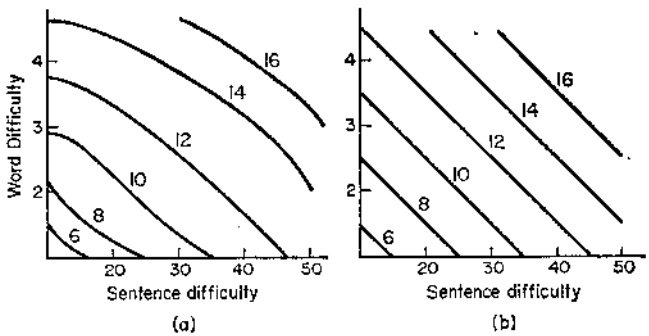

Figure 2.
(*a*) Estimated limits of word and sentence difficulty (measured in arbitrary units) acceptable at various reading ages.
(*b*) Limits of acceptable difficulty assumed by a typical readability formula.

a poor reader has such difficulty with unfamiliar words that the difficulty of long sentences is insignificant by comparison, whereas the chief difficulty for more mature readers is presented by long and complex sentences. Our contours of equal difficulty should therefore be shown as a series of curves (a). Admittedly, my own studies show that, for adult readers of periodicals, word and sentence difficulty vary together so closely that a measure of word length alone can serve as a readability predictor. None the less, the findings of Gray and Leary, which were based mainly on studies of very weak readers, are probably also valid for children.

The effect of these two assumptions taken together is illustrated in Figure 3. The word and sentence lengths of a great number of books are supposed to have been plotted upon a two-dimensional base. For each book we assume that we have discovered the average age at which children read and enjoy it. The average age at which the book is readable is represented by height above the base. Clearly, if we take enough different books the average ages will form a surface of varying height, technically termed a regression surface. It is assumed in readability formulas

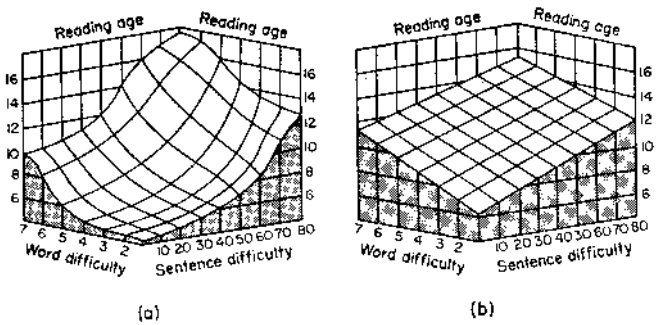

Figure 3.
(*a*) Likely nature of a regression surface representing the average reading ages at which texts of various degrees of word and sentence difficulty (measured in arbitrary units) are acceptable.
(*b*) The type of regression surface assumed in readability formulas.

that this regression surface is very nearly a plane, as shown on the right, whereas a true representation of the data would probably curve up and down, as on the left.

THE ALTERNATIVE: READABILITY TABLES

The crazy thing is that readability formulas have got so many figures in them that the equations are hardly ever used directly. Instead people prefer to read the predicted scores off a simple little chart called a nomogram. My feeling is that, if you are willing to refer to a chart, which represents a regression surface so flattened out that it is a serious distortion of the true surface, then you should be willing to refer to a table which gives the actual characteristics of readers of books having various average word and sentence lengths.

Such readability tables could be prepared in the following way.

We get teachers in a large sample of schools to ask their pupils each to write down a list of books which they have read and

thoroughly enjoyed in the past three months. The teachers then compare these lists with a list of 240 books, amalgamating six groups of 40 books each, which librarians find to be the most often borrowed by children in each of the age ranges: 6 or below, 7 to 8, 9 to 10, 11 to 12, 13 to 14, and 15 or above.

Ignoring the librarians' estimates, the teacher now writes against each book on the librarians' list the ages of any of her pupils who claim to have read and enjoyed it. The scatter of ages would be reduced if mental ages were recorded, but I suppose that in practice one would have to make do with chronological ages. Anyhow the idea of this procedure is to make sure that no more than 240 books have to be analysed for word and sentence length, and, above all, to reduce the effect of pupils' claims to have enjoyed books which are less popular, leading one to suspect that they have not really read them. Finally all the mass of data would be fed into a computer which would produce a smoothed regression surface in the form of a readability table.

REFERENCES

ABRAMS, M. (1963) *Education, Social Class and Newspaper Reading.* London: Institute of Practitioners in Advertising.

BONGERS, H. (1947) *The History and Principles of Vocabulary Control.* Woerden, Netherlands: Wocopi.

BURT, C. (1959) *A Psychological Study of Typography*. Cambridge University Press.

CARROLL, J. B. (1960) 'Vectors of prose style', in Sebeock, T.A. (ed.) *Style 'in Language.* Cambridge, Mass.: M.I.T. Press and John Wiley.

DALE, E., and CHALL, J. S. (1948) `A formula for predicting readability', *Educ. Res. Bull.,* 27, 1-20 ; 28, 37-54.

ENGLISH, H. B., and ENGLISH, A. C. (1958) *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms.* London: Longmans, Green.

FLESCH, R. F. (1948) 'A new readability yardstick', *J. Applied Psychol.* 32, 221-33.

GRAY, W. S. and LEARY, B. (1935) *What makes a book readable.* University of Chicago Press.

GUNNING, R. (1952) *The Technique of Clear Writing.* New York: McGraw-Hill.

HARRIS, Z. S. (1962) *String Analysis of Sentence Structure.* The Hague: Mouton.

McCALL W. A. and CRABBS, S. M. (1925) *Standard Test Lessons in Reading.* New York: Teachers College, Columbia University (rev. edn. 1950).

POWERS, R. D., SUMNER, W. A., and KEARL, B. E. (1958) 'A recalculation of four adult readability formulas', *J. Educ. Psychol.,* 49, 99-105.

TAYLOR, W. L. (1953) "Cloze procedure": A new tool for measuring readability', *Journalism Quart.,* 33, 42-8 and 99.

TINKER, M. A., (1965) *Bases for Effective Reading.* Minneapolis: University of Minnesota Press.

WASHBURNE, C., and MORPHETT, M. V. (1938) 'Grade placement of children's books', *Elementary School y.,* 38, 355-64.