Linguistic Research 35(Special Edition), 137-170
DOI: 10.17250/khisli.35..201809.006

Assessing the readability of fiction: A corpus analysis and readability ranking of 200 English fiction texts*

Michael Dalvean · Galbadrakh Enkhbayar*****
(Canberra College · Southern Taiwan University of Science and Technology)

Dalvean, Michael and Galbadrakh Enkhbayar. 2018. Assessing the readability of fiction: A corpus analysis and readability ranking of 200 English fiction texts. *Linguistic Research* 35(Special Edition), 137-170. Standard readability measures are based on the readability of non-fiction texts. This means that the validity of the measures when applied to fiction texts is questionable. Thus, the scores given to fiction texts using such indices may be invalid when used by English teachers to identify fiction texts of appropriate difficulty for students with various reading ability levels. This paper attempts to address this problem by 1) developing a readability measure specifically designed for fiction texts and 2) applying it to 200 English fiction texts. A corpus, consisting of 100 adults' and 100 children's texts, is used for the analysis. In the initial modeling, several standard readability measures are used as variables, and machine learning is used to create a classifier which is able to classify the corpus with an accuracy of 84%. A second classifier is then created using linguistic variables rather than standard readability measures. The latter classifier is able to classify the corpus with an accuracy of 89%, indicating that the standard readability measures are less accurate in classifying fiction texts than linguistic variables. Due to its higher accuracy, the latter classifier is then used to provide a linear complexity or 'readability' rank for each text. The ranking using the linguistic-based classifier provides an more accurate method of determining which texts to choose for students according to their reading levels than the standard readability measures. Importantly, the ranking instantiates a fine-grained increase in complexity. This means that the ranking can be used by an English teacher to select a sequence of texts that represent an increasing challenge to a student without there being a frustratingly discrete rise in difficulty. (**Canberra College · Southern Taiwan University of Science and Technology)**

Keywords readability, text complexity, fiction, machine learning, computational linguistics, corpus linguistics

---

* This research was completed while the first author was a lecturer in English as a Second Language, and the second author was an undergraduate student, at Global Leadership University, Ulaanbaatar, Mongolia. The authors would like to thank the anonymous reviewers of *Linguistic Research* for their helpful comments and suggestions. All remaining errors are, of course, the authors' own.

** First and corresponding author
*** Second author

1. Introduction

There are two basic purposes of this paper. The first is to create a method for determining the readability of fiction texts. This is something that has not been a focus of the formulas that are used in readability analysis; their focus is non-fiction texts. The second purpose is to use the method to rank a corpus of 200 English novels. The idea here is that such a ranking can be of use to English teachers in that it can help them determine 1) the appropriate text to prescribe for a student at a given level of English reading and 2) a subsequent series of texts that are of gradually increasing complexity.

Determining the readability of fiction texts has not been considered as important as determining the readability of non-fiction texts. Readability Studio (2015), a professional readability program, provides a suite of 34 readability measures – all of which were designed for, or calibrated on, non-fiction texts. Standard practice is to use such measures to determine the readability of both fiction and non-fiction texts. Whether measures calibrated on non-fiction texts can be used to validly measure the readability of fiction texts is an important question. A salient element of this paper involves showing that a readability measure specifically calibrated on a corpus of fiction texts can outperform the standard readability metrics. Whether this is because there is a fundamental difference in the nature of fiction and non-fiction which is not picked up by the traditional readability measures is an interesting question, which could be fruitfully addressed in subsequent research.

The relative neglect of fiction texts in the creation of readability measures belies the value of fiction in language development. The evidence shows that reading fiction texts has a greater positive impact on verbal ability than reading non-fiction (Mar and Rain 2015). This is not because there is any measurable difference in the range of vocabulary (McCreath et al 2017). One possible reason is that there is an emotional element involved with fiction (ibid) and learning takes place more effectively with such associations (Hascher 2010). Along with the commonly accepted benefits of reading, there is also a benefit associated with writing: Douglas and Miller (2016) found that the sophistication of reading level in business students was positively associated with the level of sophistication in their writing. Kim (2005) found that a similar synergy between

reading and writing exists in an L2 setting. Finally, the benefits for vocabulary growth of reading cannot be overestimated. For a systematic analysis of the benefits of reading for vocabulary growth, see Nation (2015).

With these observations in mind, it seems obvious that an English teacher needs to promote as much reading as possible, and that, if the aim is language development, promoting the reading of fiction is preferable to promoting the reading of nonfiction. However, how does the English teacher choose the appropriate text for a student? If the text is too difficult, the student will become frustrated (Feng et al 2013). The student should ideally be presented with new words, but not to the point that their enjoyment of the text is hampered by a lack of comprehension. The idea here is that a student needs to know at least 95%-98% of the words in a text in order to comprehend it (Laufer and Ravenhorst-Kalovski 2010). Furthermore, even if an English teacher is able to find the ideal text in terms of the student's ability, what text should she provide the student with after the student has read that text? If the next text presents a level of difficulty that is significantly above the first text, then the frustration problem emerges. Frustration can lead to lowering of self-confidence, which has been shown to be linked to demotivation in language learning (Shin and Kim 2017). However, if the second text is simpler than the initial text, then there is a possibility that the student will not be learning new vocabulary.

