Evaluating four readability formulas for Afrikaans

Carel Jansen

Centre for Language and Cognition, University of Groningen, The Netherlands
Language Centre, Stellenbosch University, South Africa
E-mail: c.j.m.jansen@rug.nl

Rose Richards

Language Centre, Stellenbosch University, South Africa
E-mail: rr2@sun.ac.za

Liezl van Zyl

Language Centre, Stellenbosch University, South Africa
E-mail: Liezl@heyplainjane.com

Abstract
For almost a hundred years now, readability formulas have been used to measure how difficult it is to comprehend a given text. To date, four readability formulas have been developed for Afrikaans. Two such formulas were published by Van Rooyen (1986), one formula by McDermid Heyns (2007) and one formula by McKellar (2008). In our quantitative study the validity of these four formulas was tested. We selected 10 texts written in Afrikaans – five articles from a popular magazine and five documents used in government communications. All characteristics included in the four readability formulas were first measured for each text. We then developed five different cloze tests for each text to assess actual text comprehension. Thereafter, 149 Afrikaans-speaking participants with varying levels of education each completed a set of two of the resulting 50 cloze tests. On comparing the data on text characteristics to the cloze test scores from the participants, the accuracy of the predictions from the four existing formulas for Afrikaans could be determined. Both Van Rooyen formulas produced readability scores that were not significantly correlated with actual comprehension scores as measured with the cloze tests. For the McKellar formula, however, this correlation was significant and for the McDermid Heyns formula the correlation with the cloze test scores almost reached significance. From the outcomes of each of these last two formulas, about 40% of the variance in cloze scores could be predicted. Readability predictions based only on the average number of characters per word, however, performed considerably better: about 65% of the variance in the cloze scores could be predicted just from the average number of characters per word.

Keywords: Afrikaans, cloze test, readability formula, text comprehension, text difficulty

1.   Introduction

For almost a hundred years now,[1] readability formulas have been used to measure objectively, or rather predict, how difficult it is to comprehend a given text.[2] Without doubt, the most famous readability formula was developed in 1948 by Rudolph Flesch. His formula predicts reading ease based on two text characteristics: average word length in number of syllables and average sentence length in number of words.[3] In the past century, readability formulas such as the Flesch formula have gained great popularity. As DuBay (2004:2) notes, around the 1980s already more than 200 different readability formulas existed. The field has exploded since computer technology has made it possible to apply readability formulas in an automated way (Benjamin 2012:63). Most word processors now have one or two built-in formulas that can inform users about the difficulty of a text they are working on, and websites exist that allow visitors to enter a text and immediately receive, often at no cost, readability scores according to a number of different formulas.[4]

Originally, most of the work on readability measurement was done with English-language material. After these early years, however, more and more language-specific formulas were developed (Klare 1974-1975:91), and with good reason to do so: texts in different languages that deal with similar topics do not necessarily share the same characteristics, and the effects of these characteristics on readability may also be quite different. Rabin (1988:66-76) presents different formulas for Chinese, Danish, Dutch, French, German, Hebrew, Hindi, Korean, Russian, Spanish, Swedish and Vietnamese (see also Klare 1974-1975:91-95). Although Rabin (1988) does not mention the existence of readability formulas for Afrikaans,[5] two years before her publication two such formulas had already been presented by Van Rooyen (1986).

In the current article, the Van Rooyen formulas are discussed in some detail, in order to assess the strengths and weaknesses of the development process and the usability of the resulting formulas. We then go into the only other two readability formulas for Afrikaans that we could identify. Both formulas were constructed more than 20 years after the pioneering work of Van Rooyen. One of these formulas was developed by McDermid Heyns (2007); the other was developed by McKellar (2008). Next, we present a study into the validity of the four existing readability formulas for Afrikaans. Should one or more of these formulas prove to produce accurate readability predictions, then these formulas would provide a good base for creating automated readability formulas that may effectively help writers to produce more comprehensible texts in Afrikaans.

---

[1]   According to Fry (2002:286), the first readability formula was published in 1923, when Lively and Pressey published their "method of measuring vocabulary burden of textbooks".

[2]   See Flesch (1948:221).

[3]   The exact Flesch formula reads as follows (see Flesch 1948:223-225; DuBay 2004:21-22):
RE = 206.835 – (1.015 x ASL) – (84.6 x AWL)
RE (Reading Ease) is predicted readability score, ASL = average sentence length (the number of words divided by the number of sentences), and AWL = average number of syllables per word (the number of syllables divided by the number of words).

[4]   For English, see, for instance, two online resources: an automatic readability checker (Readability formulas n.d.) and Juicy Studio's Readability test resource (Juicy Studio 2017). For Dutch, see the review in Jansen and Boersma (2013), as well as, for instance, the online Leesniveau Tool (Accessibility 2017).

[5]   In the same volume in which the article by Rabin appeared, Klare (1988:22) mentions the existence of formulas for Finnish and Afrikaans. However, no references are provided here other than to Rabin's overview in which formulas for both languages are missing.

Our research question is the following: How accurately do the four existing readability formulas for Afrikaans predict actual text comprehension?

2.   The existing formulas for Afrikaans

2.1.   The Van Rooyen formulas

In a study that, according to the author, was the first of its kind in Afrikaans, Rien van Rooyen presented two readability formulas she had developed based on two different types of comprehension tests (Van Rooyen 1986:59). In her article, published in Afrikaans, Van Rooyen discusses the various steps required to arrive at a formula based on empirical data, and shows how she took these steps herself.

