

This paper was downloaded from

**The Online Educational Research Journal**  
**(OERJ)**

[www.oerj.org](http://www.oerj.org)

OERJ is an entirely internet-based educational research journal. It is available to anyone who can access the web and all articles can be read and downloaded online. Anybody can submit articles as well as comment on and rate articles. Submissions are published immediately provided certain rules are followed.

# **Matching texts to learners: the usefulness of readability formulae**

**Dahlia Janan**

**(Sultan Idris University of Education, Malaysia) and**

**David Wray**

**(University of Warwick, UK)**

## **Abstract**

Making decisions about matching texts to learners is something regularly required from teachers at all levels. Texts which are too simple for the learners in question are not likely to help develop these learners cognitively: texts which are too difficult are not likely to be read at all. Making such decisions about text suitability is often described as measuring the ‘readability’ of texts, and for a long time this measurement was treated as unproblematic. However, although definitions of readability have generally recognised that it must take into account characteristics of both the text and the reader, approaches to measuring readability have tended to focus on the text alone. Measurement has often been approached through the use of readability formulae which use such features as vocabulary difficulty and sentence length. In this paper we review these formulae and offer a critique of them, based on the research literature and on our own empirical work. We offer the conclusion that readability formulae by themselves cannot ensure a match between text and learner, and they need to be supplemented by teachers’ professional judgements.

## Introduction

It has been argued that the most important pedagogic decision that teachers make is “making the match” (Fry 1977), that is, ensuring that learners are supplied with reading materials, in whatever subject, that are at an appropriate level of difficulty for them. Learners who are given reading materials that are too easy are not challenged and their learning growth can be stunted (Chall and Conard 1991). Learners who are given reading materials that are too difficult can fail to make progress (Gambrell, Wilson, and Gantt 1981), are frequently off task and may exhibit behaviour problems (Anderson, Wilkinson and Mason 1987), or may become so frustrated that they simply give up (Kletzien 1991). Making the match is therefore a crucial skill for teachers, and it is also important for those who, while not teachers, do produce written material which they desire to be read and understood by other people. It has long been considered that the successful exercise of this skill requires knowledge of the readability level of materials. The Bullock Report in 1975 commented that, “a particularly important teaching skill is that of assessing the level of difficulty of books by applying measures of readability. The teacher who can do this is in a better position to match children to reading materials that answer their needs” (DES 1975, 113). Defining and exploring this concept of readability gave rise to a significant body of research from the 1920s to the early 1990s, one of the major outcomes of which was the production of a large number of “readability formulae”, that is, approaches to analysing texts which were designed to give a quantitative measure of the “level” a reader would need to be at in order to read and understand a particular text successfully. Even into the 21<sup>st</sup> century, new readability formulae have continued to be produced, reflecting the attractiveness of such an approach to matching texts to learners.

### *Definitions of readability*

Various definitions of the concept of readability have emphasised the elements in a text which were associated with comprehension (or lack of it) on the part of the reader: that is, the understanding of words, phrases and ideas in the passage. Parts of the concept also referred to a person’s ability to read a given text at an optimum speed. Finally, the concept also included motivational factors which affected a reader’s interest in reading a text. According to Dale and Chall (1948) these three elements of the definition of readability were not separate, but interacted with each other. To explain this interaction, Gilliland (1974, 13) provided the following example:

*‘...in a scientific article, complex technical terms may be necessary to describe certain concepts. A knowledge of the subject will make it easier for a reader to cope with these terms and they, in turn, may help him to sort out his ideas, thus making the text more readable. This interaction between vocabulary and content will affect the extent to which some people can read the text with ease’.*

Thus, definitions of readability have never been entirely text-centric. However, despite the established claim put forward by Harris and Hodges (1995, 203) that, “Text and reader variables interact in determining the readability of any piece of material for any individual reader”, approaches to the measurement of readability have not generally reflected such interactive definitions. Readability measurement has instead usually involved objective estimates of the difficulty level of reading material derived from the application of formulae which generally took into account sentence and vocabulary difficulty.

### *The development of readability formulae*

As mentioned above, there was a great deal of development in readability research between the 1920s and the early 1990s. The growth of attention to this research area was caused by the urge to emphasize quantification in developing a scientifically based curriculum. From the middle of the 1990s, however, developments in this research area decreased significantly. As one example of this declining research interest, we can examine the publications included in the JSTOR online archive of journal articles. In the 15 year period from 1965 to 1980, 1298 articles referenced by the key word “readability” are available in the JSTOR collection, and from 1980 to 1995, a further 1590 articles are available. In the subsequent 1995 – 2010 period, however, only 672 new articles appeared.

This decrease in research was undoubtedly related to explorations of the use of readability formulae. Many criticisms were made of these formulae, with research suggesting that they were not reliable and valid predictors of text difficulty (e.g. Redish and Selzer 1985; Bruce, Rubin and Starr 1981). It seemed that the ideal readability concept as suggested by Dale and Chall (1948), which would involve the text and the reader, was not measured, and may not have been measurable. The readability concept tended to focus on an objective estimation of text comprehension difficulty without involving the readers of that text. Many of the assumptions made about readability, and arguments as to its weakness as a concept, are automatically associated with readability formulae, because these formulae were the best known product of this field of research. Yet there has always been a clear attractiveness about these formulae, as a brief review of their history will demonstrate.