A graded series of texts solves this problem. An English teacher can quickly determine the level at which a student is reading by getting them to read a section of the text. If the student reports that she does not understand the text, the teacher can select a text that is further towards the less complex end of the spectrum. By this means, a text at the ideal reading level can be found. This text is read, and the next text to be read is the text that is next highest in complexity. This process continues such that the student is reading texts of increasing complexity, but the gradations are such that the student does not perceive that each text is particularly harder. Furthermore, due to the repetition of new vocabulary at given and closely contiguous levels, much of the new vocabulary can be learned in context. That is, the new vocabulary can be learned without frequent recourse to a dictionary, which is similar to the way in which native speakers learn new words.

An important outcome of this paper is a ranking of 200 fiction texts

according to their reading complexity. The analysis falls into the area of readability research which has been an important issue since Flesch (1948) produced the Flesch measure of readability. Numerous methods have been subsequently derived, all of which have various advantages and disadvantages. However, the important issue here is that most measures were designed for nonfiction. Furthermore, the measures are not accurate enough to provide a fine-grained gradation between texts of differing complexity. Both of these problems are addressed in this study.

It should be noted that the term 'English teacher' is used to describe teachers of English to both native and non-native (ESL) students. The reason for this is that the concept of the 'difficulty' of a text is very similar in both contexts. In recent years, many publishers have begun including in their descriptions of children's books some index, or indices, of readability. To some extent, these measures are associated with standard grade levels. However, even within grades, there is a great deal of variation in the difficulty level of texts in standard reading schemes. Thus, the question as to what a student should read after they have read a given book is essentially the question of what level of complexity a student should be reading at after finishing a given text. The situation is very similar for the ESL teacher. As with a teacher of English to native speakers, the situation calls for a fine-grained gradation in texts so that the student can be presented with a sequence of texts that represents a slowly increasing level of complexity. Thus, theoretically, the role of the teacher of English to native and non-native students is similar. Empirical evidence that the role of English teacher's role is similar for both native and non-native students is provided by the fact that, for both native and non-native speakers, the vocabulary threshold for general comprehension of a text is 95% (van Zeeland and Schmitt 2012). This indicates that very similar processes are taking place in the cognitive processes of both the L1 and L2 English student. Thus, for the purposes of this paper, the general term 'English teacher' will be used to describe a teacher who needs to be able to accurately assess the complexity of an English text in an ESL or non-ESL setting.

2. Previous work in readability/text complexity

There has been a plethora of work in readability in the last century, with much of the work being conducted in the last 20-30 years due to the advent of easily available computing resources. As such, the following selection of previous work in the field is only a small sample. The studies have been selected on the basis that they represent a given type of or approach to the problem of reading complexity. For more comprehensive overview of the field in general see Collins-Thompson (2014).

Flesch's (1948) measure of 'Reading Ease was designed to place a given text on a scale of 0 -100, with 100 being very easy to read and 0 being very difficult. The formula is R:E: = 206:835 − 0:846wl − 1:015sl, where wl is the number of syllables per 100 words and sl is the average number of words per sentence. Variations on this formula are designed to place texts into US grade levels (Kincaid et al 1975). The various Flesch formulae are very prominent, and they have significantly influenced the field of readability analysis. The Fog index (Gunning 1968) is a similar measure, based on the formula Fog Index = 0.4 * (ASL + PHW), where ASL is the average sentence length and PHW is the number of words with more than two syllables.

The main problem with this kind of formula is that there is no account taken of the semantic element of text (Hartley 2016: 1524). Dale and Chall (1948) go some way towards addressing this issue by incorporating a measure of difficulty based on the extent to which the text contains words that are outside a list of 3000 common words. In a similar vein, Spache (1953) provides a formula for estimating the difficulty of early primary texts using the formula Index/Grade Level = (0.141 X ASL) + (0.086 X PDW) +0.839, where ASL is average sentence length and PDW is the percentage of difficult words based on Spache's word list.

In the last two decades, the Lexile (Stenner 2001) measure has become popular. This measure includes the syntactic measures based on sentence length and number of syllables, but also includes a measure based on the extent to which words in the text occur frequently in English. Such 'word frequency' measures have developed over several decades of research into corpus linguistics. The idea here is that those words that are encountered frequently are

more familiar, and texts which contain such words are likely to be easier to understand than texts with less frequently encountered words.

Ardoin et al (2010) provides an overview of the research into how accurate these standard measures are for predicting one proxy for text complexity – words read correctly per minute. The conclusion is that the standard readability formulae are not good predictors of reading difficulty.

The standard readability measures are based on a priori assessments of what makes text complex. In recent years the approach has been to use computational linguistics and machine learning to examine large corpora of texts to look for differences between texts assessed as being of different levels of difficulty. The procedure here is to take a corpus of texts that have been labeled with grade or age levels, break the texts down into various linguistic variables, and then use computer modeling methods to look for differences in how the linguistic variables vary as between grade/age levels. The important point here is that the more modern modeling methods are not based on a priori assumptions about text complexity; they are empirical in that a vast number of linguistic variables are included in the analysis and the final model is the result of the patterns that have been discovered in the data.

A good example of the machine learning approach is Xia, Kochmar and Briscoe (2016). This study used the WeeBit corpus of children's/teenagers' informational texts which has five age groupings from 7-8 to 14-16. A second corpus was derived from the Cambridge English Examinations in English as a Foreign Language. This system has texts from five different levels, which correspond to the Common European Framework of Reference (CEFR) language levels from A2 to C2. The study used a range of traditional readability formulas, measures of lexical variation and density, variables derived from the Academic Word List (Coxhead 2000), syntactic variables, and cohesion and coherence variables. Combining all these variables and using support vector machines as the modeling method, an accuracy of 80% in classifying the WeeBit corpus was achieved. Using similar variables, an accuracy of 78.5% was achieved on the Cambridge examination corpus.