First, 30 texts were selected from a series of standardised comprehension tests for Afrikaans, issued by the Institute for Psychological and Edumetric Research of the Human Sciences Research Council (IPER-HSRC). Using five-option multiple-choice tests for these 30 texts, developed by IPER-HSCR for learners in Standard 6 to 10 (presently known as Grade 8 to 12), Van Rooyen collected comprehension scores from 1 200 randomly selected Afrikaans-speaking learners from six high schools in the suburbs of Pretoria. Furthermore, among the same group of learners, so-called 'cloze tests' were taken. For this purpose, in all texts, every fifth word was deleted and learners were asked to fill each empty space with the word they expected to have been there in the original text (see section 3.4 on the validity of the cloze test). To prevent unwanted knowledge effects, no learner was asked to complete a cloze test on the same text about which he or she was requested to answer multiple-choice questions (1986:60).

As the next step, the 30 texts were linguistically analysed. In each text, 16 variables were measured: average number of words per sentence, average number of syllables per sentence, average number of letters per sentence, average number of syllables per word, average number of letters per word, number of words with one syllable per 100 words, number of words with three or more syllables per 100 words, number of prepositions per 100 words, number of conjunctions per 100 words, number of verbs per 100 words, number of adverbs per 100 words, number of adjectives per 100 words, number of pronouns per 100 words, number of abstract nouns per 100 words, number of concrete nouns per 100 words, and number of infinitive constructions per 100 words (1986:62).

After this, statistical analyses could be performed. First, correlations were calculated between the linguistic variables and both comprehension score variables: average number of correct answers to the multiple-choice questions, and average number of correctly filled blanks in the cloze tests. Next, for each comprehension score variable, multiple linear regression analyses were performed to determine the optimal combination of linguistic predictor variables (including quadratic and interaction terms). As a result, Van Rooyen could present two formulas (1986:64-65). In her first readability formula (hereafter RE-Van Rooyen-MC), Reading Ease (RE) is expressed as the predicted mean score on the IPEN-RGN multiple-choice test. In the second readability formula (hereafter RE-Van Rooyen-CL), Reading Ease (RE) stands for the predicted mean score on the cloze test.

RE-Van Rooyen-MC:

$$RE = 67.323 + (0.074*Slg*K) + (1.167*I*K) - (0.339*Slg) - (2.475*K)$$

RE = predicted mean score on the IPEN-RGN multiple-choice test; Slg = average number of syllables per sentence; K = number of concrete nouns per 100 words; I = number of infinitive constructions per 100 words.

RE-Van Rooyen-CL:

$$RE = 22.898 + (0.154*Slet) + (3.479*V) - (0.334*D) - (0.033*Slet*V)$$

RE = predicted mean score on the cloze test; Slet = average number of letters per sentence; D = number of words with three or more syllables per 100 words; V = number of pronouns per 100 words.

Van Rooyen does not present exact figures about the correlations between the scores predicted by her formulas, on the one hand, and the actual scores from the IPEN-RGN multiple-choice tests and the cloze tests, on the other hand. The approximate size of these correlations can be deduced, however, from a remark on p. 65 (translated here into English): "For both formulas the percentage of explained variance in the dependent variable is about 40 per cent". This implies that for both formulas the correlation with the actual learner scores must be around .63 (the square root of .40).[6]

Van Rooyen concludes that despite the differences in the linguistic variables used in the two formulas, the outcomes are highly comparable (1986:65). The article ends with detailed instructions on how to apply the formulas, including definitions of the linguistic characteristics that have to be counted: the numbers of syllables, sentences, concrete nouns, words, infinitive constructions, letters, words with three or more syllables, and pronouns.

Although we appreciate the work by Van Rooyen, some critical remarks should be made here, one being that it is hard to determine on exactly which data her comprehension findings were based. No information is provided, for instance, about the number of learners per text who participated in the comprehension tests. What the article does reveal about this subject suggests that the 1 200 learners who participated each carried out one comprehension test, either in multiple-choice test format or in cloze test format. That would mean that Van Rooyen obtained data from (1 200/30)/2=20 learners per text per test. It also is not clear how the learners were distributed among different age groups. Van Rooyen's decision to use IPEN-RGN tests developed for learners in Grade 8 to 12 suggests that in her study there may have been five different age groups. This would imply that per text, per comprehension test and per age group, data from on average four readers were used (1986:60). Be that as it may, all participants in this study were learners aged about 12 to 16 years, and all texts were intended to be read by this target group. As Van Rooyen herself points out, the formulas may, by implication, be regarded as valid only for texts intended for learners of about 14 tot 18 years (1986:65).

Next, a practical problem in the application of RE-CL should be mentioned. One of the variables in this formula is the number of pronouns per 100 words. According to Van Rooyen's definition of pronouns (translated here in English) "all pronouns should be included: personal pronouns such as *ek* (I), *jy* (you), *ons* (we), *dit* (it), and *hulle* (they); possessive pronouns such as *watter* (what/which), *waarmee* (with which) and *waarop* (on which); indicative pronouns such as *daarmee* (with that); and other pronouns such as *iemand* (someone), *party* (some),

---

[6]   See, for instance, Field (2009:179).

enigeen (anyone), *daar* (there) and *iets* (something)" (1986:67). From a grammatical point of view, part of this definition is confusing and incorrect. A minor problem is that *watter* (what) is not a possessive but an interrogative pronoun. More importantly, however, is that *waarmee* (with what), *waarop* (on which) and *daarmee* (with that) are not pronouns but pronominal adverbs. This makes it difficult for users of RE-CL to decide which words in a given text should be considered as pronouns. As a consequence, it may be difficult to decide what figure should be entered in the formula for "number of pronouns per 100 words".

Finally, Van Rooyen does not provide any information about the relationship between the outcomes of her comprehension tests. This leaves it unclear to what extent a text that according to the results from the multiple-choice tests should be considered as difficult would also qualify as difficult according to the results from the cloze tests, and vice versa. Furthermore, Van Rooyen justified her claim about the comparability of the outcomes of the two formulas (1986:65) with the finding that the linguistic variables in both formulas explain about 40% of the variance in the comprehension scores they predict. This similarity, however, is not relevant for the possible resemblance between the predictions from the two formulas. How comparable these predictions are can only transpire from their mutual correlation. The reader is not informed, however, about the relationship between the outcomes of the two Van Rooyen formulas.

