REASSESSING THE ACCURACY AND USE OF READABILITY FORMULAE

[1]Dahlia Janan & [2]David Wray
[1]Faculty of Languages and Communication
Universiti Pendidikan Sultan Idris
[2]Institute of Education
University of Warwick

[1]Corresponding author: *dahlia@fbk.upsi.edu.my*

ABSTRACT

Purpose – The purpose of the study is to review readability formulae and offer a critique, based on a comparison of the grading of a variety of texts given by six well-known formulae.

Methodology – A total of 64 texts in English were selected either by or for native English speaking children aged between six and 11 years. Each text was assessed using six commonly used readability formulae via the Words Count website (http://www.wordscount. info/) which provides automated readability indices using FOG, Spache, SMOG, Flesh-Kincaid and Dale-Chall. For the ATOS formula, the Renaissance Learning website was used (http://www. renlearn.com/ar/overview/atos/). Statistical tests were then carried out to check the consistency among the six formulae in terms of their predictions of levels of text difficulty.

Findings – The analysis demonstrated significantly different readability indices for the same text using different formulae. 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. This finding suggests that readability formulae need to be used carefully to support teachers' judgements about text difficulty rather than as the sole mechanism for text assessment.

Significance – Making decisions about matching texts to learners is something regularly required from teachers at all levels. Making

such decisions about text suitability is described as measuring the 'readability' of texts, and for a long time, this measurement has been treated as unproblematic and achieved using formulae which use such features as vocabulary difficulty and sentence length. This study suggests that the use of such readability formulae is more problematic than may at first appear. Although the study was carried out with native English speaking children using texts in English, it is argued that the lessons learnt apply equally to Malay speakers reading Malay language texts.

Keywords: Readability formula, reading material, measurement.

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 & Conard, 1991). Learners who are given reading materials that are too difficult can fail to make progress (Gambrell, Wilson & Gantt, 1981), are frequently off task and may exhibit behaviour problems (Anderson, Wilkinson & Mason, 1987), or simply give up (Kletzien, 1991). Making the match is therefore a crucial skill for teachers, indeed for anyone who produces written material they desire to be read and understood by others. The successful exercise of this skill requires knowledge of readability, a concept which gave rise to a significant body of research from the 1920s to the early 1990s. One of the major outcomes was the production of a large number of "readability formulae", that is, approaches to analysing texts 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 21st century, new readability formulae have continued to be produced, reflecting the attractiveness of such an approach in matching texts to learners.

Definitions of Readability

Various definitions of the concept of readability have emphasised the elements in a text associated with comprehension on the part of

the reader. Parts of the concept also referred to a person's ability to read a 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 are not separate, but interact with each other. Thus, definitions of readability have never been entirely text-centred. 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, instead, usually involved objective estimates of the difficulty level of reading material derived from a study of the text alone.

The Development of Readability Formulae

There was a great deal of development in readability research between the 1920s and the early 1990s, caused by an urge to emphasise quantification in developing a scientifically based curriculum. From the middle of the 1990s, however, developments in this research area decreased significantly. In the JSTOR archive of journal articles, for example, 1,298 articles referenced by the key word "readability" were logged between 1965 and 1980, and a further 1,590 between 1980 and 1995. From 1995 to 2010, however, only 672 new articles have appeared.

This decrease in research has undoubtedly been related to criticisms of the use of readability formulae. Research has suggested that they are not reliable or valid predictors of text difficulty (e.g., Redish & Selzer, 1985; Bruce, Rubin & 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. Readability instead tended to focus on an objective estimation of text difficulty without involving the readers of that text. Many of the assumptions about readability, and arguments as to its weakness as a concept, are associated with readability formulae, because these formulae are the best known products of this field of research.

The earliest readability formulae were produced between 1921 and 1934, including examples 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 set of formulae was produced in the years between 1934 and 1952, by, for example, 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 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, with new formulae, such as McLaughlin's (1969) SMOG. More recently, a number of computerised formulae have been developed, such as the Lexile Framework (Lennon & Burdick, 2004) and ATOS (Milone, 2008). The Lexile Framework formula uses variables, such as average sentence length and word frequency (Stenner, A. J., Burdick, H., Sanford, E. & Burdick, D., 2006), while ATOS 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. They are still popular and are now much easier to use with most being available on the internet, and even Microsoft Word providing inbuilt readability measures. Recent research has employed these formulae for a variety of purposes (e.g., Burke & Greenberg, 2010; Gallagher, Fazio & Gunning, 2012). However, they have continued to receive heavy criticism over the years.