The important point about this kind of approach is that it makes no claims about the likelihood of any given variable being related to the complexity of the text. A vast number of variables is included in the modeling procedure and the

machine learning algorithm is used to determine the complexity of the a given text based on the patterns in the data. It is this approach that is used in the current study.

3. Materials and methods

3.1 The corpus

There are many studies that have applied readability assessment methods to fiction. However, most of the measures used were initially calibrated on nonfiction and were then applied to fiction. There is very little research that has been devoted to specifically designing metrics to assess the reading complexity of fiction. The current study differs from most in that it focuses entirely on designing a measure for fiction complexity. For this reason, the modeling corpus consists entirely of fiction texts.

A second major respect in which the current study differs from the mainstream approach is that it does not attempt to categorise texts as being associated with a given grade or age level. Instead, it attempts to provide a relative measure of complexity. This is by no means unique: Flesch (1948) was one of the early methods which assigned a similar measure of 'reading ease' between 0 and 100. There have been several others that do not associated text difficulty with age or grade level. However, most of the latest methods using computational linguistics and machine learning use either age or grade levels as the dependent variable in the creation of models designed to classify texts by complexity. The problem here is that it is difficult to assign an age or grade level to many fiction texts. Certainly the works of Enid Blyton might be confined to readers of primary school age. However, the ostensibly 'primary' text, Carroll's *The Adventures of Alice in Wonderland,* can be read by children, but it can also be enjoyed by older children and adults who can enjoy the author's mathematical and philosophical allusions. Similarly, Graeme's *The Wind in the Willows* has social and political elements which make it enjoyable for those beyond 'primary' level. In short, many primary-level books can be understood by primary level children, but to assign a specific age or grade is difficult. This

is not the case with a history text written specifically for year 6 students.

The approach taken in this study is not to attempt to assign an age or grade to texts but to select one set of texts that are generally considered to be able to be comprehended by primary level children and another that are considered to be beyond the comprehension level of primary children. Thus, the corpus is a binary design, consisting of a group of 100 typically primary school-level texts and a group of 100 typically 'adult-level' texts. The idea here is that there should be a sufficiently large gap between the two groups in terms of complexity that the modeling procedure will detect a strong signal. As the modeling procedure demonstrates, this intuition is supported in that the model is able to correctly classify the 200 texts with an accuracy of 89%, demonstrating that there is a significant difference between the two groups of texts which, it can be argued, represents a difference in complexity.

The primary school corpus was selected on the basis of the appearance of constituent books on reading lists of primary schools in the UK, the US and Australia for students in year 3 and above. Such lists typically include books that are traditionally read by children, such as Grimm's fairy tales, and classic books written specifically for children such as Enid Blyton's *Famous Five* series, as well as modern children's books such as the early Harry Potter books. Where possible, the suggested age range of books was checked with the current publisher. Any book that had an age range recommendation where the lower limit of the age range was greater than 12 was excluded. The idea here is that there should be a general impression on the part of schools and publishers that the book can be understood by primary school students. One caveat in relation to the selection of texts is that the final selection consists of only those for which an extract was available in digital form.

The adult group was selected by consulting college, senior high school and advanced placement reading lists as well as publishers' websites. Any book that was recommended for readers for young adults was excluded. Where a specifically mentioned work by an author was not available in digital form, another work by the same author was sought.

The extracts for books sourced from such repositories as the Gutenberg project (www.gutenberg.com) were significantly larger than the extracts from the later works, largely due to copyright considerations. For most of the books, the

extract used consists of one chapter. In some cases this was not possible. However, for all cases the extract was at least 300 words (min = 377; max =26389; mean= 3147; stdev=3183) which is above the standard cut-off of 200 words for validity of many readability formulas. The actual word count for each extract is given in Appendix 2.

Thus, the corpus consists of two groups of texts which, theoretically, are separated by a 'sophistication gap'. Interestingly, despite this gap, the traditional measures of readability indicate that there is not a great difference in difficulty as between the two groups. Using an average grade level derived from 29 standard reading measures (Readability Studio 2015) the average grade level of the primary texts is 5.7 while that for the post-primary texts is 7.6 ($p<.001$; two tailed test). This indicates that on the basis of standard readability measures, the books read by adults are less than two years ahead of those read by primary school children. However, as we shall see, there is a great deal of variation in the actual complexity of texts. Thus, while it is some books often read by adults are quite simple, others are quite complex.

3.2 The variables

The dependent variable was coded as 0 for those texts that were in the primary category and 1 for those that were in the post-primary category.

The independent variables included traditional readability metrics, syntactic variables and semantic variables. Thirty-four standard readability metrics were derived from a professional readability program, Readability Studio (2015). However, several of the metrics were not appropriate because they provided a range for texts rather than a discrete number. Several metrics were discarded because they provided an open-ended category for some levels (eg: 19 years +). Other tests were discarded because they could not generate a valid score for all texts. Twenty-four standard readability metrics remained for use in the modeling..

Several other measures were derived from the Readability Studio output. These were a measure of the rate of use of the passive voice, the rate of use of of long (6+ characters) words, the rate of use of SMOG hard words, the rate of

use of of Fog hard words, the rate of use of Dale-Chall unfamiliar words, the rate of use of Spache unfamiliar words, and the rate of use of Harris-Jacobson unfamiliar words. Readability Studio also produces a mean grade score for each text based on the scores generated from the individual readability metrics. This variable was also included in the modeling.