2.2.   The McDermid Heyns formula

Twenty-one years after Van Rooyen published her readability formulas, Jacques McDermid Heyns read a paper in Afrikaans (unpublished, but available from the authors of this article)[7] at the 2007 Joint Annual Conference of the Southern African linguistic organisations LSSA/SAALT/SAALA, held in Potchefstroom. Apparently unaware of the work done by Van Rooyen (1986), McDermid Heyns claimed that, until that time, no readability formulas had been developed for Afrikaans. To fill this perceived gap, McDermid Heyns developed three formulas after the example of the Flesch formula (see note 3). His first formula was based on a sentence-by-sentence analysis of one text of 1 660 words, namely the Afrikaans version of the "Universal Declaration of Human Rights". The second formula was derived from a text-by-text analysis of 90 texts written in Afrikaans (altogether 28 206 words). For developing both formulas McDermid Heyns first counted the number of sentences, words and syllables, and then used the outcomes as linguistic predictors in multiple linear regression analyses with readability scores for the texts as the dependent variable. McDermid Heyns does not, however, explain what the source of his readability scores was. In view of the information from McDermid Heyns that he "adapted" his formulas according to the analyses of the sentences and texts, it seems probable that he used readability scores directly derived from applying the Flesch formula to the texts he had analysed (see also McKellar 2008; Pienaar 2009:58). Comparing the standard deviations in the outcomes of the two regression analyses, McDermid Heyns decided in favour of his first formula. According to McKellar (2008) however, a third and final formula based on the work by McDermid Heyns[8] was included in *Skryfgoed* (2008), a collection of proofing tools for Afrikaans. Unfortunately Microsoft updates did not make it possible to

---

[7]   We thank Martin Puttkammer, supervisor of Jacques McDermid Heyns at the time of his studies at North West University, for providing us with this paper.

[8]   No information is available on the way this formula was developed.

use this formula in new versions of *Skryfgoed*.[9,10] The final McDermid Heyns formula is as follows:

RE-McDermid Heyns = 138.8989 – (1.0052 x ASL) – (35.4562 x AWL)

RE = predicted readability score, ASL = number of words per sentence, and AWL = number of syllables per word.

2.3.   The McKellar formula

One year after McDermid Heyns had presented his formula, Cindy McKellar introduced a new formula in a poster session (in English) at the 2008 conference of the Pattern Recognition Association of South Africa (PRASA), held in Cape Town (McKellar 2008). For developing her formula, McKellar followed the example of the way the Flesch formula was developed. In her analyses, however, text characteristics were added from two other formulas for English: the Dale-Chall formula, including percentage of familiar words in a text, and the Automated Readability Index, including average word length in characters (see DuBay 2004:23, 49). McKellar analysed 200 texts in Afrikaans, collected from sources such as textbooks, storybooks and web pages. For each text, a computer program counted all sentences, words, syllables, characters, familiar words, brackets (in pairs), numbers and symbols. Familiar words were identified using frequency counts in the *Media24* corpus from *Pharos Dictionaries*. McKellar then used the outcomes of the text analyses as linguistic predictors in multiple linear regression analyses with readability scores for the texts as the dependent variable. McKellar explains how these readability scores were collected as follows: "Each of the 200 texts was given a readability index manually. This was done on a scale of 0 to 100, where 0 represents an extremely difficult text and 100 an extremely easy one. The indexing was performed by five different persons, in such a way that each text was indexed by at least two persons. In cases where the indexes differed, the average index was calculated." The formula that, according to McKellar, predicted the readability scores best is shown below:

RE-McKellar = 168.6145 – (0.7326 x ASL) – (52.2133 x AWL) – (0.1077 x %FW) – (0.4539 x NB) – (0.7061 x NNS)

RE-McKellar = predicted readability score, ASL = number of words per sentence, AWL = number of syllables per word, %FW = percentage of words that can found in the list comprising the 50 most frequently used words in the *Media24* corpus, NB = number of brackets used (counted in pairs), and NNS = number of numbers and symbols (anything that does not consist of alphanumerical characters).[11]

Finally, the formula was evaluated with 20 different texts, selected from various sources. These texts were indexed by seven people who were not involved in the development of the formula. The accuracy of the predictions from the McKellar formula was compared with the accuracy of the predictions from the final McDermid Heyns formula that, according to McKellar, was "based on the Flesch reading ease score". It was found that the McKellar formula performed best. The correlation between the outcomes of this formula and the assessments of the

---

[9]   Personal correspondence with Martin Puttkammer and Gerhard van Huyssteen (North West University, Potchefstroom).

[10]   For the most recent version, see *Skryfgoed* (2017).

[11]   We assume that NB is meant to be calculated as number of pairs of brackets per 100 words, and NNS as number of numbers and symbols per 100 words.

readability judges was .91; for the McDermid Heyns formula a correlation coefficient of .90 was found. The existence of the Van Rooyen formulas is not mentioned in this article.

Comparing the development of the four formulas discussed above, it is striking that it was only for the Van Rooyen formulas that actual text comprehension data were used. It appears that McDermid Heyns did not present his texts to a sample of readers from the target group at which the texts were aimed. McKellar based her readability scores on estimates from judges. From the information provided in her article it is unclear how often the judges did or did not agree on their assessments. More importantly, it is unclear to what extent these judges may be considered to be experts in this field; hence it is impossible to tell to what extent their assessments would correctly predict the levels of text difficulty for the intended readership.

Since, as far as we know, no independent evaluation studies have been carried out to test the validity of the four existing formulas for Afrikaans, we decided to perform such a study, and to use actual comprehension data from readers speaking Afrikaans as their first language.

