An analysis of readability metrics on English exam texts

Sibeko Johannes
Nelson Mandela University
Johannes.Sibeko@mandela.ac.za

Menno van Zaanen
South African Centre for Digital Language Resources
Menno.vanZaanen@nwu.ac.za

Abstract

Readability metrics provide information on how difficult a text is to read. This information is relevant, for instance, to identify suitable texts for learner readers. Readability metrics have been developed for several languages, but no such metrics have been developed for the indigenous South African languages. One of the limitations in the development of the metrics is the availability of texts in these languages for which the readability is known. To resolve this issue, we would like to consider texts that are used in final year exams of language subjects at highschool. We expect these texts to have consistent readability throughout the years. Additionally, in South Africa, language subjects may be taught both as home language or first additional language. We expect there to be differences in readability between the exam texts for these subjects. To test these assumptions, in this article, we compute readability scores using nine existing readability metrics for the final year exams of English home language and English first additional language. The results show that indeed the readability of the texts is consistent over the years and significantly different between the two subjects. Generalizing over these results, we expect that we can use final year exam texts of other languages to develop readability metrics for the indigenous South African languages in future work. An analysis of the performance of the readability metrics on the English texts serves as a starting point to identify useful text properties to use for the development of the readability metrics for the indigenous South African languages.

Keywords: English, readability metrics, text readability, highschool exam texts

1 Introduction

The research presented in this article forms part of a bigger project that aims to develop readability metrics for indigenous South African languages. To develop these metrics, we consider using educational texts, such as reading comprehension and summary writing texts used in final year exams, as these are expected to have known or at least consistent readability. Currently, however, it is still unclear whether these exam texts indeed have consistent readability and as such are suitable for the development of readability metrics. The explorative research described here investigates readability of English comprehension and summary exam texts used in South Africa.

We focus on English since tried and tested metrics for measuring text readability in English exist. If the results for the English exam texts are as we expect, then we assume that we can use the same type of exam texts for the indigenous South African languages. The analysis of text readability using English readability metrics may also provide information on what text properties to investigate further when developing text readability metrics for the indigenous South African languages.

In the case of South African official languages, as far as could be ascertained, only Afrikaans has readability metrics. The four Afrikaans readability metrics are based on the English readability metrics (*see* Jansen, Richards and Van Zyl 2017). Fashioned after Afrikaans, we learn from the already established scholarship of text readability in English.

South African schools offer English on three levels (DBE 2012). The English home language (HL) subject is aimed at learners who start school with English competency skills such as listening, speaking, reading, and writing (DBE 2011b). The English as the first additional language (FAL) is proposed for learners who start school with some exposure to English (DBE 2011a; 2012; 2016), whereas English as the second additional

language (SAL) is intended for learners who start school with no exposure or competency skills in English (DBE 2011c). The content, teaching schedule, and the overall curricula for these English subjects are governed by the Curriculum and Assessment Policy Statements (CAPS). Curricula for other official languages are translated from the generic English CAPS (Van der Walt 2010; De Vos, Van der Merwe and Van der Mescht 2014; Tshesane 2014; Probert and De Vos 2016; Van der Walt 2018). In this article, SAL examinations are excluded for two reasons. First, SAL examinations are set at provincial level and there is no certainty that the same rigorous processes followed at National level are followed. Second, SAL examination papers are inconsistently uploaded on the DBE website and many examination papers could not be located.

Given that the FAL subject is aimed at learners with lower proficiency than those in the HL subject (DBE 2011a, p.8; 2011b, p.8), we expect that the texts used for reading comprehension in the FAL examination will be easier to read than texts used in the HL examination. Grade 12 teachers preparing learners for the final examinations are encouraged to source different texts and adapt them to their learners' levels (DBE 2017, p.5). Unfortunately, the guidelines do not specify the text characteristics that teachers can adapt, so the selected texts by the teachers cannot be used reliably in this research. Additionally, examination guidelines do not include any information on whether readability metrics are used to prepare examination papers.

In order to understand the readability of texts used in the English highschool subjects, this article sets out (i) to check whether the readability of the English reading comprehension and summary writing exam texts is consistent (that is, whether there are no differences between the readability of the texts of the different examination opportunities and whether there are differences between the HL and FAL exam texts), and (ii) to investigate whether different readability metrics are consistent with these results in order to get an idea of what text properties (used in the metrics) might be useful for the development of similar metrics for other languages.