A number of psycholinguistic variables were generated from the Recursive Inspection of Text (RiotScan) software (Boyd 2013). These included measures of linguistic abstraction based on Mergenthaler's (1996) method, and concreteness, age of acquisition, imageability, familiarity and uncertainty based on the Gilhoolie and Logie (1980) word norms. RiotScan also provided ten syntactic and associated variables based on word and sentence lengths as well as word uniqueness.

Thus, there were 58 readability, psycholinguistic, syntactic, and linguistic variables included in the analysis. Details of the variables are provided in Appendix 1.

3.3 Modeling

Modeling was done with WEKA (Frank et al 2016). The modeling method used was logistic regression. The reason for this was that we need to have an ordering of texts that represents a continuous gradation. If we were to use a non-linear method, such as support vector machines or neural networks, the actual binary classification accuracy might be higher than that achieved with logistic regression, However, it is unlikely that the scores generated using such methods would yield a true linear gradation of complexity for all variables. The reason for this is that such methods are able to model highly non-linear phenomena. While this is a strength for classification *per se*, it is a disadvantage if we want genuinely linear gradations in all variables. To put this in practical terms, it is quite likely that a non-linear induction method could place two texts closely together in terms of their dependent value 'score' but that discontinuities occur in the constituent independent variables. In such a situation a student might, for example, go from a text with a low level of complex words to a text with a higher level of complex words while the syntactic complexity goes in the other direction. Certainly the overall complexity will have risen only

incrementally, but the point is that the student will be faced with an increase in the complexity of vocabulary that is disconcerting. We want as smooth as possible a transition from one text to another on all variables so that differences in complexity are not overly perceptible.

The first stage in the modeling procedure was to see how effective the existing readability metrics are at distinguishing between primary and post-primary texts. This process involved creating a univariate logistic model for each of the standard 24 readability metrics and a further univariate model using a variable derived from the grade means of each text generated by Readability Studio. A further logistic model was created using all 25 variables. Finally, the Simple Logistic module in WEKA was used to create a model using variable selection. This modeling method uses LogitBoost (Landwehr et al 2005) to select the most predictive variables. Variables are added until the cross validated accuracy of the model ceases to fall. The classification accuracy and Root Mean Squared Error based on leave-one-out cross validation (n = 200) for these 26 models are presented in Table 1.

Table 1. Accuracy data for modeling with readability metrics

Model	**Acc′ %**	**RMSE**
Bormuth Cloze Mean	70.5	0.4269
Bormuth Grade Placement	71.0	0.4262
Coleman-Liau (grade levels)	69.0	0.4362
Coleman-Liau (pred. Cloze scores)	71.5	0.4360
Danielson-Bryan 1	69.5	0.4551
Danielson-Bryan 2	69.0	0.4552
Degrees of Reading Power	70.5	0.4269
Degrees of Reading Power (GE)	67.5	0.4242
EFLAW	59.5	0.4835
Farr, Jenkins, Paterson	70.5	0.4457
Flesch Reading Ease	72.5	0.4236
FORCAST	70.5	0.4465
Gunning Fog	68.0	0.4519
LIX (index values)	70.0	0.4442
Modified SMOG	76.5	0.4132

New Fog Count	63.0	0.4843
PSK Dale-Chall	78.5	0.3843
PSK Farr, Jenkins, Paterson	65.0	0.4620
PSK Gunning Fog	69.5	0.4467
PSK Flesch	69.5	0.4362
RIX (index values)	72.5	0.4411
SMOG	77.5	0.4078
SMOG (simplified)	77.0	0.4096
Spache Revised	66.0	0.4494
Mean Gade Level	67.5	0.4415
All 25 Variables	83.5	0.3604
Simple Logistic	84.0	0.3220

The variable with the greatest classification accuracy in the univariate modeling was PSK Dale-Chall, which achieved a correct classification rate of 78.5% (RMSE = .3843). The highest accuracy overall was achieved with the Simple Logistic model, with an accuracy of 84% (RMSE = .3220). Variables and Coefficients for this model are presented in Table 2.

Table 2. Simple logistic model using readability metrics

Variable	**Coeff'**
Bormuth Cloze Mean	0.03
Coleman-Liau (pred. Cloze scores)	0.05
Danielson-Bryan 1	-0.35
Danielson-Bryan 2	0.04
EFLAW	-0.04
Farr, Jenkins, Paterson	0.03
New Fog Count	-0.06
PSK Dale-Chall	2.96
PSK Farr, Jenkins, Paterson	-0.69
SMOG	0.66
Constant	-24.89

The benchmark for subsequent modeling is thus 84%. That is, using machine learning, a model can be created using the existing readability metrics which are

able to classify the texts in the corpus with an accuracy of 84%. As we will see, it is possible to improve on this by adding linguistic and syntactic variables such that a model with a classification accuracy of 89% can be created.

This stage of the modeling procedure involved the creation of a logistic regression model with all 58 independent variables. Leave-one-out cross validation (n = 200) was used to test the model. This model achieved a-cross validated accuracy of 84% (RMSE = .3971). Subsequent models were created by excluding one group of variables and leaving the others in. Table 3 shows the cross-validated accuracy of the models thus created.

Table 3. Accuracy metrics for initial modeling

Model #	**Variables Excluded**	**No. of Variables remaining**	**Acc (%)**	**RMS Error**
1	n/a	58	84.00	0.3971
2	Syntactic	45	80.50	0.4414
3	Word List/Grammatical	54	85.00	0.3867
4	Readability Metrics	34	85.50	0.3755

Given that the fourth model had the highest accuracy, albeit by a small margin, this model was chosen for further modeling.