The earliest readability formulae were produced between 1921 and 1934, including such examples as those from Thorndike (1921) and Vogel and Washburne (1928). At that time, primary attention was given to vocabulary as the basis for predicting readability, and emphasis was placed on *Thorndike’s Teacher’s Word Book* as the basis for judging vocabulary difficulties and unfamiliarity (Klare 1963).

The next sets of readability formulae were produced in the years between 1934 and 1952, by educators such as Dale and Tyler (1934) and McClusky (1934). This period also saw the advent of the much better known Flesch (1948), Dale and Chall (1948), and Gunning (1952) formulae, whose more recent iterations are still used today. The focus of these formulae was on including more and different factors as variables, with less dependence on the Thorndike word count.

Later formulae tended to be developed for much more specialised purposes, for example, specific audiences, such as primary school students, e.g. Spache (1953). Formulae continued to develop despite criticisms of their reliability. The arrival of cloze procedure as a tool for measuring readability in the mid-1950s stimulated the development of new criteria, new formulae, computerized versions, and the continued testing of text variables (Dubay 2004). Bormuth (1966), for example, showed that changes in a number of text variables in addition to vocabulary and sentence length could affect comprehension. Bormuth further claimed that cloze testing was appropriate for measuring not only the difficulty of the whole passage but also the difficulty of individual words, phrases, and clauses.

Readability formulae continued to develop with new formulae such as McLaughlin’s (1969) *Simple Measure of Gobbledygook* (SMOG). Chall and Dale (1995) updated their 1948 formula, largely by updating its essential list of 3,000 easy words that had first been

assembled 47 years earlier. More recently, a greater number of computerized formulae have been developed, such as the Lexile Framework (Lennon and Burdick 2004) and ATOS (Milone 2008). The Lexile Framework formula uses several variables, such as average sentence length and word frequency (Stenner et al 2006). The ATOS readability formula was formed with the purpose of providing an “open” formula that would be available to the educational community free of charge and began with an extensive study of readability. It includes three variables, words per sentence, average difficulty level of words, and characters per word (Milone 2008; Renaissance Institute 2000).

To sum up, readability formulae have gone through several phases and changes. Early formulae were mainly dependent on Thorndike’s Word List. Following this, they developed by adding new variables. They are undoubtedly still popular and are now much easier to use with most being available on the internet. Even Microsoft Word provides inbuilt readability measures (using the Flesch Reading Ease and the Flesch-Kincaid Grade Level formulae). However, they have still received heavy criticism over the years.

### *Criticisms of readability formulae*

Readability research has focused on devising procedures and instruments that can reliably and validly distinguish easier from more difficult reading materials. Underpinning this research has been a belief that reading difficulty is influenced by four factors in the reading materials themselves, namely content, stylistic elements, format and organization. Stylistic elements appeared to be the most ‘amenable to reliable quantitative measurement and verification’ (Chall 1974, 156). These elements included factors such as vocabulary load, sentence structure, idea density, and human interest, which have all been found to be significantly related to reading difficulty.

Vocabulary load has been construed as a combination of vocabulary diversity and vocabulary difficulty. According to Chall (1974, 157), the fewer different words in a text, the easier that text was to read. Ways to measure vocabulary difficulty included either reference to a set list of words or were approached through word length.

Sentence structure was also found to be significantly related to comprehension difficulty (Chall 1974). The best way to measure sentence structure was believed to be by sentence length (Chall 1974). Generally, the longer the sentences were, the harder the text was deemed to be. Apart from examining sentence structure, researchers were also interested in estimating sentence difficulty by the number of complex sentences, the number of simple sentences, and sentence length estimated by a count of syllables.

Sentence structure and vocabulary load have been the most commonly used variables in readability formulae. Accordingly, several assumptions were developed that:

- the smaller the number of different words, the easier the material was to read;
- the larger the proportion of unfamiliar or long words in a text, the harder it was for the reader to grasp the meaning;
- the longer the sentences, the harder the text;
- the simpler the sentences, the easier the text (Chall 1974).

These assumptions have underpinned a series of criticisms of readability formulae in particular, and of readability in general. Bruce, Rubin and Starr (1981) pointed out that readability formulae did not take into account current knowledge about the reading process.

They included sentence length and word difficulty but were not able to measure other factors that make a text difficult, such as the degree of discourse cohesion, the number of inferences demanded, the number of items to remember, the complexity of ideas, rhetorical structure, dialect and required background knowledge. Also, they attempted to measure text difficulty isolated from other elements such as the context of a text's use, the reader's motivation, interest, competitiveness, value and reading purpose (Bruce, Rubin and Starr 1981).

A further weakness has a statistical basis. Stoke (1978) examined seven types of readability formulae namely the Flesch, FOG, SMOG, Power-Sumner and Kearl, Farr-Jenkins-Peterson, Dale-Chall, and a simple count of "hard" words. He found that although these formulae produced a high inter-correlation, they gave widely differing grade levels for the same texts. In other words, these formulae agreed on which texts were difficult, but not on the level of that difficulty.