2 Background

Measuring text readability can be approached from different perspectives. One perspective depends on readers' characteristics (Nouwens, Groen and Verhoeven 2016; Duff 2019, p.562-3; Kärbla, Uibu and Männamaa 2019; 2020; Phillips Galloway et al. 2020, p.4). From this perspective, the readability of a text depends on how well a reader can either understand the literal meaning of the text, infer meaning from the text, or use evaluative techniques to comprehend the text (Basaraba et al. 2013; Tennent 2014; Kärbla, Uibu and Männamaa 2020). As such, since text readability is viewed in relation to the specific reader, it is used interchangeably with text difficulty and reading difficulty (*see* Collins-Thompson (2014)).

Another perspective, which relates to the readability metrics used in this study, does not view text readability in relation to the reader. Instead, text readability is viewed as a subcategory of text complexity (Amendum, Conradi and Hiebert 2018, p.122), which focuses on independent linguistic factors that can be manipulated (Mesmer, Cunningham and Hiebert 2012, p.235) as opposed to how the text interacts with the reader (McNamara, Louwerse and Graesser 2002; Meyer 2003; Stahl 2003; Stenner et al. 2006; Benjamin 2012; Spencer et al. 2019). Readability metrics are described as mathematical formulas obtained through regression analysis (Mc Laughlin 1969, p.640) that are used to measure readability (Heydari 2012, p.423; Begeny and Greene 2014, p.198). They focus on the style of writing (Courtis 1987, p.20) as manifested, among others, through word and sentence lengths (Stevens, Stevens and Stevens 1992), syllable counts (Kate et al. 2010, p.547), and wordlists (Vajjala and Meurers 2014, p.3). Readability formulas generally output estimated grades or levels of education appropriate for each text, but other numeric values may also be computed.

Table 1: Extracts and summary information from the 2016 November HL and FAL examination texts.

	HL	FAL
Extract	'Hand gestures are really a powerful aspect of communication, from both the speaker's and the listener's end,' says Dr Carol Kinsey Goman, body language expert. Last year, a study analysing human gestures found that the most popular, prolific speakers used an average of 465 hand gestures, which is nearly twice as many as the least popular speakers used. Other research has found that people who 'talk' with their hands tend to be viewed as warm, agreeable and energetic, while those who are less animated are seen as logical, cold and analytical.	South Africa ranks as one of the top thirty driest countries in the world. This knowledge should encourage a new approach towards the way we use our fragile water resources. As South Africans, we have had to change our behaviour to adapt to electricity cuts, so the water crisis demands a change in our habits relating to water usage. South Africa loses billions of Rands annually through leaking taps and water pipes. It is important to repair or replace damaged water connections and washers to stop all leaks.
Sentences	3	5
Tokens	91	89
Syllables	153	142

In the South African context, studies on text readability of health documents (Joubert and Githinji 2014; Leopeng 2019; De Wet 2021) and textbooks evaluations (Sibanda 2013; Wissing, Blignaut and Van Den Berg 2016) using classical readability metrics have been conducted. However, according to our knowledge, there are no empirical studies investigating the readability of reading comprehension and summary writing texts in the domain of South African basic education.

3 Methodology

3.1 Material

The South African Department of Basic Education (DBE) affords grade 12, which is the final grade in the South African basic education schooling system, candidates two examination opportunities. The end-of-year grade 12 examinations are written in November of each year. Until 2018, supplemental examinations were written in February/March. Since 2019, the supplemental examinations are written in May/June. From 2016, low performing learners who could not cope with the grade 12 curriculum were allowed to write three subjects at the end of the academic year (November session) and complete the remaining three subjects in May/June of the following year. The Multiple Examination Opportunities Policy was discontinued after the May/June 2019 examinations (DBE 2019).

Although some exam material is not available on DBE's website for public access, most of the texts used in the exams can be found there.

Our data set comprises 48 exam texts composed of twelve HL and twelve FAL November texts from 2008 to 2019, eight HL and eight FAL February/March texts from 2011 to 2018, and four HL and FAL May/June texts from 2016 to 2019. The exam texts were manually extracted from the PDF documents, which were downloaded from DBE's website. Headings were manually punctuated to ensure the correct identification of sentence boundaries. Footnotes, endnotes, and source references were manually removed from the text. Table 1 provides example extracts from the 2016 November exam texts for the HL and the FAL examinations including some of the textual properties that are used in readability metrics.

Table 2: Classical readability formulas used in the study.