Criticisms of Readability Formulae

Bruce, Rubin and Starr (1981) pointed out that formulae did not take into account current knowledge about the reading process. They included sentence length and word difficulty but did not 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 and required background knowledge. Also, they attempted to measure text difficulty in isolation from other elements, such as the context of a text's use and the reader's motivation and interest (Bruce et al. 1981).

It has also been suggested (Davison & Kantor, 1982) that changes made to a text, on the basis of readability formulae, to make it easier to read (such as splitting complex sentences into component clauses and changing vocabulary items), may actually make it harder to understand. As a result of such critique, 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).

Given such a level of criticism, it might be expected that the use of readability formulae would have diminished 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 & O'Connor, 2011; Freda, 2005). Badarudeen and Sabharwal (2010), for example, reported on their use of a variety of formulae to judge the readability of medical patient education materials. Their critique on the use of formulae was 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 & Sabharwal, 2010: 2574). 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.

METHODOLOGY

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

The aim of this element of the study is 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 are:

- How consistent are the formulae in distinguishing difficult texts from easier texts?
- How consistent are the formulae in suggesting a readability level for individual texts?

Participants and Texts

The study involved 32 randomly selected UK native English-speaking school children aged from six to 11 years. These children represented a range of abilities in terms of reading, although there were no non-readers amongst them. For the purposes of the wider study, the reading proficiency of these sample children was important, but for the particular element of the study discussed here, it was not a significant factor as it was the texts which were the main focus, not the readers. The children were requested to bring along to a meeting with the researcher any book, or other reading material, that they had 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; it was also due to 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 was 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. Comparability between readability formulae should not, therefore, be taken for granted. In this study, our aim was to compare the outcomes of the formulae on a set body of texts.

Procedures

The 64 extracts were analysed using the six readability formulae via the Words Count website (http://www.wordscount.info/), which provided automated readability indices using FOG, Spache, SMOG, Flesh-Kincaid and Dale-Chall. For ATOS, the Renaissance Learning website was used (http://www.renlearn.com/ar/overview/atos/). All six formulae were originally produced for use with texts aimed at native English-speaking children and their output scores were all expressed as 'grade levels': that is, US grade school levels (The accepted method to derive chronological ages from these figures is simply to add six to each. This grade five indicates 11 year old children). The six readability scores for each of the 64 extracts was then listed and entered into SPSS software for later analysis.

Analysis

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 difficulty levels. The Spearman's rank order correlation coefficient was used in this procedure.
- *Comparison of the grade levels*. The aim was to demonstrate the extent to which formulae was in concurrence in predicting the grade levels of the 64 texts. Paired-sample T-Tests were used for this purpose.

RESULTS AND DISCUSSION

Consistency Estimation of the Formulae

Table 1 presents the results of the comparison between the order of difficulty of the 64 texts produced by each formula.

Table 1

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
SMOG	.98**	.93**	.83**	-.41**	.70**
FOG		.95**	.84**	-.47**	.74**
Flesch-Kincaid			.88**	-.32*	.68**
Spache				-.14	.68**
Dale-Chall					-.49**

Number of texts N= 64
*** p< .01 * p< .05*

Table 1 shows that a *very high* statistically significant correlation is found between 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 is 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 common predictor variable. 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 is 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 is the use of the number of polysyllabic words in a text.

The ATOS formula did have a moderately 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. 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 is 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' from that used by the Spache formula.

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

Grade level predictions

We also calculated the average grade levels predicted by each of the six formulae. Table 2 shows this analysis.

Table 2

The Mean Text Grade Levels Predicted by the Six Readability Formulae

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

The data showed 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 were carried out to identify whether there are any statistically significant differences between these mean grade levels. Table 3 shows the results.