Using the Simple Logistic module in WEKA with the 34 variables in model 4, the result is a logistic model of 9 variables. The cross-validated classification accuracy of this model was 89%, with an RMS error of .2854. The sensitivity (ability of the model to identify complex texts) was 86%, while the specificity (ability of the model to identify simple texts) was 92% (Kappa = .78, z =11.05, $p<.001$). Thus, the model is able to efficiently classify the texts in the corpus. Variables and coefficients for this model are provided in Table 4.

Table 4. Variables and coefficients of final model

Variable	Coefficient
Unique_Ratio	-0.061
Hapax_Ratio_Sent	-0.0282
AvgWordLengthPerSentence	-2.181
COVSentenceLength	1.3419
GLNorms_AOA_M	1.2958

GLNorms_Familiar_M	2.0285
GLNorms_Concrete_SDM	-1.8924
% of Fog hard words	0.1368
% of Dale-Chall unfamiliar words	0.3921
Constant	-5.0775

The equation represented in Table 4 is, in essence, a formula for determining the readability of a given text. Certainly, it contains more variables than the standard readability formulas. However, it is essentially very similar in nature in that it provides a means of determining the readability of a text according to a number of variables derived from the text. Thus, the first aim of this paper, that of developing a readability measure specifically designed for fiction texts, has been achieved.

Although it is not usual in a machine learning-oriented paper to provide a traditional analysis of the parameters of a logistic model, it is worthwhile making the observation that three of the variables in the model are the opposite in sign to that which we would expect from the raw data. AvgWordLengthPerSentence and GLNorms_Concrete_SDM have positive associations with the dependent variable on a univariate basis, while GLNorms_Familiar_M has a negative association. This indicates that there are likely to be some suppressor effects in the model (Ray-Mukherjee et al 2014). This is an interesting observation, and may be worthwhile examining in future research of a more theoretical nature. Another observation is that the number of variables is considerably less than the number of variables included in some large-scale machine learning studies while the classification accuracy is relatively good. The fact that a small number of variables can be found to be highly predictive of text complexity has been noted previously (Solnyshkina et al 2017).

These are interesting observations and are worthwhile considering in subsequent research. The important observation from the point of view of the current analysis is that the high level of classification accuracy indicates that the formula derived from the modeling is likely to provide a good indication of the relative difficulty of the texts in the corpus.

4. Discussion

The purposes of this paper have been to devise a method of determining the readability of a corpus of fiction texts and to provide a ranking of the corpus based on the readability method. The readability measure is clearly effective: the classification accuracy of 89% suggests that a significant element of what constitutes complexity has been captured by the modeling. Furthermore, the use of a linear modeling method means that it is reasonable to assume that texts ranked on the basis of their score should reflect differences in complexity on a continuous scale. In this section we will consider some of the important implications of the ranking.

Appendix 2 contains the ranking by complexity of 200 texts. The scores are standardized so that the average is represented by 0 with a standard deviation of 1. Negative scores represent texts of lower than average complexity and positive scores represent texts of higher than average complexity. The least complex text in the corpus is RL Stine's *Revenge of the Living Dummy* (-2.11), while the most complex is Edgar Allan Poe's *The Fall of the House of Usher* (2.80). The text closest to the average is Dorothy Parker's *The Waltz* (.01). It is interesting to note that the two stories at the extremes of the scale are from the horror genre, although the Stine text has elements of humour, which the Poe text decidedly does not.

To get an insight into the extent to which the model is able to discern the differences in complex and less complex text, it is worthwhile considering an extract from the text scoring the highest in complexity and the text scoring the lowest.

Following is a selection from the most complex text, Poe's *The Fall of the House of Usher*:

> There was an iciness, a sinking, a sickening of the heart—an unredeemed dreariness of thought which no goading of the imagination could torture into aught of the sublime. What was it—I paused to think—what was it that so unnerved me in the contemplation of the House of Usher? It was a mystery all insoluble; nor could I grapple with the shadowy fancies that crowded

> upon me as I pondered. I was forced to fall back upon the unsatisfactory conclusion, that while, beyond doubt, there are combinations of very simple natural objects which have the power of thus affecting us, still the analysis of this power lies among considerations beyond our depth (Poe, 2015 [1839]: 95).

Following is a selection from the least complex text, Stine's *Revenge of the Living Dummy*:

> You may wonder why my best friend, Molly and I were in the old graveyard late at night.
> I shivered as I thought about what we were doing. Wind howled through the trees, and pale streaks of lightning cracked the sky.
> "Hurry, Molly," I whispered, hugging myself as the moon disappeared behind the clouds. "It's going to storm."
> "I am hurrying, Britney," Molly said. "But the ground . . . it's really hard."
> We were digging a grave. We took turns. One of us shoveled while the other stood lookout. I felt cold raindrops on my forehead. I kept my eyes on the low picket fence near the street. Nothing moved. The only sounds were the scrape of the shovel in the dirt and a drumroll of thunder, deep but far away.
> Across from me, an old gravestone made a creaking sound as it tilted in the wind (Stine, 2008: 143).

Both are excellent examples of writing in that each author has directed his writing at his target audience. The complexity of Poe's writing is due to his desire to depict not only his narrator's mental landscape but also that of the highly complex central character, Roderick Usher, via the narrator's mind. This requires depicting high orders of intentionality. Creating and maintaining high-order intentional constructs have been found to be highly cognitively demanding (Powell et al 2010). It is therefore to be expected that Poe's writing should be complex. Stine, on the other hand, has deftly aimed his writing at primary school students and as such uses simple, direct, concrete language.