Formula	Calculation
Kincaid	$= 0.39\left(\frac{\#tokens}{\#sentences}\right) + 11.8\left(\frac{\#syllables}{\#tokens}\right) - 15.59$
Flesch	$= 206.835 - 1.015\left(\frac{\#tokens}{\#sentences}\right) + 84.6\left(\frac{\#syllables}{\#tokens}\right)$
SMOG	$= 3.1291 + 1.043\sqrt{\#polysyllabic\ words\ x\ \frac{30}{\#sentences}}$
Fog	$= 0.4[\left(\frac{\#tokens}{\#sentences}\right) + 100\left(\frac{\#complex\ words}{\#words}\right)]$
Coleman-Liau	$= 0.0588\left(\frac{\#letters}{\#samples}\right) - 0.296\left(\frac{\#sentences}{\#samples}\right) - 15.8$
ARI	$= 4.7\left(\frac{\#letters}{\#words}\right) + 0.5\left(\frac{\#words}{\#sentences}\right) - 21.43$
LIX	$= \left(\frac{\#long\ words}{\#words} x\ 100\right) + \left(\frac{\#words}{\#sentences}\right)$
RIX	$= \frac{\#long\ words}{\#sentences}$
Dale-Chall	$= 0.0496\left(\frac{\#words}{\#sentences}\right) + 11.8\left(\frac{\#difficult\ words}{\#words}\right) x\ 0.1579 + 3.6365$

3.2 Procedure

To evaluate the readability of the different texts, we compute the readability according to nine well-known readability metrics, namely, Flesch-Kincaid Grade Level (Kincaid) (Kincaid et al. 1975), Flesch Reading Ease (Flesch) (Flesch 1948), Simple Measure of Gobbledygook (SMOG) (Mc Laughlin 1969), Gunning Fog index (Fog) (Gunning 1952; 1969), lisbarhetindex (LIX) and Rate index (RIX) (Anderson 1983), Automated Readability index (ARI) (Senter and Smith 1967; Kincaid and Delionbach 1973), Coleman-Liau index (Coleman and Liau 1975), and the Dale-Chall index (Dale and Chall 1948). The formulas used in each of the metrics are presented in Table 2.

We have used the Python readability package (version 0.3.1) to compute these. All of these metrics have been developed specifically for English texts. Note that for all metrics, lower scores imply easier to read texts, except for the Flesch metric which shows higher scores for easier to read texts.

'Polysyllabic words' as used in SMOG, and 'complex words' as used in Fog, refer to words with more than two syllables (Eltorai et al. 2015, p. 831; Harden 2018, p. 37). Fog does not count proper nouns and three-syllable words formed by adding suffixes such as -es and -ed.

In SMOG, one uses three samples of ten sentences each, one from the beginning of the text, one from the middle and one from the end of the text (Mc Laughlin 1969, p. 639; Zhou, Jeong and Green 2017, p. 100). The summed results from the samples are then used in the formula. The Coleman-Liau formula divides the text into shorter pieces of 100 words each. The 100-word pieces of text are each analysed individually and the averages are used in the calculations. The LIX and the RIX formulas use 'long words' to signify words with more than six characters. It is suggested that for calculation of both LIX and RIX, ten samples of ten sentences be used for the analysis (Anderson 1983, p. 495). As the exam texts are below 100 sentences each, no sampling was necessary.

Table 3: Mean scores for text properties used in the metrics.

	February		May		November		Overall	
	FAL	HL	FAL	HL	FAL	HL	FAL	EngHL
Tokens	915.25	1173.75	891.00	1157.00	962.42	1134.67	933.70	1151.42
Syllables	1347.25	1726.50	1315.00	1774.75	1373.42	1727.00	1353.87	1734.79
Syllable/ word	1.47	1.47	1.48	1.54	1.43	1.52	1.45	1.51
Sentences	55.63	64.38	54.00	63.75	59.17	59.08	57.13	61.63
Words/ sentence	16.71	18.27	16.53	18.30	16.45	19.48	16.54	18.88
Letters	4400.25	5562.63	4283.50	5648.75	4534.75	5539.50	4443.48	5565.42
Letters/ word	4.81	4.74	4.81	4.89	4.72	4.87	4.77	4.83
Long words	221.13	278.50	204.50	308.25	220.75	296.92	218.13	292.67
Complex words	121.63	166.88	113.75	193.25	108.83	179.58	113.74	177.63
Difficult words	281.75	375.00	269.50	395.25	294.00	388.33	285.09	385.04

Dale-Chall uses 'difficult words' to signify words that do not appear in the wordlist of 3000 frequently used words. Commonly used words are identified as words in the list together with plurals of basic words in the list, -s, -ed, -ing, and -ied verbs, -ly adverbs, names of people and organisations with organisation names only being counted two times per 100-word sample, abbreviations, and compound words if both words appear on the list (Barry and Stevenson 1975, p. 219). For our data sets, we used the basic setting of the readability package which samples four evenly spaced 100-word samples for each text. This type of sampling is recommended (Dale and Chall 1948, p. 37).