It has also been suggested (Davison and Kantor 1982) that changes made to a text, on the basis of readability formulae, in order to make it easier to read, may actually make it harder to understand. Such changes included splitting complex sentences into component clauses and changing vocabulary items, amongst others. Davison and Kantor (1982) argued strongly against the use of readability formulae as a guide to writing graded texts, and urged experimental research to define the real factors constituting readability. As a result, two professional associations in the USA, the *International Reading Association* and the *National Council of Teachers of English*, called for the cautious use of readability formulae and, indeed, a moratorium on their use (Michelson 1985; Anderson et al 1985).

Other criticisms have been made of the use of readability formulae:

- grade-level formulae were designed for children's school books and not the adult material to which they have often been applied (Redish and Selzer 1985; Redish 2000);
- no commonly used formulae were developed for technical materials (Redish 2000);
- readability formulae only measure what can be counted (Redish 2000);
- they assume that all readers are more alike than different (Redish 2000);
- most of what makes a document usable is not included in readability formulae (Chambers 1983; Redish and Selzer 1985; Redish 2000);
- readability formulae do not work on forms, web pages, or documents with lots of lists (Redish 2000);
- they are not very reliable as predictors of reading comprehension (Chambers, 1983; Fuchs, Fuchs and Deno 1983; Redish 2000; Stoke 1978; Sydes and Hartely 1997; Templeton, Cain and Miller 1981);
- changing a text to improve readability scores does not automatically result in improved comprehension (Chambers 1983; Pichert and Elam 1985; Redish 2000; Sydes and Hartley 1997).

Given such a level of critique, it might be expected that the use of readability formulae would have diminished hugely and, to a large extent, that is true in educational contexts. In other fields, however, readability formulae are still used heavily (e.g. Cronin, O'Hanlon and O'Connor 2011; Freda 2005). Badarudeen and Sabharwal (2010), for example, report on their use of a variety of formulae to judge the readability of medical patient education materials. Their critique of the use of formulae is limited to the observation that "there is no consensus as to which readability formula is best suited for assessing patient education materials. In general, it is preferable to use more than one readability method to improve the validity of the

results” (Badarudeen and Sabharwal 2010, 2574). Guo, Zhang and Zhai (2011) report on their integration of a readability index into the Twitter search engine and have the following to say about readability formulae: “Although they have their limitations, such as overemphasis on observable, morphological or syllabic features of word or text, neglect of readers’ interest and pre-existing knowledge, readability formulas provide valuable information, and are becoming more popular than ever” (103). It appears that, despite the critiques, readability formulae are still perceived to have a useful function in a number of fields. It was partly to re-examine this functionality that we carried out the present study.

## **The present study**

The study reported here was a small part of a much broader study which attempted to revisit the conceptual model of readability and to modify this to take account of recent developments, particularly in conceptions of the reading process (see Janan, 2011 for a full account).

The aim of this element of the study was to make comparisons between the outcomes of a number of readability formulae when applied to a range of texts. The texts involved were those selected by the researchers as likely to be suitable for young readers and those selected by these readers themselves. The readability formulae used were a mixture of the most widely used formulae developed from the 1950s to the modern era (2008). The research questions were:

- How consistent were the formulae in distinguishing hard texts from easier texts?
- How consistent were they in suggesting a readability level for individual texts?

## ***Participants and texts***

The study involved 32 randomly selected school children aged from 6 to 11. These children were requested to bring along to a meeting with the researcher any book, or other reading material, that they enjoyed reading and which they thought was neither too easy nor too difficult for them.

Randomly selected extracts of these 32 texts (between 100 and 400 words depending on the overall length of the text) were then put through a number of readability formulae: FOG (Gunning, 1952), Spache (1953), SMOG (McLaughlin, 1969), Flesch-Kincaid (Kincaid et al., 1975), Dale-Chall (Chall & Dale, 1995) and ATOS (Milone, 2008). One of the reasons for selecting these six formulae was their popularity over time, but also the fact that they are ‘open standard’, that is they can be applied to any material without the payment of a fee. The aim was to derive a readability index for these initial texts which could then be used as a benchmark index to guide the selection of further reading texts for these children. Subsequently another text was selected for each child which, as part of the overall design of the research, was planned to be slightly difficult, and whose difficulty was again measured by the six formulae. We thus had a bank of 64 texts, an extract of each of which had been checked with a number of readability formulae

## ***Comparing readability formulae***

Although there are common factors which most readability formulae include in their measurement procedures, there are nevertheless some differences between formulae in terms

of their major focus points. Some were designed to measure the readability of different types of text for different readers at particular ability levels. The FOG formula, for example, was developed specifically for adult level materials, whereas the Dale-Chall and Flesch Kincaid formulae were developed for materials for readers from primary school to adulthood. The Spache formula, on the other hand was designed to be used with materials aimed at primary school children.