The direct application of this information is that an English teacher

struggling to find a text that is accessible to a student who finds Harry Potter difficult to read might recommend a text by Stine. On the other hand, a student for whom Boccaccio's *The Decameron* is easily comprehended might be encouraged to try Poe.

An interesting issue as we trace the increase in complexity in the ranking is that we find that a famously 'complex' novel, Faulkner's *As I lay Dying,* scores the lowest out of all the post-primary texts. At -.6, it is within one standard deviation of the mean and is therefore in the average range of complexity. Initially, this seems odd given its reputation as a difficult novel. However, if we consider the difficulty of the text, as opposed to the structure of the novel, it becomes clear that the text is quite simple. The extract used in the modeling depicts the impressions of three of the characters, all poor rural Southerners, as they each individually describe the same scene. The language is matter-of-fact and direct. Interestingly, it is in the present tense. In short, the text itself is not complex. This is not to say that the structure of the novel does not have its difficulties.

At the other end of the spectrum, two 'primary' texts' RM Ballantyne's *A Coral Island* and Kenneth Graeme's *Wind in the Willows* score at .14 and .13 respectively, putting them above the score of texts by such noted authors as Camus, Plath, Flaubert, Lessing and Parker. Certainly it is true that all these texts are well within the average range of 0 +/- 1 and are therefore within one standard deviation of the average. However, it does seem incongruous that two ostensibly 'primary' texts score higher than some canonical works of literature. The explanation is that these works use a level of language sophistication that belies their status as primary-level texts. The subject matter of the text may seem at the primary level. However, a cursory glance at either of these texts shows that they are written in a style that is not characteristic of modern primary-level texts. One of the benefits of the ranking is that an English teacher can avoid the mistake of assigning such texts on the ground that they are primary texts to students who have lower levels of reading skills and, instead, select texts that are substantially easier according to the ranking.

In relation to the 'average' level of complexity for some canonical works of literature, it is worth mentioning that the scale represents text complexity, not necessarily literary value or worth. It is quite possible that the reason for the

effectiveness of some writers is that they are able to express complex ideas in a simple way. This can explain how canonical works of literature can score below some primary-level books on a raw measure of complexity. Consider, for example, that Golding's *Lord of the Flies* has precisely the same complexity level of the primary level novel of which it is a parody, *A Coral Island*. Thus, Golding has mimicked the language level of *The Coral Island* while expressing sentiments that are the polar opposite of the sentiments expressed therein. This is an outstanding feat of craftsmanship. This idea, that artistic greatness may be associated with a moderate level of complexity, has also been noted in the fields of music and poetry (Dalvean 2016).

The important point here is that the actual level of complexity of a given text is difficult to assess subjectively. We tend to assume that, because a text is associated with the primary level that it will be simple. But the analysis shows that there is a great deal of variation in the complexity of supposedly simple books.

The important point is that the English teacher can benefit from the ranking by using it to find the text level at which a given student is reading and progressively select the next most complex text as the student finishes each text. At no stage is there a danger that the student will be confronted with a text that is significantly beyond them yet the teacher can be confident that they are learning to comprehend increasingly complex text. This scenario is made possible by the fact that the gradation from one text to another is fine enough to avoid significant jumps in complexity between texts.

5. Applications in the classroom setting

5.1 Selecting the starting text

An English teacher can be of great assistance to a student if s/he can match a text with a student's ability level. Research indicates that the ideal text level for a reader, the lexical threshold, is a text where between 95% and 98% of the words are known (Schmitt et al 2017). What this suggests is that the teacher should attempt to determine the student's match with a given text by

determining whether the student knows at least 95%-98% of a text. The ranking can be used to determine the text which provides the student with this vocabulary level.

The problem, however, is that the standard way of matching a student with a text of given vocabulary level requires that we know the vocabulary levels of both the student and the text. A more direct procedure is to get the student to read a given selection of a text and underline each instance of each word they do not know. The student's reading vocabulary for the given text is measured by taking the raw number of unknown words and dividing by the total number of words. Importantly, each instance of an unknown word should be noted. The reason for this is that repeated words will be factored into the calculation by dividing by the total number of words in the extract. The following extract demonstrates the correct way of marking the text by a student who knows all the words except 'colt(s)' and 'cart'

> I wish you to pay attention to what I am going to say to you. The colts who live here are very good colts, but they are cart-horse colts, and of course they have not learned manners. You have been well-bred and well-born; your father has a great name in these parts, and your grandfather won the cup two years at the Newmarket races; your grandmother had the sweetest temper of any horse I ever knew, and I think you have never seen me kick or bite. I hope you will grow up gentle and good, and never learn bad ways; do your work with a good will, lift your feet up well when you trot, and never bite or kick even in play (Sewell 1999 [1877]: 1).

The total number of words in the extract is 124, and the number of underlined words is 4. Thus, the percentage of unknown words is 4/124 = 3.22%. The percentage of known words is therefore 100 − 3.22 = 96.78. Thus, this text is in the ideal range of 95% - 98% of words known by this student.

It should be noted that the word 'cart' in 'cart-horse' is underlined rather than 'cart-horse' as a whole, because the student knows the meaning of 'horse'. However, if the student did not know the meaning of the word 'horse' then each instance of the word 'horse' would also need to be underlined, and the hyphenated 'cart-horse' would need to be counted as two words which the

student did not know. In general, students may need several attempts at this procedure in order to generate a score which can be used by the teacher to determine the student's lexical level.