3.3 Analysis

To analyze the performance of the different readability metrics on the HL and FAL exam texts, we will first provide mean and standard deviation values of each of the metrics. Additionally, we investigate correlations between the results of the different metrics. As the different metrics aim to describe the same property of the text, we expect there to be relatively high significant correlations.

Once the descriptive statistics are provided and discussed, we create linear regression models for each of the readability metrics. This indicates the relationship between the readability of the HL and FAL texts, including the years and months of the exams. We expect these analyses to identify significant differences between the HL and FAL texts and we expect no significant differences based on the years and months.

4 Results

4.1 Descriptive statistics

To get a better understanding of the behavior of the metrics on the texts, first, an overview of the textual properties used in the metrics discussed in this article is presented in Table 3. Second, we provide mean and standard deviations for the different metrics for both the HL and FAL texts in Table 4.

Table 4: Means and standard deviations (within brackets) of the different readability metrics for the English HL and FAL exam texts.

Metric	HL	FAL
Kincaid	9.53 (1.49)	7.99 (1.10)
Flesch	60.18 (8.60)	67.18 (6.20)
SMOG	12.26 (1.12)	10.74 (0.94)
Fog	13.68 (1.57)	11.51 (1.25)
ARI	10.76 (1.59)	9.28 (1.35)
Coleman-Liau	11.06 (1.30)	10.42 (1.17)
LIX	44.22 (4.30)	39.97 (3.61)
RIX	4.80 (0.92)	3.88 (0.67)
Dale-Chall	9.85 (0.48)	9.30 (0.60)

Table 5: Correlation between the different readability formulas. All correlations are significant (p<.0001).

Metrics	Kincaid	Flesch	SMOG	Fog	ARI	Coleman-Liau	LIX	RIX	Dale-Chall
Kincaid	1.00	-.95	.94	.96	.96	.85	.95	.96	.78
Flesch	-.95	1.00	-.93	-.92	-.87	-.93	-.91	-.88	-.78
SMOG	.94	-.93	1.00	.99	.87	.81	.89	.89	.73
Fog	.96	-.92	.99	1.00	.90	.79	.91	.92	.74
ARI	.96	-.87	.87	.90	1.00	.85	.96	.97	.81
Coleman-Liau	.85	-.93	.81	.79	.85	1.00	.89	.84	.84
LIX	.95	-.91	.89	.91	.9	.89	1.00	.99	.84
RIX	.96	-.88	.89	.92	.97	.84	.99	1.00	.81
Dale-Chall	.78	-.78	.73	.74	.81	.84	.84	.81	1.00

In Table 4, we see consistent differences in the scores where FAL texts are considered to be easier than the HL texts. These results do not consider the influence of the different months or years. This will be investigated in more detail with the linear regression models. The results of Pearson's correlations between the results of the different formulas are presented in Table 5.

All of these correlations are significant (p<.0001). We see that most pairs of metrics show strong positive correlations, except for the Flesch metric, which (in contrast to the other metrics) shows strong negative correlations as higher values mean easier-to-read texts. The lowest absolute correlations are found for Dale-Chall and SMOG (r=.73), and Dale-Chall and Fog (r=.74) metrics (although these are still considered strong correlations). Overall, these results show that the metrics provide very similar behavior.

We also present density plots (see Figure 1) for the different metrics for both HL and FAL texts. This shows that the readability scores are generally normally distributed.

4.2 Linear regression analyses

To investigate the influence of the subject (HL and FAL), and year and month of the exam on the readability, we created linear regression models for each of the readability metrics. For this, we use Subject (HL vs FAL), Year, and Month as independent variables (we also consider the possibility of interaction between the last two variables in the model) and the readability score as the dependent variable.

These results do not consider the influence of the different months or years.