One of the major differences between formulae, however, lies in the inclusion or not of a word familiarity factor in the calculations. There is a question about the validity of the means used to determine this familiarity, with the most common means being the use of a pre-set list of 'easy' or 'familiar' words. Perera (1980) noted that such word lists were based on frequency counts done in the United States, although the formulae were still used in Britain, where patterns of vocabulary use were different. A comparison of the revised Spache (1974) list (American) with a British frequency count of children's written vocabulary (Edwards and Gibbon 1973) reveals some discrepancies. Words such as *bonfire*, *doll*, *fairy*, *football* and *mummy* are listed as familiar words in the British list but not the American, whereas words like *cabin*, *candy*, *gift*, *parade* and *neighbourhood* are listed as familiar words to American children but not to British. It has also been suggested that, 'average word frequency is not a good predictor because many words are common at a certain age or level, but then become uncommon – such as "kitten". But in cases like these, infrequency at higher grade level does not make them difficult words' (Milone, 2008, 6).

Comparability between readability formulae should not, therefore, be taken for granted. In this study our aim was to try to establish as agreed a level of difficulty for each text as was possible, and therefore we applied more than one formula. The formulae that we used were FOG (Gunning 1952), Spache (1953), SMOG (McLaughlin 1969), Flesch-Kincaid (Kincaid et al 1975), Dale-Chall (Chall and Dale 1995) and ATOS (Milone 2008). One of the reasons for selecting these six formulae was their popularity over time, but also the fact that they are 'open standard', that is they can be applied to any material without the payment of a fee.

### ***Procedures***

The 64 texts were photocopied and then saved as RTF files, using Optical Character Recognition software. The texts were analysed using the six readability formulae via the Words Count website (<http://www.wordscount.info/>). This website provided automated readability indices using FOG, Spache, SMOG, Flesch-Kincaid, and Dale-Chall. For ATOS the Renaissance Learning website was used (<http://www.renlearn.com/ar/overview/atos/>). The six readability scores, expressed in terms of US grade levels, for each of the 64 texts were then listed and entered into SPSS software for later analysis. (It is generally accepted that to transform a US grade level to a chronological age, one simply adds 6.)

### ***Analysis***

The six readability formulae are each calculated differently and each makes use of a slightly different range of predictor variables. There are five predictor variables involved: word length, grade level of word, sentence length, unfamiliar or difficult words and polysyllabic words. Table 1 shows which of these predictor variables are used by each formula.

*Table 1 about here*



The variable most frequently used by these formulae is sentence length. But it should be noted that none of the predictor variables is used by all the formulae. The SMOG formula is unique in that it uses only one predictor variable, that of polysyllabic words.

Statistical tests were carried out to check the consistency and the relationships between the six formulae in terms of their predictions of levels of text difficulty. These statistical analyses involved:

*Consistency estimation.* The aim of this was to demonstrate the consistency among the formulae in ranking the texts in order of their levels difficulty. The Spearman rank order correlation coefficient was used in this procedure.

*Comparison of the grade levels.* The aim was to demonstrate the extent to which formulae agreed with one another in predicting the grade levels of the 64 texts. Paired-sample T-Tests were used for this purpose.

## **Results**

### *Consistency Estimation of the Formulae*

The consistency estimation of the six formulae in predicting the difficulty levels of the texts was assessed by the use of Spearman's rank order correlation coefficient. Table 2 presents the results of the comparison between the order of difficulty of the 64 texts produced by each formula.

*Table 2 about here*

Table 2 shows that a *very high* statistically significant correlation was found among the SMOG, FOG, Spache, and Flesch-Kincaid formulae in predicting the grade level of the texts involved. These formulae produced *almost* (but not quite) the same results in judging whether the text was easy or difficult to read. These high correlations were achieved *in spite of* the fact that these four formulae did not all share a single predictor variable in common. Three of them did share the use of sentence length as a variable, but this variable was also used by the Dale-Chall and ATOS formulae, whose correlations were not so high.

The *highest* statistically significant correlation was between the SMOG and the FOG formulae ( $\rho=.98$ ). In other words, the SMOG and FOG formulae produced virtually the same results in ranking the 64 texts in order of reading difficulty. The only common predictor variable to these two formulae was the use of the number of polysyllabic words in a text.

The ATOS formula did have a *high* statistically significant correlation ( $\rho= .68$  or higher) with the SMOG, FOG, Spache, and Flesch-Kincaid formulae. It should be noted that this formula used two predictor variables, word length and grade level of words, which none of the other formulae used.

The Dale Chall formula, on the other hand, actually showed a negative correlation with the results of all the other formulae, although this was a *medium* statistically significant correlation only with the SMOG and FOG formulae. This means that the Dale-Chall formula

was likely to predict certain texts as easy or difficult, the reverse of the way they would be judged by other formulae. This was surprising, because this formula shares the use of the two predictor variables, sentence length and unfamiliar words, with several other formulae. In judging unfamiliar words, however, Dale-Chall does use a different list of ‘easy words’ to that used by the Spache formula.

Generally, the data here suggest that, although there was some consistency in ordering texts according to difficulty levels between the FOG, SMOG and Flesch-Kincaid formulae, the consistency levels among the other formulae varied.

#### *Grade level predictions*

We also calculated the average grade levels predicted by each of the six formulae. Table 3 shows this.

#### *Table 3 about here*

The data show that the six formulae yielded different results for the mean text grade levels predicted for the same 64 texts. The Dale-Chall formula had the *highest* mean grade level (9.88), whereas the ATOS had the *lowest* (3.13). This indicates a range of predictions for the difficulty levels of texts concerned here of over six and a half chronological years. Texts which the Dale-Chall formula predicted were suitable for fifteen years olds were recommended by ATOS as suitable for nine year olds.