3.   Method

To evaluate the four existing formulas for Afrikaans, ten texts written in Afrikaans were selected. First, for each text all linguistic characteristics included as linguistic predictors in one or more of the four formulas were measured. After this, comprehension was measured for each of these texts, using cloze test scores from a total of 149 Afrikaans-speaking participants with varying levels of education. Finally, using both the outcomes of the four formulas and the cloze test scores from the participants, the accuracy with which each of the four formulas could predict cloze scores was determined. More detailed information regarding the method is presented in sections 3.1 to 3.5.

3.1.   Texts

Ten texts of approximately 300 words each were used. Five texts were sourced from the *Huisgenoot*, a popular Afrikaans magazine, and five texts were sourced from official government communications in Afrikaans. All texts were retyped, in order to make them appear similar and to reduce possible interference with readability by, for instance, layout and letter font. For an overview of all ten texts, including their titles and numbers of words, see Appendix 1.[12]

3.2.   Participants

One hundred and fifty Afrikaans-speaking people living in the vicinity of Stellenbosch and Cape Town took part in this study. One participant completed only the very first parts of the cloze tests with which he was presented, hence his data were not used in the statistical analyses. The remaining 149 participants varied in gender (42 masculine, 45 feminine, 62 no information available); ethnicity (32 coloured, 32 white, 2 black, 83 no information available); highest level of education (1 Grade 9, 4 Grade 11, 74 Grade 12, 5 certificate, 45 university degree, 18 diploma, 2 other); and mean age (whole group of participants: 31.16 years). For more detailed information, see Table 1.

---

[12]   All materials are available from the authors upon request.

Table 1: Participants' gender, ethnicity, level of education and mean age (between brackets)

		Grade 9	Grade 11	Grade 12	Certificate	University degree	Diploma	Other
Masculine	**Coloured**	0	2 (50.50)	9 (30.11)	2 (24.00)	1 (24.00)	2 (19.50)	0
	White	0	0	4 (25.25)	0	6 (42.67)	2 (58.00)	0
	Black	0	0	1 (18.00)	0	0	0	0
	No ethnicity information	0	0	11 (27.36)	0	2 (21.50)	0	0
Feminine	**Coloured**	1 (48.00)	1 (46.00)	4 (20.75)	0	7 (37.71)	3 (47.33)	0
	White	0	0	5 (22.40)	1 (49.00)	12 (37.33)	2 (57.00)	0
	Black	0	0	1 (48.00)	0	0	0	0
	No ethnicity information	0	0	3 (29.00)	0	1 (40.00)	4 (40.75)	0
No gender information	**No ethnicity information**	0	1 (52.00)	36 (21.92)	2 (51.50)	16 (27.81)	5 (55.20)	2 (29.00)
Total		**1 (48.00)**	**4 (49.75)**	**74 (24.46)**	**5 (40.00)**	**45 (32.84)**	**18 (47.22)**	**2 (29.00)**

3.3.   Measures: Formula scores

In order to determine the readability scores according to the four formulas, all relevant linguistic characteristics were calculated for each text.

- For the Van Rooyen RE-MC: average number of syllables per sentence, number of concrete nouns per 100 words, and number of infinitive constructions per 100 words;

- for the Van Rooyen RE-CL: average number of letters per sentence, number of words with three or more syllables per 100 words, and number of pronouns per 100 words;

- for the McDermid Heyns RE: number of words per sentence, and number of syllables per 100 words;

- for the McKellar RE: number of words per sentence, number of syllables per word, percentage of words that can found in the list comprising the 50 most frequently used words in the *Media24* corpus,[13] number of brackets used (counted in pairs), and number of numbers and symbols (anything that did not consist of alphanumerical characters); and

- number of characters per word (as per Jansen and Boersma; 2013:59; see section 3.5).

---

[13]   We thank Gerhard van Huyssteen for providing us with this list.

Next, the scores found for these characteristics were entered in the respective formulas, and the resulting RE-scores were determined.[14] For the outcomes per text, see Appendix 1.

3.4.   Measures: Cloze tests

In order to measure actual text comprehension, five cloze tests were taken for each text. In a cloze test a number of words in a given text are deleted and replaced by blanks or dashes. All blanks or dashes have the same length. Participants are requested to substitute each blank or dash with the word they expect to have been there in the original text. The number of correct answers is expressed as a percentage of the total number of omitted words in the text. The average percentage for a sample of readers from the target group of the text is interpreted as a valid measure of the comprehensibility of the original text.[15]

As Horton (1974-1975:250) states, both the construct validity and the concurrent validity of the cloze procedure have been established. The construct validity of the cloze test addresses the subject's "ability to deal with the linguistic structure of the language; it is related to the ability of the subject to deal with the relationships among words and ideas". The concurrent validity refers to "the variances shared among cloze tests, reading comprehension tests, reading gain tests, and verbal intelligence tests [which] probably [are] a measure of the reader's ability to deal with the relationships among words and ideas".

The construct validity of the cloze test may be explained as follows: reading may be viewed as a "psycholinguistic guessing game", "a selective process [involving] partial use of minimal available language cues selected from perceptual input on the basis of the reader's expectation" (Goodman 1967:126-127; see also Jansen and Boersma 2013:51-52). Readers scan a page and pick up graphic cues guided by their prior choices, language knowledge, cognitive styles and the strategies they have learned. They try to relate the resulting perceptual image to syntactic, semantic and phonological cues stored in memory and make a tentative choice about a possible interpretation of the perceptual input. Then they test their choice for grammatical and semantic acceptability in the context developed by earlier interpretation choices. If the tentative choice turns out to be unacceptable, the process starts anew; if the choice proves to be acceptable, reading continues and expectations are formed about what lies ahead (Goodman 1967:127, 134-135; see also Alderson 2000:17, 19).