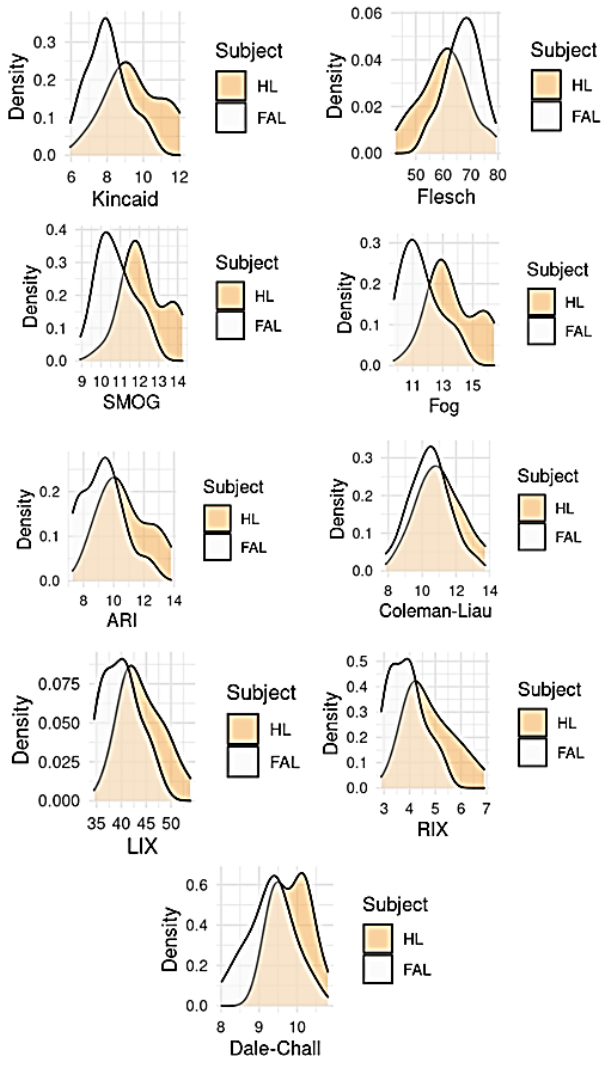

Figure 1: Data distributions density plots for all of the metrics, separated by the HL and FAL texts.

Table 6: Linear regression results. The difference indicates the estimated difference between the HL and FAL values. Error indicates the standard error. F(1, 23) provides the results of the F test with the corresponding p values.

	Difference	error	F(1, 23)	p
Kincaid	-1.54	0.38	16.8471	.0004
Flesch	7.01	2.08	11.3126	.0027
SMOG	-1.53	0.32	23.2077	<.0001
Fog	-2.16	0.44	24.5655	<.0001
ARI	-1.48	0.42	12.4651	.0018
Coleman-Liau	-0.65	0.31	4.3258	.0489
LIX	-4.25	1.13	14.0233	.0011
RIX	-0.92	0.23	15.3211	.0007
Dale-Chall	-0.55	0.12	21.5054	.0001

This will be investigated in more detail with the linear regression models. Although the results are consistent, the grade levels as used in these metrices are not purposed for the South African schooling system. As such, the actual grade levels cannot be determined using these metrics at this point. Both Year and Month are sum-coded, allowing investigation of the influence of these variables with respect to the mean values. Note that although the readability scores are strongly correlated, they are not combined in any model.

For all metrics, the residuals conform fairly well to the normality assumption (according to their histogram and Q-Q plots), although the Kincaid, Coleman-Liau, and Fog metrics show slight deviations from the normal distribution. Similarly, the homoscedasticity for all models is also good (according to the residual plots). The information of all the linear regression models can be found in Table 6.

As you can see, all the linear regression models indicate significant differences between the readability values for both subjects (HL vs FAL) for each of the metrics. Note that the models (according to ANOVA analyses) did not show any significant differences resulting from the year and month variables ($p>.05$), except the model for Coleman-Liau, which showed a significant influence of Year ($F(11, 23) = 2.3303$; $p=.04$) and the model for Dale-Chall, which showed a significant interaction between Year and Month ($F(10, 23) = 3.2075$; $p=.01$). Note that Dale-Chall already showed the largest deviations in correlations with other metrics, which may be due to the significant influence of the interaction between the Year and Month variables.

5 Discussion

Our results indicate that grade 12 examiners for the HL and FAL subjects have selected texts that are consistent over time (no significant differences between year and month) and different for each level (significant differences between HL and FAL). This corresponds to the viewpoint that the HL subject is more complex and caters for learners with higher language competency than those in the FAL subject, as it is supported by the text readability of English exam texts.