Having determined the student's lexical level, the next stage is to ensure that the student is reading a text that is within his/her lexical threshold. In the above case, the text is ideal. However, if the rate of unknown words had been above 5%, the teacher would be warranted in choosing a text from the list that ranks lower in complexity. Then, the above procedure of underlining unknown words would be repeated with the simpler text. If the student understands 95% - 98% of the simpler text, then that text is the one that the student should be reading at that point in the development of his/her reading skill.

A rate of unknown words lower than 98% may be an indication that a harder text is required because this would provide an opportunity for vocabulary development. However, this is dependent on the student's frustration threshold; a student with a high frustration threshold is more willing to take on the challenge while a low frustration threshold indicates a text with a reading vocabulary of significantly less than 2% may be in order.

5.2 Choosing subsequent texts

The above procedure demonstrates how the ranking can be used by an English teacher to select the ideal starting text for a student with a given reading skill level embarking on a reading program. The next issue to be addressed is the selection of subsequent texts. Here, the issue is quite straightforward in that the teacher needs simply to select the text which appears next to the starting text at the higher level of difficulty. It should be noted that the difference between most of the texts in the ranking is quite small, and this means that a student who is presented with a text that is not interesting to him/her could potentially skip that book and go on to the next text in the ranking. Consider, for example, a student who enjoyed William Alexander's *Ghoulish Song* (score = 1.29). The next text in the ranking is Beverley Cleary's *Ramona and Her Father* (Score = -1.28). If the student did not find this text interesting, it would be quite possible for the student to skip it and go to the next text in the ranking, Iain Lawrence's *Lord of*

the Nutcracker Men (score = -1.27). A caveat which should be noted is that large jumps, in which several novels are skipped, could result in students being presented with texts beyond their lexical level. However, as will be discussed below, it may be that the problems presented by a significant jump in complexity can be overcome by an increase in interest and motivation.

5.3 Interest and motivation

An issue that is not specifically addressed in this paper is the extent to which a given text appeals to a student such that that they are motivated to persevere with it. It is clear that a student who is interested in the subject matter of a text will be more likely to put the required effort into completing it. The question is, however, how can we determine whether a student will be interested in a given text? In the current study, the ranking of texts in terms of readability may need to be supplemented by the English teacher's judgment about the extent to which a given text is likely to be of interest to a student. Thus, it may be appropriate to skip some texts in the ranking, with the caveat that skipping too many contiguous texts may lead to a significant increase in difficulty between texts that are presented to the student. However, what is gained in interest could compensate for a significant increase in difficulty.

In subsequent research it may be worthwhile to determine what factors can contribute to the likelihood of students reading for pleasure. Certainly, the avoidance of frustration is an important element of reading for pleasure, and this is the basic idea behind determining the readability of a text. However, if the subject matter of the text is not intrinsically interesting to a student, it may be that readability on its own cannot provide sufficient motivation for the student. In one study of the interestingness of children's reading material, the interestingness of sentences accounted for thirty times as much variance in sentence recall as readability (Anderson et al 1987).

To increase the likelihood that students find texts interesting, students could be matched to texts on the basis of their preferences derived from surveys, interviews and reflective writing assignments. Methods of doing this could be an interesting extension of the current work in future research.

5.4 Appropriateness

It is understood that the kind of ranking produced in this paper may be used to select texts for students of school age. For this reason it is important that appropriateness as well as complexity be taken into account. The listing provides an indication as to whether each text is from the primary group of texts. It is recommended that, for texts that are not from the primary group, the teacher research whether the text is appropriate for the age range of the student cohort under consideration. The ultimate decision about the appropriateness of a text should be determined by school policy and procedure. However, there may be situations in which the school has not had the opportunity to make a decision about a particular book. In such situations it is worthwhile trying to determine if a given book is on a 'banned books' list such as the annual 'Banned Books' list produced by the American Library Association (www.alair.ala.org). Such lists are often designed to draw attention to issues associated with intellectual freedom. However, the lists usually provide a comprehensive account of why there were calls for the banning of the book, and this kind of information can be used by an English teacher to determine the appropriateness of a text for a given student or cohort.

6. Conclusion

In this paper we have demonstrated a method of scoring fiction texts by reading complexity. We examined a number of syntactic, linguistic and reading complexity variables for their ability to classify a corpus of 200 literary texts made up of 100 primary and 100 post-primary texts. The modeling procedure demonstrated that the syntactic and linguistic/grammatical features were more predictive than the standard readability measures. The final model consisted of one syntactic variable and five linguistic variables, and was able to correctly classify 89% of the 200 texts in the corpus. We used the linear scores generated by the model to rank the 200 texts. The result is a ranking of literary texts by reading complexity.

We went on to describe how such a ranking could be effectively used by an

English teacher. We described how a given student's reading vocabulary could be determined and how to use the text ranking to select a text appropriate for a student's reading vocabulary. Finally, provided ways in which an English teacher could use the ranking in a practical classroom setting.

We contend that the information presented in this study is likely to be of immediate practical benefit to the practicing English teacher.

References

Anderson, Jonathan. 1983. Lix and rix: Variations on a little-known readability index. *Journal of Reading* 26(6): 490-496.