What participants in a cloze test are asked to do strongly resembles what is requested from readers of the text on which the cloze test is based. To be able to guess correctly which words were deleted from such a text, participants in a cloze test must possess the same types of skills and knowledge as are required from readers. Both cloze test participants and readers must make use of their language knowledge, cognitive styles and strategies, and syntactic, semantic and phonological cues stored in memory. They must make tentative choices about the words that may have been deleted, and they must test these choices for grammatical and semantic acceptability in the context of the test or the text.

---

[14]   In cases where there was doubt, for instance when counting the number of pronouns per 100 words for the Van Rooyen RE-CL formula (see section 2.1), the definition provided by the developer of the formula was followed as closely as possible. As a close approximation for the average number of letters per sentence (Slet), as in Van Rooyen (1986:67), the average number of characters per sentence excluding spaces was determined.

[15]   See Fry (1989:295) for this and other possible validity measures.

Jones (1997) provides a critical appraisal of the use of cloze tests to measure readers' understanding of texts used in the field of accounting. His most important argument goes against the construct validity of the cloze test. In his view "the skills necessary to infer missing words from accounting narratives may be very different from the skills necessary to comprehend accounting text" (Jones 1997:118). In his appraisal Jones, however, does not pay attention to the similarities already mentioned between a reader's tasks and what is asked of participants in a cloze test. These similarities perfectly explain the high level of concurrent validity that follows from the strong correlations found between cloze test scores and scores in traditional reading comprehension tests, as reported in Bormuth (1967; 1968), Kamalski (2007) and Gellert and Elbro (2013), for instance.

The cloze test can be administered in different ways. One of the choices to be made is between the fixed-ratio method and the rational fill-in method. The fixed-ratio method requires that the nth word be replaced by a blank or a dash (for instance the first, sixth, eleventh word, etc.). When applying the rational fill-in method, the researchers themselves decide which words will be left out, for instance, only verbs, nouns, adjectives, technical words and/or prepositions (Kobayashi 2002:573; O'Toole and King 2011:129). Already in 1953, the founding father of the cloze test, Wilson Taylor, advocated following the fixed-ratio method (Taylor 1953:419-420). Taylor's plea is supported by the results reported in Bachman (1985). Bachman compared cloze scores of 910 students in total. He reports that the fixed-ratio method and the rational fill-in method led to comparable outcomes, and that scores collected with both methods strongly correlated with other comprehension measures (1985:544, 546).

A disadvantage of the fixed-ratio method, however, may be that – depending on the number of the starting word – different cloze tests versions may result in different outcomes: not all words are equally redundant, hence some blanks may be more difficult to fill than others (Bachman 1985:537-538; Kobayashi 2002:581-583), and the meaning of a score depends on which items are chosen from a text (Abraham and Chapelle 1992:474). An easy solution for this problem is to create different cloze versions, as suggested by Bormuth (1967:2) and Jongsma (1971:26): one version starting with deleting the first word, a second version starting with deleting the second word, and so on. As O'Toole and King (2010:314) conclude from their study into the impact of the location of the first deleted word: "it would be wise for teachers and material developers to exhaustively sample the text of the passages whose more general readability they wish to estimate. The generation of five cloze tests from a passage is relatively easy and exact scoring would enable a much clearer picture of the ease or difficulty of the passage to emerge."

Another decision that has to be made when administering a cloze test concerns the way in which the participants' answers are scored. The first possibility is *exact scoring*: a word filled in at a given dash is only considered correct if it matches the word that was deleted from the original text, leaving out of account minor differences in spelling. The second option is to apply *conceptual scoring*. Following this approach, an answer is also correct if it may be considered as a synonym of the word from the original text. At first sight, conceptual scoring may seem preferable: filling in a synonym seems like a fair indication of text comprehension. Several studies, however, found strong relationships between the outcomes of the two different ways of scoring (see, for instance, McKenna 1976:142; Litz and Smith 2006:55, 68; O'Toole and King 2011:135-139). Taylor (1953:432) was already in favour of exact scoring, as it helps to avoid "the problem of coder reliability that so plagues content analysis". O'Toole and King (2011:140) also conclude that for measuring the comprehensibility of different texts, exact scoring is more appropriate than conceptual scoring; it may be regarded as "easier, less

subjective, sufficiently reliable [and] highly correlated with conceptual scoring" and "[exact scoring] does not advantage particular sections of the reading population". Having compared 24 cloze test studies, Watanabe and Koyama (2008) conclude that conceptual scoring leads to more reliable outcomes than exact scoring. It should be noted, however, that their meta-analysis only refers to studies with cloze tests taken in another language than the participants' mother tongue. It seems hard to predict what the results would have been if outcomes had been analysed from studies with cloze tests in the first language of participants. Owing to a larger vocabulary of such participants, their answers may have varied to a greater extent than was the case in the studies that Watanabe and Koyama (2008) refer to – and so might the outcomes of the conceptual scoring method.

In view of the mentioned considerations, we decided to follow the exact scoring approach and to apply the fixed-ratio method. The exact scoring approach serves the purposes of both reliability and the speed of scoring. The fixed-ratio method does not require any subjective decisions about the words that should or should not be deleted. To prevent that chance decisions about the starting word would influence the test results, five different cloze test versions were developed for each text. In all versions, every fifth word was deleted – in the first versions starting with the first word, in the second versions starting with the second word, etc. In this way we ensured that each word in every text would play a role in the cloze scores we collected, in the same way that each word in the original texts plays a role for readers of these texts.