#### *Individual paired formulae comparisons*

We then examined the differences between pairs of formulae in terms of the mean grade levels they produced for the 64 texts. Paired-sample T-tests was carried out to identify whether there were statistically significant differences between these mean grade levels. Table 4 shows the results.

#### *Table 4 about here*

Table 4 shows that the *highest* difference between the text mean grade levels was between the Dale Chall and ATOS formulae (difference = 6.75), with the difference being statistically significant ( $t=21.98$ ,  $df=62$ ,  $p<.01$ ). The only comparison which showed *no* statistically significant difference between the mean grade levels was that between the Flesch-Kincaid and FOG formulae, where the mean difference was .09 ( $t=-40$ ,  $df=62$ ,  $p=.69$ ). It can therefore be concluded that only the Flesch-Kincaid and the FOG formulae produced similar results for the grade levels of these 64 texts.

The largest differences in the mean grade levels predicted by these formulae were between the Dale-Chall formula and the others, with the Dale-Chall producing mean grade levels for the 64 texts of between 3 and 6 years higher.

In summary, the results of the formulae reliability analyses suggest that despite the fact that the SMOG, FOG, Flesch-Kincaid, Spache and ATOS formulae were found to correlate quite strongly when predicting the grade level of these texts in terms of the rank orders of difficulty they produced, there were some widely differing grade level scores being produced by all the formulae. In other words, although the SMOG, FOG, Flesch-Kincaid, Spache and ATOS formulae generally agreed on which texts were easier or more difficult than other texts, they still assigned individual texts to different grade levels.

In the case of the Dale-Chall formula, not even a consistency of rank ordering was found with the other formulae. Dale-Chall tended to grade texts as at a higher level than the other formulae, but it also was prone to assign a text as easy, whereas the rest of the formulae predicted it as difficult.

## **Discussion**

A definition of readability as the “*ease with which a reader can read and understand a given text*” (Oakland and Lane 2004, 244) suggests the need to consider both reader and text in making judgements about reading ‘ease’. However, the measurement of readability has not generally reflected this definition and instead has focused on features in text language which appear to make texts easy or difficult to read (Harrison 1984). Pikulski (2002) has, thus, argued that: ‘Readability continues to be among the most discussed, misunderstood, and misused concepts in reading. It is all too commonly, but erroneously, thought to be a precise numerical score, obtained through the use of readability “formulas,” that indicates the level of difficulty of a text’ (1).

Measurements of text features, made through the application of readability formulae, have been extremely popular, but heavily criticized in terms of their validity and reliability. The formulae, it has been argued, fail to measure comprehension (Duffy 1985), fail to include a range of components vital to comprehension such as subject knowledge, motivation for reading, text genre, context and purpose of reading (Schriver 2000) and research has suggested that various formulae tend to produce significantly different results on the same text and an average score, taken over a passage, can conceal a wide range of variations of difficulty within a passage (Sydes and Hartley 1997).

The findings of the present study support those of Sydes and Hartley (1997). The analysis of our sample of 64 texts, carried out with six readability formulae (ATOS, Dale-Chall, Flesch-Kincaid, FOG, SMOG, and Spache), has demonstrated significantly different readability indices for the same text. It appeared that some of the formulae (but not all) were consistent in their ranking of texts in order of difficulty but were not consistent in their grading of each text, with up to a six year discrepancy between them. The study results are also consistent with those of Stoke (1978), in suggesting that among these formulae, there are some which classify a text as easy, whereas others classify it as difficult, and vice versa.

Our findings raise two major questions. Firstly, if readability formulae focus only on one half of the reader-text relationship at the heart of reading, how can we reconceptualise readability to focus upon both aspects? And secondly, if readability formulae have so many weaknesses (and our study is certainly not the first to point these out), then why have they continued to be used so widely, in such a range of areas?

### ***The readability paradigm***

Problems in readability research and in the use of readability formulae seem to result from a general failure to follow through on definitions, which have always insisted that there are two sides to any reading, and readability, event – the text and the reader. The actual measurement of readability has tended to be approached from within a particular paradigm, that is, that readability exists independently of a particular reader, and that the reader's comprehension can be predicted from an examination of text characteristics. This essentially positivist paradigm has viewed reading comprehension as an input and output process, put simply, getting meaning from the page. However, conceptualisations of reading and reading comprehension have changed and are now viewed as meaning-construction processes (Ruddell and Unrau 2004). Meaning no longer comes from the text, but from readers who bring their social and cultural backgrounds to an interaction with the text. Accordingly, the movement in reading research has suggested that an interpretivist approach is an appropriate alternative paradigm within which to study these processes. Reading research more recently has tended to focus on what happens in readers' minds during reading, and has employed error and miscue analysis (Goodman and Goodman 1977) and think aloud protocols (Pressley and Afflerbach 1995) to explore reading and comprehension processes as they happen. This has not proved unproblematic and the use of both error/miscue analysis and think aloud protocols have had their critics (McKenna and Picard 2006; Cotton and Gresty 2006). Nevertheless, the alternative paradigm has deeply affected views of reading, and, in turn, views of and research into readability.