From this result, we hope that this will also be the case for the exam texts of the indigenous languages. As mentioned in the background section, the indigenous languages' curricula are often translated from the English curricula. Because of this, we hope that the selections of texts in the indigenous languages also mirror the English practices in as far as readability is concerned.

The nine metrics investigated in this article all show similar behaviour. First, the correlation results indicate no significant differences between any of the metrics. All metrics correlate strongly when considering the readability of the exam texts. Second, the metrics provide similar linear regression models with only minor differences.

There are two unexpected findings. First, Coleman-Liau shows a strong positive correlation with other metrics in the study, but the linear regression model indicates that there are significant differences between the different years. This may be because Coleman-Liau splices texts into pieces of 100 words each and then uses the averages to calculate the overall outcome.

Second, Dale-Chall linear regression model also shows statistically significant differences in terms of Years and Months. Unfortunately, this explorative research does not explore these peculiarities in detail. Nonetheless, if one is to use a metric fashioned after the Dale-Chall index,

a list of frequently used words would need to be generated in the language of choice. At this point, such lists do not yet exist in the indigenous South African official languages. An exception is McKellar's list of frequently used Afrikaans words compiled for the Afrikaans text readability metric (Jansen, Richards & Van Zyl 2017, p.154).

6 Conclusion

In this article, we explored the readability of the English HL and FAL highschool exam texts used in South Africa. We used nine classical readability metrics to investigate the readability of each text. We showed that the nine metrics are significantly and positively correlated to each other. Linear regression models showed that there are consistent significant differences between the HL and the FAL texts. Moreover, the models did not identify significant differences of the HL or FAL texts used at the different examinations.

One of the aims of the exploratory research described in this article was to get a sense of whether exam texts can be used in the development of text readability metrics for South African indigenous languages. Given that the indigenous language curricula are translated from the English generic curricula we may expect similar readability characteristics for the indigenous languages exam texts. That is, the HL and FAL texts for the indigenous South African languages can be expected to be significantly different while texts used at different examinations are expected to show no significant differences in readability.

The results found in this article also indicate areas for future work. For instance, one could explore reasons for the lower correlation between the Dale-Chall index and the SMOG and Fog indexes. Additionally, lists of frequently used words could be compiled for each indigenous language to explore how corpus-based metrics, such as the Dale-Chall index, affect text readability outcomes in the official indigenous South African languages. Furthermore, the syllable-based metrics show the need for the development of computational linguistic tools for the indigenous languages, such as syllabifiers to automatically identify syllables.

Acknowledgements

This research was funded in part by the National Research Foundation (NRF) South Africa (Grant number: 128875). It forms part of a larger study by the first author supervised by the second author. The Grant holder acknowledges that opinions, findings, and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the authors, and that the sponsors accept no liability whatsoever in this regard.

References

Amendum, SJ, Conradi, K & Hiebert, E 2018, 'Does text complexity matter in the elementary grades? A research synthesis of text difficulty and elementary students' reading fluency and comprehension', *Educational Psychology Review,* vol. 30, no. 1, pp. 121-151.

Anderson, J 1983, 'Lix and rix: Variations on a little-known readability index', *Journal of Reading,* vol. 26, no. 6, pp 490-496.

Barry, JG & Stevenson, TE 1975, 'Using a computer to calculate the Dale-Chall Formula', *Journal of Reading*, Vol. 19, no. 3, pp. 218-222.

Basaraba, D, Yovanoff, P, Alonzo, J & Tindal, G 2013, 'Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist?', *Reading and Writing,* vol. 26, no. 3, pp. 349-379.

Begeny, JC & Greene, DJ 2014, 'Can readability formulas be used to successfully gauge difficulty of reading materials?', *Psychology in the Schools,* vol. 51, no. 2, pp. 198-215.

Benjamin, RG 2012, 'Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty', *Educational Psychology Review,* vol. 24, no. 1, pp. 63-88.

Coleman, M & Liau, TL 1975, 'A computer readability formula designed for machine scoring', *Journal of Applied Psychology,* vol. 60, no. 2, pp. 283-284.

Collins-Thompson, K 2014, 'Computational assessment of text readability: A survey of current and future research', *ITL-International*

Journal of Applied Linguistics, vol. 165, no. 2, pp. 97-135.

Courtis, JK 1987, 'Fry, smog, lix and rix: Insinuations about corporate business communications', *The Journal of Business Communication,* vol. 24, no. 2, pp. 19-27.

Dale, E & Chall, JS 1948, 'A formula for predicting readability: Instructions', *Educational research bulletin***,** Vol. 27, No. 2, pp. 37-54.