3.5.   Procedure

The participants in this study were approached by colleagues and students from Stellenbosch University. All participants completed a consent form, which included information about the purpose, procedures, risks and benefits of the study, and reminded them of the confidentiality of their responses and their rights as participants.[16] After this, each participant was presented with a pack of two cloze tests: one of these cloze tests emanated from one of the *Huisgenoot* texts, and the other cloze test was derived from one of the government texts. As a result, 149*2=298 cloze scores could be calculated for 5*10=50 cloze test versions, amounting to on average 5*298/50=29.8 cloze scores per text. After this, each completed text was scored. Following the definitions from the developers of the four formulas as carefully as possible, the scores were determined for all linguistic characteristics included in the formulas. Finally, statistical analyses were performed relating the cloze scores to the outcomes for the linguistic characteristics and the predictions from the formulas. One extra characteristic was added: the number of characters per word. In a comparable study carried out in the Netherlands into the validity of Dutch readability formulas, it was found that this characteristic strongly correlated with cloze scores (Jansen and Boersma 2013:59). In view of this outcome, we decided to measure the correlation between this characteristic and the cloze scores we collected.

4.   Results

Table 2 presents the correlations between the readability predictions from the four formulas and the average cloze test scores for the ten texts. The predictions and cloze scores for each text can be found in Appendix 1.

---

[16]   Permission for conducting this study was granted by the Stellenbosch Research Ethics Committee: Human Research (Humanities): Proposal number SU-HSD-002010; NHREC registration number REC-050411-032.

Table 2: Correlations between readability predictions from the four existing Afrikaans formulas and average cloze scores

	RE-Van Rooyen-MC	RE-Van Rooyen-CL	RE-McDermid Heyns	RE-McKellar	Average cloze score
RE-Van Rooyen-MC	1	r=-.12 p=.73	r=.01 p=.98	r=.01 p=.97	r=-.21 p=.56
RE-Van Rooyen-CL		1	r=.92 p<.001	r=.88 p=.001	r=.34 p=.33
RE-McDermid Heyns			1	r=.99 (p<.001)	r=.61 p=.06
RE-McKellar				1	r=.64 p=.04
Average cloze score					1

Table 2 shows that for the predictions of the McKellar formula a statistically significant (p<.05) correlation with average cloze scores (r=.64; p=.04) was found; and for the McDermid Heyns formula the correlation almost reached significance (r=.61; p=.06). By implication, from the outcomes of these formulas, about 40% of the variance in the cloze scores could be predicted ($100*(0.64^2)=40.96\%$, and $100*(0.61^2)=37.21\%$). For both Van Rooyen formulas the correlations with average cloze scores were considerably lower and not significant (r=-.21; p=.56 and r=.34; p=.33).

Furthermore, the very weak and non-significant correlation (r=-.12; p=.73) between the outcomes of the two Van Rooyen formulas is remarkable, as is the strong, almost perfect correlation between the scores according to the McDermid Heyns formula and the McKellar formula (r=.99; p<.001). Striking also are the strong and significant correlations between the RE-Van Rooyen-CL formula on the one hand, and the McDermid Heyns formula (r=.92; p<.001) and the McKellar formula (r=.88; p=001) on the other hand.[17]

Table 3 presents the outcomes of correlation analyses, including the cloze scores, on the one hand, and the text characteristics included in the four formulas, on the other hand.

---

[17]   These outcomes may seem to contrast with the differences in correlations found between the cloze scores and the Van Rooyen-CL formula (r=.34) and the correlations between the cloze scores and the McDermid Heyns formula (r=.61) and the McKellar formula (r=.64). This contrast may, however, be explained by the limited number of texts involved in determining these correlations.

Table 3: Correlations between cloze scores and text characteristics included in the four existing Afrikaans formulas

	Average cloze score	
	r	**p**
Syllables per sentence	-.54	.11
Concrete nouns per 100 words	.20	.58
Infinitive constructions per 100 words	-.58	.08
Letters per sentence	-.55	.10
Words with ≥3 syllables per 100 words	-.57	.09
Pronouns per 100 words	.67	.03
Words per sentence	-.18	.61
Syllables per word	-.72	.02
Percentage of words in list of 50 most frequently used words	.42	.23
Pairs of brackets per 100 words	-.01	.99
Numbers and symbols per 100 words	-.07	.84
Characters per word	-.81*	.005

As Table 3 shows, significant correlations with average cloze score were found only for number of pronouns per 100 words (r=.67; p=.03) and average number of syllables per word (r=-.72; p=.02). Entering both predictors in a regression analysis with cloze score as dependent variable revealed that a formula including these two predictors would not lead to predictions significantly correlating with cloze scores. The correlation, however, between average cloze score and number of characters per word proved to be strong and significant: r=-.81; p=.005. In other words, $100*(-0.81^2)=65.61\%$ of the variance in the cloze scores could be predicted from just the average number of characters per word.

Regression analyses revealed that adding any other text characteristic to this predictor would not lead to stronger correlations with average cloze score.

5.   Conclusions

At present, four readability formulas for Afrikaans exist: two formulas developed by Van Rooyen (1986), one formula developed by McDermid Heyns (2007) and another formula developed by McKellar (2008). This study tested how accurately these formulas would predict actual text comprehension. Both Van Rooyen formulas proved to produce readability scores that were not significantly correlated with actual comprehensibility scores as measured with cloze tests. For the McKellar formula, however, this correlation was significant; and for the McDermid Heyns formula the correlation with the cloze test scores almost reached significance. From the outcomes of both the McDermid Heyns formula and the McKellar formula, about 40% of the variance in cloze scores could be predicted. Readability predictions based only on the average number of characters per word, however, proved to perform considerably better. About 65% of the variance in the cloze scores could be predicted from just the average number of characters per word.

Publications on the development of the existing formulas revealed a number of shortcomings that might help to explain the relatively low level of accuracy of the predictions. Van Rooyen used data only from participants in Grade 8 to Grade 12, and all texts she used were intended for this target group. As a consequence, her two formulas may be valid only for readers with comprehension levels comparable to those of learners aged about 14 to 18 years. In our study, however, text comprehension data were collected from a different and wider group of speakers of Afrikaans.