Table 3

Paired-Sample T-Tests of the Differences between the Mean Grade Levels of the Six Formulae

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

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

Table 3 shows that the *highest* difference between the text mean grade levels is 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 shows *no* statistically significant difference between the mean grade levels is that between the Flesch-Kincaid and FOG formulae, where the mean difference is .09 (t=-40, df=62, p=.69). It can therefore be concluded that only the Flesch-Kincaid and the FOG formulae produce similar results for the grade levels of these 64 texts.

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 ranking the texts in order of difficulty, some widely differing grade level scores were 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 being at a higher level than the other formulae, but it was also prone to assigning a text as easy, whereas the rest of the formulae predicted it as difficult.

A definition of readability as the "*ease with which a reader can read and understand a given text*" (Oakland & 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). Measurements of text features, through the application of readability formulae, have been popular, but heavily criticised in terms of validity and reliability. The formulae, it has been argued, fail to measure comprehension (Duffy, 1985) and 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).

The findings of the present study support these criticisms. 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 appears that some of the formulae (but not all) are consistent in their ranking of texts in order of difficulty but are not consistent in their grading of each text, with up to a six year discrepancy among them. 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 on both aspects? 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 wide 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 & Unrau, 2004). Meaning no longer comes from the text, but from readers who bring their social and cultural backgrounds into 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. Research on reading more recently has tended to focus on what happens in readers' minds during reading, and has employed

error and miscue analysis (Goodman & Goodman, 1977) and think aloud protocols (Pressley & 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 & Picard, 2006; Cotton & 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, D. (2011). For our purposes here, 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. They suggest, however, that, while not offering a definitive picture of the reading difficulty of a particular text, formulae 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 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.

Application to the Malaysian context

The research discussed in this article was carried out with native English-speaking children, using English language texts. What is its relevance, therefore, to the Malaysian context?

It should, firstly, be recognised that the selection of English language texts for non-native English speakers, is itself a common occurrence in Malaysia, although many studies of English reading (e.g., Yaacob & Pinter, 2005) do not mention the criteria by which texts were chosen for the Malay speaking students they researched. Abdullah and Hasim (2007) reported one study of the readability levels of English language texts used with Malay language speakers and the measures of readability used in their work were derived from the familiar readability formulae as critiqued in our present study. A more extensive understanding of readability in English language texts seems to be a worthy aim for teachers in the Malaysian context as in other international contexts.

The research literature on the study of readability of Malay language texts is quite limited. The most influential research in the area is from a 1982 doctoral study (Md Yunus, 1982) which has been used by a few researchers since that date (e.g., Arifin, Halim & Bakar, et al., 2013; Abdullah, 2013) because its major outcome was a Malay language readability formula, the only one of its kind as far as we can tell. This formula is based on the analysis of 300 word extracts from texts, on which the following calculations were applied:

Y (Readability index) = -13.988 + 0.3793 × (words per sentence) + 0.0207 × (no. of syllables)

The Y index relates to an educational level; thus a readability score of 1 should indicate that the text is suitable for reading by Primary 1 students.

Our current study has questioned the usefulness of such formulae in assessing readability in English Language texts, and the same critique should apply to readability measurement in Malay language texts, especially in the light of the observation of Lee and Low (2014) that Malay and English orthographies work in similar ways, both being based on alphabetic scripts.

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 a reading class. Their prime 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 weigh in 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

Abdullah, I. H., & Hasim, R. S. (2007). Readability analysis of Malaysian short stories in English. *E-Bangi Jurnal*, *2*. Retrieved from http://pkukmweb.ukm.my

Abdullah, K. (2013). A study on the use of electronic stories in Malay language teaching at the lower primary level. *Journal of Arts and Education, 1*, 12–26. Retrieved from http://www.accessinterjournals.org

Anderson, R. C., Hiebert, E. H., Scott, J. & Wilkinson, L. (1985). *Becoming a nation of readers*. Washington, DC: National Institute of Education.

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.

Arifin, Z., Halim, Z., & Bakar, K. (2013). The Readability of Balaghah Textbook of Malaysian Higher Certificate of Religion (Stam): A pilot study. *Advances in Natural and Applied Sciences*, *7*, 192–199. Retrieved from http://www.aensiweb.com

Badarudeen, S., & Sabharwal, S. (2010). Assessing readability of patient education materials: Current role in orthopaedics. *Clinical Orthopaedics and Related Research, 468*, 2572–80.