This does not mean that text is no longer seen as important in readability, but rather that a way forward might be to view readability (and reading) from both positivist and interpretivist paradigms. Judgements about the difficulty levels of texts can only be made by taking into account the characteristics of the texts themselves *and* the characteristics of the readers who read them. A fuller examination of this principle, and an exploration of it in action, can be found in Janan (2011). For our purposes in this article, the implication is that readability formulae cannot tell us everything we need to know about the process of matching a reader to a text, and the information that some formulae do give us about textual features needs to be tempered by a close knowledge of the reader and his/her background, motivations, purposes for reading, attitudes etc.

### ***The continuing popularity of readability formulae***

The findings presented here have shown that there are problems related to the reliability and validity of the formulae used to assess readability. Consistency levels among the six formulae used in this study varied, and one of the formulae (Dale-Chall) stood out from the rest as being inconsistent in predicting the level of text difficulty. The remaining five formulae, notwithstanding their conflicting results in assessing the difficulty levels of individual texts, did achieve a relatively high correlation between themselves in terms of their rank ordering of the texts in terms of difficulty. This suggests that, while not offering a definitive picture of the reading difficulty of a particular text, they can generally be relied upon to distinguish between easier and harder texts. Their ability to do this is, in essence, the key to their continuing popularity. Most users of readability formulae will not, in fact, need to assign a precise readability level to any individual text, but will need to be able to judge whether certain texts are likely to be easier or harder to read than other texts. In classroom settings, and this is crucial, it is highly unlikely that the information given by a readability formula will be the only information a teacher will use in suggesting a particular text for a child to

read. Teachers know their children and will take this knowledge into account, even sub-consciously, as they make the decision about matching books to readers.

## Conclusion

In this article we have explored the phenomenon of readability formulae. This exploration has led us to outline the main areas in which these apparently simple and useful tools have been found to be inadequate. We have added our own evidence to this critique but we have concluded by recognising that, whatever their faults, readability formulae may still have a place in the armoury of a busy teacher of reading. Their chief advantage is their simplicity of operation, and here the benefits of a technologically rich era have played a very significant role. Any text or extract can very quickly be scanned into a computer, and put through a number of readability formulae, thus providing a teacher with a quick initial indication of text difficulty. Many publishers of texts for children, of course, will already provide such information for teachers.

It needs to be accepted, however, that this is an *initial* indication. Teachers also need to bring to bear their professional judgements, in terms of their knowledge of the children they teach, in making judgements about what might be suitable texts for these children to read. Readability formulae may have a place in a busy classroom, but it can never be as the *only* source of information about text difficulty.

## References

- Anderson, R., Wilkinson, I. & Mason, M. 1987. *Do errors on classroom reading tasks slow growth in reading?* Center for the Study of Reading Technical Report No. 404. Champaign: University of Illinois at Urbana-Champaign
- Anderson, R. C., E. H. Hiebert, J. Scott and L. Wilkinson. 1985. *Becoming a Nation of Readers*. Washington, DC: National Institute of Education.
- Badarudeen, S. and S. Sabharwal. 2010. Assessing readability of patient education materials: current role in orthopaedics. *Clinical orthopaedics and related research*. 468: 2572-80.
- Bormuth, J. R. 1966. Readability: A New Approach. *Reading Research Quarterly*. 1: 79-132.
- Bruce, B., A. Rubin and K. Starr. 1981. Why Readability Formulas Fail. *IEEE Transactions on Professional Communication*. 24: 50-52.
- Chall, J. S. and S. Conard. 1991. *Should textbooks challenge students? The case for easier or harder books*. New York: Teachers College Press.
- Chall, J. S. and E. Dale. 1995. *Readability revisited: the new Dale-Chall readability formula*. Cambridge, Mass.: Brookline Books.
- Chall, J. S. 1974. *Readability: an appraisal of research and application*. Epping, Essex: Bowker, for The College of Librarianship Wales.