A remarkable result we found pertaining to the two Van Rooyen formulas was the almost absent correlation between the mutual predictions of these formulas. Van Rooyen (1986) does not mention in her study how predictions from her two formulas were correlated. In future studies it would be interesting to see how the outcomes of the Van Rooyen formulas relate to each other, and also to the outcomes of the McDermid Heyns formula and the McKellar formula.

In the introduction of this article, we alluded to the possibility that, in order to help writers to produce better texts in Afrikaans, automated readability formulas could be created. Should one decide to try to create automated readability formulas for Afrikaans, then a natural option would be to start from existing formulas. Of course, using these formulas for automated applications would only be useful if they prove to produce accurate readability predictions. From this study we cannot conclude that starting from existing formulas for Afrikaans would be a viable option.

A more favourable possibility for the development of automated readability formulas would be to use actual text comprehension data such as those collected in this study. We realise, however, that comprehension data from more texts and more text types would be needed. Therefore, we are in the process of collecting cloze test scores from altogether 225 new participants in respect of 15 new texts in three new text types (medical pamphlets, daily newspapers and insurance brochures). The combined set of data from the present study and the forthcoming study can be related to text characteristics measured by means of advanced language technology.[18] We trust that this will lead to the creation of readability formulas that will outperform the existing formulas for Afrikaans.[19]

An important proviso here is that, in the presentation of such new formulas for Afrikaans, the objections against readability formulas raised in the literature should be adequately addressed. DuBay (2004:28-29), for instance, points at the role that prior knowledge and the motivation of individual readers play in text comprehension. In readability formulas, however, such individual reader characteristics are not taken into account (Schriver 2000:138-139; Jansen and Lentz 2008:7). Dreyer (1984:335) and Duffelmeyer (1985:393) bring up the unsubstantiated high level of precision suggested by the decimal figures in many formulas. Schriver (2000:139-140) remarks that authors may be tempted to "write to the formulas" by simply adding more periods in order to make the sentences shorter. Such authors are misled by the suggestion that a correlational relationship, in this case between sentence length and text comprehensibility, would be equal to a causal relationship. Such text characteristics are indices and not causes of text difficulty (see, for example, Singer 1988:vii).

---

[18]   Such language technology is presently available for Afrikaans at the Centre for Text Technology (CTexT) at the Potchefstroom Campus of the North-West University (North-West University 2017).
[19]   A comparable project in the Netherlands is LIN (LeesbaarheidsIndex voor het Nederlands) (NWO 2017).

Despite obvious shortcomings, a possible benefit of using readability formulas is that they remind writers who are not trained as professional authors to be aware of readability issues.[20] Being confronted with unfavourable readability predictions may be an important first step toward creating texts that readers understand and appreciate: forewarned is forearmed (see also Schriver 2000:140). Automated readability formulas grounded in advanced research and presented with sufficient caution may be a real asset for users of Afrikaans.

References

Abraham, R.G. and C.A. Chapelle. 1992. The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal* 76(4): 468-479.

Accessibility. 2017. Leesniveau tool. Available online: https://www.accessibility.nl/kennisbank/tools/leesniveau-tool (Accessed 17 April 2017).

Alderson, J.C. 2000. *Assessing reading*. Cambridge: Cambridge University Press.

Bachman, L.F. 1985. Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly* 19(3): 535-556.

Benjamin, R.G. 2012. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review* 24(1): 63-88.

Bormuth, J.R. 1967. *Cloze readability procedure*. CESIP Occasional report 1. Los Angeles, CA: University of California.

Bormuth, J.R. 1968. Cloze test readability: Criterion reference scores. *Journal of Educational Measurement* 5(3): 189-196.

Dreyer, L.G. 1984. Readability and responsibility. *Journal of Reading* 27(4): 334-338.

DuBay, W.H. 2004. *The principles of readability*. Available online: http://www.impact-information.com/impactinfo/readability02.pdf (Accessed 17 April 2017).

Duffelmeyer, F.A. 1985. Estimating readability with a computer: Beware of the aura of precision. *The Reading Teacher* 38(4): 392-394.

Field, A. 2009. *Discovering statistics using SPSS*. Los Angeles: Sage.

Flesch, R. 1948. A new readability yardstick. *Journal of Applied Psychology* 32(3): 221-233.

Fry, E. 1989. Reading formulas: Maligned but valid. *Journal of Reading* 32(4): 292-297.

Fry, E. 2002. Readability versus levelling. *The Reading Teacher* 56(3): 286-291.

Gellert, A.S. and C. Elbro. 2013. Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment* 31(1): 16-28.

---

[20]   For a plea in defense of the advantages of readability formulas, if wisely used, see Fry (1989).

Goodman, K. 1967. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist* 6(4): 126-135.

Horton, R.J. 1974-1975. The construct validity of cloze procedure: An exploratory factor analysis of cloze, paragraph reading, and structure-of-intellect tests (abstract). *Reading Research Quarterly* 10(2): 248-250.

Jansen, C. and N. Boersma. 2013. Meten is weten? Over de waarde van de leesbaarheidsvoorspellingen van drie geautomatiseerde Nederlandse meetinstrumenten. *Tijdschrift voor Taalbeheersing* 35(1): 47-62.

Jansen, C. and L. Lentz. 2008. Hoe begrijpelijk is mijn tekst? De opkomst, neergang en terugkeer van de leesbaarheidsformules. *Onze Taal* 77(1): 4-7.

Jones, M.J. 1997. Critical appraisal of the cloze procedure's use in the accounting domain. *Accounting, Auditing & Accountability Journal* 10(1): 105-128.

Jongsma, E.R. 1971. *The cloze procedure: A survey of the research*. Indiana University: Bloomington (IN).

Juicy Studio. 2017. Readability test. Available online: http://juicystudio.com/services/readability.php (Accessed 17 April 2017).