Bruce, B., Rubin, A. & Starr, K. (1981). Why Readability Formulas Fail. *IEEE Transactions on Professional Communication, 24*, 50–52.

Burke, V., & Greenberg, D. (2010) Determining readability: How to select and apply easy-to-use readability formulas to assess the difficulty of adult literacy materials. *Adult Basic Education and Literacy Journal*, *4*, 34–42.

Chall, J. S., & Conard, S. (1991). *Should textbooks challenge students? The case for easier or harder books*. New York: Teachers College Press.

Chall, J. S., & Dale, E. (1995). *Readability revisited: the new Dale-Chall readability formula*. Cambridge, Mass: Brookline Books.

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology*, *37*, 45–54.

Cronin, M., O'Hanlon, S., & M. O'Connor. (2011). Readability level of patient information leaflets for older people. *Irish Journal of Medical Science*, *180*, 139–142.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11–28.

Dale, E., & Tyler, R. (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* (pp. 80–107). Costa Mesa, California: Impact Information.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly, 17*, 187–209.

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.

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*, 151–156.

Fry, E. (1977). *Elementary reading instruction*. New York: McGraw-Hill.

Gallagher, T. L., Fazio, X., & Gunning, T. (2012). Varying readability of science-based text in elementary readers: Challenges for teachers. *Reading Improvement, 49*, 93–112.

Gambrell, L. B., Wilson, R. M. & Gantt, W. (1981). Classroom observations of task-attending behaviors of good and poor readers. *Journal of Educational Research*, *74*, 400–404.

Goodman, K. S., & Goodman, Y. (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.

Harris, T. L., & Hodges, R. (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.

Janan, D. (2011). *Towards a new model of readability* (Unpublished doctoral dissertation). University of Warwick, Coventry.

Kincaid, Fishburne, J., R, Rogers, R., & Chissom, B. (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.

Lee, L. W., & Low, H. M. (2014). The development and application of an online Malay language corpus-based lexical database. *Kajian Malaysia*, *32* (Supp. 1), 151–166.

Lennon, C., & Burdick, H. (2004). *The Lexile Framework as an approach for reading measurement and success*. Durham, NC: Metametrics.

McClusky, H. Y. (1934). A quantitative analysis of the difficulty of reading materials. *Journal of Educational Research, 28*, 276–282.

McKenna, M. C., & 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.

Md. Yunus, K. R. (1982). An assessment of structural variables in Malay: A readability formula. *Dissertations from ProQuest.* Paper 1263. Retrieved from http://scholarly repository.miami.edu

Michelson, J. (1985). IRA, NCTE take stand on readability formulae. *Reading Today*, 2(3), 1.

Milone, M. (2008). *The development of ATOS: The Renaissance readability formula*. Wisconsin Rapids, WI: Renaissance Learning.

Oakland, T., & Lane, H. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing, 4*, 239–252.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: the nature of constructively responsive reading*. Hillsdale, N.J.: Lawrence Erlbaum.

Redish, J., & Selzer, J. (1985). The place of readability formulas in technical communication. *Journal of Reading Behavior, 32,*46–52.

Renaissance Institute. (2000). *The ATOS readability formula for books and how it compares to other formulas*. Madison, WI: Renaissance Institute.

Ruddell, R. B., & Unrau, N. (2004). Reading as a meaning-construction process: The reader, the text, and the teacher. In R.B., Ruddell, N. Unrau, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 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.

Stenner, A. J., Burdick, H., Sanford, E. & Burdick, D. (2006). How accurate are Lexile text measures. *Journal of Applied Measurement, 7*, 307–322.

Thorndike, E. L. (1921). Word knowledge in the elementary school. In W. H. Dubay, (Ed.), *Unlocking language: the classic readability studies* (pp. 9–43). Costa Mesa, California: Impact Information.

Vogel, M., & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. In W. H. Dubay, (Ed.), *Unlocking language: the classic readability studies* (pp. 58–66). Costa Mesa, California: Impact Information.

Yaacob, A., & Pinter, A. (2008). Exploring the effectiveness of using big books in teaching primary English in Malaysian classroom. *Malaysian Journal of Learning and Instruction*, 5, 1–20.