- Chambers, F. 1983. Readability Formulae and the Structure of Text. *Educational Review*. 35: 3-13.
- Cotton, D. and K. Gresty. 2006. Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology*. 37: 45-54
- Cronin, M., S. O'Hanlon and M. O'Connor. 2011. Readability level of patient information leaflets for older people. *Irish Journal of Medical Science*. 180: 139-142.
- Dale, E. and J. S. Chall. 1948. A Formula for Predicting Readability. *Educational Research Bulletin*. 27: 11-28.
- Dale, E. and R. Tyler. 1934. A Study of the Factor Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability. In W. H. Dubay Ed., *Unlocking Language: The Classic Readability Studies*. 80-107. Costa Mesa, California: Impact Information.
- Davison, A. and R. N. Kantor. 1982. On the Failure of Readability Formulas to Define Readable Texts: A Case Study from Adaptations. *Reading Research Quarterly*. 17: 187-209.
- Department of Education and Science 1975. *A Language for Life*. London: Her Majesty's Stationery Office
- Dubay, W. H. 2004. *The Principles of Readability*. Costa Mesa, CA: Impact Information.
- Duffy, T. M. 1985. Readability formulas: What's the use? In T. M. Duffy & R. M. Waller Eds., *Designing Usable Texts* 113-143. New York: Academic Press.
- Edwards, R. P. A. and V. Gibbon. 1973. *Words Your Children Use*. London: Burke Books.
- Flesch, R. 1948. A new readability yardstick. *Journal of Applied Psychology*. 32: 221-233
- Freda, M. C. 2005. The readability of American Academy of Pediatrics patient education brochures. *Journal of pediatric health care*. 19: 151-6.
- Fry, E. 1977. *Elementary reading instruction*. New York: McGraw-Hill.
- Fuchs, L. S., D. Fuchs and S. L. Deno. 1983. *The nature of inaccuracy among readability formulas*. Minneapolis: Institute for Research on Learning Disabilities, Minnesota University
- Gambrell, L. B., R. M. Wilson and W. Gantt. 1981. Classroom observations of task-attending behaviors of good and poor readers. *Journal of Educational Research*. 74: 400-404.
- Gilliland, J. 1974. *Readability*. London: Hodder Arnold.
- Goodman, K. S. and Y. Goodman. 1977. Learning about Psycholinguistic Processes by Analyzing Oral Reading. *Harvard Educational Review*. 40: 317-33
- Gunning, R. 1952. *The technique of clear writing*. New York, NY: McGraw-Hill

- Guo, S., G. Zhang and R. Zhai. 2011. Integrating readability index into Twitter search engine. *British Journal of Educational Technology*. 42: 103–105
- Harris, T. L. and R. Hodges. 1995. *The literacy dictionary: the vocabulary of reading and writing*. Newark, Del.: International Reading Association.
- Harrison, C. 1984. Readability in the United Kingdom. *Journal of Reading*. 29: 521-529.
- Kincaid, J., R. Fishburne, R. Rogers and B. Chissom. 1975. *Derivation of new readability formulas Automated Readability Index, Fog Count and Flesch Reading Ease Formula. for Navy enlisted personnel. Research Branch Report 8-75*. Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN.
- Klare, G. R. 1963. *The Measurement of Readability*. Ames: Iowa State University Press.
- Kletzien, S. B. 1991. Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly*. 26: 67–86.
- Lennon, C. and H. Burdick. 2004. *The Lexile Framework as an approach for reading measurement and success*. Durham, NC: Metametrics Inc.
- McClusky, H. Y. 1934. A quantitative analysis of the difficulty of reading materials. *Journal of Educational Research*. 28: 276-282.
- McKenna, M. C. and M. Picard. 2006. Assessment: Revisiting the Role of Miscue Analysis in Effective Teaching. *The Reading Teacher*. 60: 378-380.
- McLaughlin, G. H. 1969. SMOG grading - a new readability formula. *Journal of Reading*. 13: 639-646.
- Michelson, J. 1985. IRA, NCTE take stand on readability formulae. *Reading Today*, 2, No 3: 1.
- Milone, M. 2008. *The development of ATOS: The Renaissance readability formula*. Wisconsin Rapids, WI: Renaissance Learning.
- Oakland, T. and H. Lane. 2004. Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests. *International Journal of Testing*. 4: 239-252.
- Perera, K. 1980. The Assessment of Linguistic Difficulty in Reading Material. *Educational Review*. 32: 151-161.
- Pichert, J. W. and P. Elam. 1985. Readability formulas may mislead you. *Patient Education and Counseling*. 7: 181-191.
- Pikulski, J. 2002. *Readability*. Boston: Houghton Mifflin.
- Pressley, M. and P. Afflerbach. 1995. *Verbal protocols of reading: the nature of constructively responsive reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Redish, J. and J. Selzer, J. 1985. The place of readability formulas in technical communication. *Journal of Reading Behavior*. 32: 46-52.
- Redish, J. 2000. Readability formulas have even more limitations than Klare discusses. *ACM Journal of Computer Documentation*. 24: 132-137.
- Renaissance Institute. 2000. *The ATOS readability formula for books and how it compares to other formulas*. Madison, WI: Renaissance Institute
- Ruddell, R. B. and N. Unrau, N. 2004. Reading as a Meaning-Construction Process: The Reader, the Text, and the Teacher. In Ruddell, R.B., Unrau, N. & H. Singer Eds., *Theoretical Models and Processes of Reading*. 1462-1521. Newark, Delaware: International Reading Association.
- Schriver, K. A. 2000. Readability Formulas in the New Millennium: What's the Use? *ACM Journal of Computer Documentation*. 24: 138-140.
- Spache, G. 1953. A New Readability Formula for Primary-Grade Reading Materials. *Elementary School Journal*. 53: 410-413.
- Spache, G. 1974. *Good reading for poor readers*. Champaign, IL: Garrard
- Stenner, A. J., H. Burdick, E. Sanford and D. Burdick. 2006. How Accurate Are Lexile Text Measures. *Journal of Applied Measurement*. 7: 307-322.
- Stokes, A. 1978. The reliability of readability formulae. *Journal of Research in Reading*. 1: 21-34.
- Sydes, M. and J. Hartley. 1997. A Thorn in the Flesch: Observations on the Unreliability of Computer-Based Readability Formulae. *British Journal of Educational Technology*. 28: 143-145.
- Templeton, S., C. Cain and J. Miller. 1981. Reconceptualising Readability: The Relationship between Surface and Underlying Structure Analyses in Predicting the Difficulty of Basal Reader Stories. *The Journal of Educational Research*. 74: 382-387.
- Thorndike, E. L. 1921. Word Knowledge in the Elementary School. In Dubay, W. H. Ed., *Unlocking Language: The Classic Readability Studies*. 9-43. Costa Mesa, California: Impact Information.
- Vogel, M. and C. Washburne. 1928. An Objective Method of Determining Grade Placement of Children's Reading Material. In Dubay, W. H. Ed., *Unlocking Language: The Classic Readability Studies*. 58-66. Costa Mesa, California: Impact Information.