DBE 2011a, *English first additional language, Further Education and Training Phase Grades 10–12: Curriculum and Assessment Policy Statement (CAPS),* Government printing works, Pretoria.

DBE 2011b, *English home language, Further Education and Training Phase Grades 10–12: Curriculum and Assessment Policy Statement (CAPS),* Government Printing Works, Pretoria.

DBE 2011c, *English second additional language, Further Education and Training Phase Grades 10–12: Curriculum and Assessment Policy Statement (CAPS)*, Government Printing Works, Pretoria.

DBE 2012, *National policy pertaining to the programme and promotion requirements of the National Curriculum Statement Grades R – 12,* Government printing works, Pretoria.

DBE 2016*, Curriculum and Assessment Policy Statement Foundation Phase Grades 1-3: English Second Additional Language,* Government printing works, Pretoria.

DBE 2017, *Official Languages: First Additional Language Examination Guidelines,* Department of Basic Education, viewed 13 Aug 2021, https://www.education.gov.za/Portals/0/CD/2017%20NSC%20Exam%20Guidelines/AMENDED%20FAL%20GR%2012%20Exam%20Guidelines%202017.pdf?ver=2017-02-22-095618-000

DBE 2019, *The discontinuation of the multiple examination opportunities (MEO) dispensation with effect from 2020*, Circular E 29 of 2019, Department of Basic Education, viewed 26 Aug 2021, https://www.education.gov.za/Portals/0/Documents/Publications/Circular%20E29%20of%202019.pdf?ver=2019-10-28-142703-507

De Vos, M, Van der Merwe, K & Van der Mescht, C 2014, 'A linguistic research programme for reading in African languages to underpin CAPS.' *Journal for Language Teaching* vol. *48, no.* 2, pp. 149-177.

De Wet, A 2021, The development of a contextually appropriate measure of individual recovery for mental health service users in a South African context, doctoral thesis, Stellenbosch University, Stellenbosch.

Duff, D 2019, 'The effect of vocabulary intervention on text comprehension: Who benefits?', *Language, speech, and hearing services in schools,* vol. 50, no. 4, pp. 562-578.

Eltorai, AE, Naqvi, SS, Ghanian, S, Eberson, CP, Weiss, APC, Born, CT & Daniels, AH 2015, 'Readability of invasive procedure consent forms', *Clinical and translational science*, vol. 8, no. 6, pp. 830-833.

Flesch, R 1948, 'A new readability yardstick', *Journal of applied psychology,* vol. 32, no. 3, pp. 221-233.

Gunning, R 1952, '*The technique of clear writing*', McGraw-Hill: New York.

Gunning, R 1969, 'The fog index after twenty years', *Journal of Business Communication,* vol. 6, no. 2, pp. 3-13.

Harden, S 2018, Comparison of readability indices with grades 1-5 narrative and expository texts, doctoral thesis, Wayne State University, Michigan.

Heydari, P 2012, 'The validity of some popular readability formulas', *Mediterranean Journal of Social Sciences,* vol. 3, no, 2, pp. 423-423.

Jansen, C, Richards, R & Van Zyl, L 2017, 'Evaluating four readability formulas for Afrikaans', *Stellenbosch Papers in Linguistics Plus*, vol. *53*, pp.149-166.

Joubert, K & Githinji, E 2014, 'Quality and readability of information pamphlets on hearing and paediatric hearing loss in the Gauteng Province, South Africa', *International journal of pediatric otorhinolaryngology,* vol. 78, no. 2, pp. 354-358.

Kärbla, T, Uibu, K & Männamaa, M 2019, 'National Estonian-language tests: What is measured in text comprehension tasks?', *New Trends and Issues Proceedings on Humanities and Social Sciences,* vol. 6, no. 5, pp. 8-16.

Kärbla, T, Uibu, K & Männamaa, M 2020, 'Teaching strategies to improve students' vocabulary and text comprehension', *European Journal of Psychology of Education***,** vol. 35, pp. 1-20.

Kate, R, Luo, X, Patwardhan, S, Franz, M, Florian, R, Mooney, R, Roukos, S & Welty, C 2010, 'Learning to predict readability using diverse linguistic features', *23rd International Conference on Computational Linguistics' (Coling 2010)*, August 2010, Beijing, pp. 546-554.

Kincaid, JP & Delionbach, LJ 1973, 'Validation of the Automated Readability Index: A follow-up', *Human Factors,* vol. 15, no. 1, pp. 17-20.