Kamalski, J.M.H. 2007. Coherence marking, comprehension and persuasion: On the processing and representation of discourse. Unpublished PhD dissertation, Utrecht University, Utrecht, The Netherlands.

Klare, G.R. 1974-1975. Assessing readability. *Reading Research Quarterly* 10(1): 62-102.

Klare, G.R. 1988. The formative years. In B.L. Zakaluk and S.J. Samuels (eds.) *Readability: Its past, present, and future*. Newark, DE: International Reading Association. pp. 14-34.

Kobayashi, M. 2002. Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal* 86(4): 571-586.

Litz, D. and A. Smith. 2006. Semantically acceptable scoring procedures (SEMAC) versus exact replacement scoring methods (ERS) for 'cloze tests': A case study. *Asian EFL Journal Quarterly* 8(1): 41-68.

Lively, B.A. and S.L. Pressey. 1923. A method of measuring vocabulary burden of textbooks. *Educational Administration and Supervision* 9(7): 389-398.

McDermid Heyns, J. 2007. *Readability statistics for Afrikaans*. Paper read at the LSSA/SAALT/SAALA Joint Annual Conference, 4 July, North-West University, Potchefstroom, South Africa.

McKellar, C.A. 2008. *A readability formula for Afrikaans*. Poster presented at PRASA, 27-28 November, Cape Town, South Africa.

McKenna, M. 1976. Synonymic versus verbatim scoring of the cloze procedure. *Journal of Reading* 20(2): 141-143.

North-West University. 2017. Centre for Text Technology: Overview. Available online: http://humanities.nwu.ac.za/ctext (Accessed 17 April 2017).

NWO (Netherlands Organisation for Scientific Research). 2017. LIN: A validated reading level tool for Dutch. Available online: https://www.nwo.nl/en/research-and-results/research-projects/i/56/8956.html (Accessed 17 April 2017).

O'Toole, J.M. and R.A.R. King. 2010. A matter of significance. Can sampling error invalidate cloze estimates of text readability? *Language Assessment Quarterly* 7(4): 303-316.

O'Toole, J.M. and R.A.R. King. 2011. The deceptive mean: Conceptual scoring of cloze entries differentially advantages more able readers. *Language Testing* 28(1): 127-144.

Pienaar, M. 2009. *Die leesbaarheid van akademiese tekste: 'n Tekslinguistiese ondersoek.* Unpublished master's thesis, North-West University, Potchefstroom, South Africa.

Rabin, A.T. 1988. Determining difficulty levels of text written in languages other than English. In B.L. Zakaluk and S.J. Samuels (eds.) *Readability: Its past, present, and future.* Newark, DE: International Reading Association. pp. 46-76.

Readability formulas. N.d. Automatic readability checker. Available online: http://www.readabilityformulas.com/free-readability-formula-tests.php (Accessed 17 April 2017).

Schriver, K. 2000. Readability formulas in the new millennium: What's the use*? ACM Journal of Computer Documentation* 24(3): 138-140.

Singer, H. 1988. Foreword. In B.L. Zakaluk and S.J. Samuels (eds.) *Readability: Its past, present, and future*. Newark, DE: International Reading Association. pp. vii-ix.

Skryfgoed. 2008. Afrikaans grammar checker. Potchefstroom, South Africa: Centre for Text Technology.

Skryfgoed. 2017. Afrikaans grammar checker. 5th ed. Potchefstroom, South Africa: Centre for Text Technology. Available online: https://spel.co.za/en/product/afrikaanse-skryfgoed-5-2/ (Accessed 17 April 2017).

Taylor, W.L. 1953. 'Cloze Procedure'. A new tool for measuring readability. *Journalism Quarterly* 30(3): 415- 433.

Van Rooyen, R. 1986. Eerste Afrikaanse leesbaarheidsformules. *Communicatio: South African Journal for Communication Theory and Research* 12(1): 59-69.

Watanabe, Y. and D. Koyama. 2008. A meta-analysis of second language cloze testing research. *Second Language Studies* 26(2): 103-133.

Appendix 1. Text titles, number of words, RE-scores and cloze test scores

Text title	Number of words	RE-Van Rooyen-MC	RE-Van Rooyen-CL	RE-McDermid Heyns	RE-McKellar	Average cloze score
Ek ly onder my pa se leuens (*Huisgenoot*, 13-11-2014)	299	66.0578	14.0337	57.1227	64.8002	57.7700
'n Genie vol grille en geite (*Huisgenoot*, 3-11-2014)	302	79.7175	31.2863	63.9573	69.4749	46.9800
Vat vyf vir versoening! (*Huisgenoot*, 13-11-2014)	299	67.8638	30.3198	63.6587	72.9366	52.8200
Boere laat my weer in liefde glo (*Huisgenoot*, 13 -11-2014)	289	65.6138	51.2001	74.3631	82.2170	44.8300
Drama agter die skerms (*Huisgenoot*, 13-11-2014)	294	78.3179	34.6916	69.5144	74.8426	49.6800
Stellenbosch rioolsuiweringswerke brei uit (Stellenbosch munisipaliteit nuusbrief, Uitgawe 3: 2013)	305	76.6216	16.3768	51.4059	51.0935	51.5500
Boodskap van die Burgemeester, Helen Zille (Kaapstad 2007–2012, *Vyf jaar plan*)	311	87.5143	14.7882	49.7354	54.9957	39.5300
Budget 2014: Die publiek se gids (National Treasury)	292	66.0518	9.1839	40.0897	35.3946	39.8200
Kaapstad 2007–2012 (Stad Kaapstad, GOP en Prestasiebestuurskantoor)	284	85.8242	12.0045	47.4450	47.5282	37.8200
Meer inligting oor elektrisiteitstariewe (Swartland munisipale nuusbrief, No. 10)	304	65.6616	8.1873	33.6674	30.1779	34.1100