### **Appendix: Readability formulae used in this study**

**FOG** (Gunning, 1952): this has become known as the easiest of all readability indices to work out, and this fact explains its popularity. The formula uses the variables of: 1. average number of words per sentence; and 2. percentage of polysyllabic words.



**Spache** (Spache, 1953): this has been widely used in United Kingdom largely because it was designed to be suitable for reading material below a difficulty level of about eleven years old. It uses the following variables: 1. the average number of words per sentence; and 2. the number of unfamiliar words, that is those not found in the original Dale list of 769 easy words.

**SMOG** (McLaughlin, 1969): this is the easiest and the quickest formula of all to work out by hand. It uses a single variable - the number of polysyllabic i.e. three or more syllable. words in 30 sentences.

**Flesh-Kincaid** (Kincaid et al, 1975): this formula involve two main factors: 1. sentence length calculated by dividing the total number of words in a passage by the total number of sentences, and 2. the number of longer words, calculated by dividing the number of syllables by the number of words in the passage.

**Dale-Chall** (Chall & Dale, 1995): The 1995 revision of the original 1948 version of this formula replaces reference to the Dale list of 769 easy words with the use of a new Dale-Chall list of 3000 common words. Average sentence length is also used as a contributing factor.

**ATOS** (Milone, 2008): ATOS is an ‘*open standard*’ readability formula which focuses on three variables: 1. words per sentence; 2. average grade level of words; and 3. average characters per word. Milone (2008) claims that use of the average grade level of words in the text has proved to be a better predictor of text difficulty than have the two other variables.

All of the above formulae produce outcomes expressed in terms of a US grade level.

## Figures

*Table 1: Predictor variables used in the six readability formulae*

Formula	Predictor variables				
	Word length	Grade level of words	Sentence length	Unfamiliar / Difficult words	Polysyllabic words
<i>FOG</i>			*		*
<i>Spache</i>			*	*	
<i>SMOG</i>					*
<i>Flesch-Kincaid</i>	*		*		
<i>Dale-Chall</i>			*	*	
<i>ATOS</i>	*	*	*		

Table 2: Spearman's rank order correlation coefficients rho. between the SMOG, FOG, Flesch-Kincaid, Spache, Dale-Chall and ATOS formulae

Formula	FOG	Flesch-Kincaid	Spache	Dale-Chall	ATOS
<b>SMOG</b> <i>Significance 2-tailed.</i>	.98 **	.93 **	.83 **	-.41 **	.70 **
<b>FOG</b> <i>Significance 2-tailed.</i>		.95 **	.84 **	-.47 **	.74 **
<b>Flesch-Kincaid</b> <i>Significance 2-tailed.</i>			.88 **	-.32 *	.68 **
<b>Spache</b> <i>Significance 2-tailed.</i>				-.14	.68 **
<b>Dale-Chall</b> <i>Significance 2-tailed.</i>					-.49 **

Number of texts N= 64

\*\*  $p < .01$  \*  $p < .05$

Table 3: The mean text grade levels predicted by the six readability formulae

Formulae	Number of texts	Mean text grade level predicted	Standard Deviation
<b>SMOG</b>	64	6.64	2.31
<b>FOG</b>	64	5.80	2.40
<b>Flesch-Kincaid</b>	64	3.96	2.29
<b>Spache</b>	64	4.05	0.69
<b>Dale-Chall</b>	64	9.88	1.20
<b>ATOS</b>	64	3.13	1.59

Table 4: Paired-sample T-Tests of the differences between the mean grade levels of the six formulae

Formulae	FOG	Flesch-Kincaid	Spache	Dale-Chall	ATOS
<b>SMOG</b> Differences between means t Significance two-tailed.	.84	2.68	2.59	3.24	3.51
	6.63	18.19	-8.02	16.67	17.22
	**	**	**	**	**
<b>FOG</b> Differences between means t Significance two-tailed.		1.84	1.75	4.08	3.51
		17.22	7.49	9.96	12.63
		**	**	**	**
<b>Flesch-Kincaid</b> Differences between means t Significance two-tailed.			.09	5.92	.83
			-.40	15.48	3.90
				**	**
<b>Spache</b> Differences between means t Significance two-tailed.				5.83	.92
				30.59	3.90
				**	**
<b>Dale-Chall</b> Differences between means t Significance two-tailed.					6.75
					21.98
					**

$df = 62$ , \*\*  $p < .01$ , \*  $p < .05$