Kincaid, JP, Fishburne Jr. RP, Rogers, RL & Chissom, BS 1975, *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*, vol. 56, pp. i-39.

Leopeng, MT 2019, Translations of informed consent documents for clinical trials in South Africa: are they readable?, masters dissertation, University of Cape Town, Cape Town.

Mc Laughlin, GH 1969, 'SMOG grading-a new readability formula', *Journal of reading,* vol. 12, no.8, pp. 639-646.

Mcnamara, DS, Louwerse, MM & Graesser, AC 2002, Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension, Technical report, Institute for Intelligent Systems, University of Memphis.

Mesmer, HA, Cunningham, JW & Hiebert, EH 2012, 'Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future', *Reading research quarterly,* vol. 47, no. 3, pp. 235-258.

Meyer, BJ 2003, 'Text coherence and readability', *Topics in Language Disorders,* vol. 23, no. 3, pp. 204-224.

Nouwens, S, Groen, MA & Verhoeven, L 2016, 'How storage and executive functions contribute to children's reading comprehension', *Learning and Individual Differences,* vol. 47**,** pp. 96-102.

Nyman, P, Kearl, BE & Powers, RD 1961, 'An attempt to shorten the word list with the Dale-Chall readability formula', *Educational Research Bulletin*, vol. 40, no. 6, pp.150-152.

Phillips Galloway, E, Uccelli, P, Aguilar, G & Barr, CD 2020, 'Exploring the cross-linguistic contribution of Spanish and English academic language skills to English text comprehension for middle-grade dual language learners', *AERA Open,* vol. 6, no. 1, pp. 1-20.

Probert, T, & De Vos, M 2016 'Word recognition strategies amongst isiXhosa/English bilingual learners: The interaction of orthography and language of learning and teaching.' *Reading & Writing-Journal of the Reading Association of South Africa*, vol. 7, no. 1, pp.1-10.

Senter, R & Smith, EA 1967, 'Automated readability index', University of Cincinnati, Ohio.

Sibanda, L 2013, A case study of the readability of two Grade 4 Natural Sciences textbooks currently used in South African schools, masters thesis, Rhodes University, Grahamstown.

Spencer, M, Gilmour, AF, Miller, AC, Emerson, AM, Saha, NM & Cutting, LE 2019, 'Understanding the influence of text complexity and question type on reading outcomes', *Reading and writing,* vol. 32, no. 3, pp. 603-637.

Stahl, SA 2003, 'Vocabulary and readability: How knowing word meanings affects comprehension', *Topics in language disorders,* vol. 23, no. 3, pp. 241-247,

Stenner, AJ, Burdick, H, Sanford, EE & Burdick, DS 2006, 'How accurate are Lexile text measures?', *Journal of Applied Measurement,* vol. 7, no. 3, pp. 307-322.

Stevens, KT, Stevens, KC & Stevens, WP 1992, Measuring the readability of business writing: The cloze procedure versus readability formulas, *The Journal of Business Communication (1973),* vol. 29, no. 4, pp. 367-382.

Tennent, W 2014, *Understanding reading comprehension: Processes and practices,* Sage, London.

Tshesane, TMJ 2014, Evaluating the functionality of the translated Sepedi Home Language CAPS for Grade 10-12, masters research report, University of the Witwatersrand, Johannesburg.

Vajjala, S & Meurers, D 2014, 'Readability assessment for text simplification: From analysing documents to identifying sentential simplifications', *ITL-International Journal of Applied Linguistics,* vol. 165, no. 2 pp. 194-222.

Van Der Walt, C 2010, 'Of shoes-and ships-and sealing-wax: A dynamic systems approach to language curriculum orientation', *Southern African Linguistics and Applied Language Studies,* vol. 28, no. 4, pp. 323-337.

Van Der Walt, C 2018, 'The category Language Structures and Conventions in the CAPS for English First Additional Language: a critical analysis', *Journal for Language Teaching,* vol. 52, no. 1, pp. 170-200.

Wissing, GJ, Blignaut, AS & Van Den Berg, K 2016, 'Using readability, comprehensibility and lexical coverage to evaluate the suitability of an introductory accountancy textbook to its readership', *Stellenbosch Papers in Linguistics,* vol. 46, pp. 155-179.

Zhou, S, Jeong, H & Green, PA, 2017, 'How consistent are the best-known readability equations in estimating the readability of design standards?', *IEEE Transactions on Professional Communication*, vol. 60, no. 1, pp. 97-111.