Anderson, Richard, Larry Shirey, Paul Wilson, and Linda Fielding. 1987. Interestingness of children's reading material. In Richard Snow and Michael Farr (eds.), *Aptitude, learning and instruction: 3. Cognitive and affective process analyses*, 287-299. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ardoin, Scott, Jessica Williams, Theodore Christ, Cynthia Klubnik, and Claire Wellborn. 2010. Examining readability estimates' predictions of students' oral reading rate: Spache, lexile, and forcast. *School Psychology Review* 39(2): 277-285.

Björnsson, Carl-Hugo. 1983. Readability of newspapers in 11 languages. *Reading Research Quarterly* 18(4): 480-497.

Bormuth, John. 1966. Readability: A new approach. *Reading Research Quarterly* 1(3): 79-132.

Boyd, Ryan. 2013. RIOT Scan: Recursive Inspection of Text Scanner (Version .0.21) [Software]. [http://riot.ryanb.cc]. Accessed 18/02/2017.

Carver, Ronald. 1985. Is the degrees of reading power test valid or invalid? *Journal of Reading* 29(1): 34-41.

Coleman, Meri and Ta Lin Liau. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology* 60(2): 283-284.

Collins-Thompson, Kevyn. 2014. Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics* 165(2): 97-135.

Coxhead, Averil. 2000. A new academic word list. *TESOL Quarterly* 34(2): 213-238.

Dale, Edgar and Jeanne Chall. 1948. A formula for predicting readability: *Educational Research Bulletin* 27: 37-54.

Dalvean, Michael. 2016. Ranking canonical English poems. *Empirical Studies of the Arts* 34(1): 103-125.

Danielson, Wayne and Sam Dunn Bryan. 1963. Computer automation of two readability formulas. *Journalism Quarterly* 40(2): 201-206.

Douglas, Yellowlees and Samantha Miller. 2016. Syntactic complexity of reading content directly impacts complexity of mature students' writing. *International Journal of Business Administration* 7(3): 71-80.

Fang, Irving. 1966. The "Easy listening formula". *Journal of Broadcasting and Electronic Media* 11(1): 63-68.

Farr, James, James Jenkins, and Donald Paterson. 1951. Simplification of Flesch reading ease formula. *Journal of Applied Psychology* 35(5): 333-357.

Feng, Shi, Sidney D'Mello, and Arthur Graesser. 2013. Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin and Review* 20(3): 586-592.

Flesch, Rudolph. 1948. A new readability yardstick. *Journal of Applied Psychology* 32(3): 221-231.

Flesch, Rudolph. 1979. *How to write plain English: A book for lawyers and consumers*. New York, NY: Harper and Row.

Frank, Eibe, Mark Hall, and Ian Witten. 2016. The WEKA workbench. online appendix for d*ata mining: practical machine learning tools and techniques*, Morgan Kaufmann, Fourth Edition, 2016. [https://www.cs.waikato.ac.nz/~ml/weka/]. Accessed 24/05/2017.

Gilhooly, Ken and Robert Logie. 1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods and Instrumentation* 12(4): 395-427.

Gunning, Robert. 1968. *The technique of clear writing* (revised edition). New York, NY: McGraw Hill.

Harris, Albert Josiah and Milton Jacobson. 1982. *Basic reading vocabularies*. New York, NY: Macmillan.

Hartley, James. 2016. Is time up for the Flesch measure of reading ease? *Scientometrics* 107(3): 1523-1526.

Hascher, Tina. 2010. Learning and emotion: perspectives for theory and research. *European Educational Research Journal* 9(1): 13-28.

Kim, Sun-Young. 2005. An exploration of the ESL college learners' perspectives on integrating L2 reading and writing: dynamics in perspective changes. *Linguistic Research* 22(2): 61-85.

Kincaid, Peter, Robert Fishburne, Richard Rogers, and Brad Chissom. 1975. *Derivation of new readability formulas: Automated reliability index, fog count, and Flesch reading ease formula for navy enlisted personnel* (Branch Report 8-75). Millington, TN.

Landwehr, Niels, Mark Hall, and Eibe Frank. 2005. Logistic model trees. *Machine Learning*, 59(1-2): 161-205.

Laufer, Batia and Geke Ravenhorst-Kalovski. 2010. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22(1): 15-30.

Mar, Raymond and Marina Rain. 2015. Narrative fiction and expository nonfiction differ-

entially predict verbal ability. *Scientific Studies of Reading* 19(6): 419-433.

McAlpine, Rachel. 2005. *Global English for global business*. Wellington, N.Z.: CC Press

McCreath, Graham, Cormac Linehan, and Raymond Mar. 2017. Can differences in word frequency explain why narrative fiction is a better predictor of verbal ability than non-fiction? *Discourse Processes* 54(5-6): 373-381.

McLaughlin, Harry. 1969. Clearing the SMOG. *Journal of Reading* 13(3): 210-211.

Mergenthaler, Erhard. 1996. Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology* 64(6): 1306-1315.

Nation, Paul. 2015. Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language* 27(1): 136 -145.

Poe, Edgar Allan. 2015 [1839]. The Fall of the House of Usher. In *The annotated Poe.* Connecticut: Harvard University Press.

Powell, Joanne, Penelope Lewis, Robin Dunbar, Marta García-Fiñana, and Neil Roberts. 2010. Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia* 48(12): 3554-3562.

Powers, Richard, William Sumner and Bryant Kearl. 1958. A recalculation of four adult readability formulas. *Journal of Educational Psychology* 49(2): 99-104.

Ray-Mukherjee, Jayanti, Kim Nimon, Shomen Mukherjee, Douglas Morris, Rob Slotow, and Michelle Hamer. 2014. Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution* 5(4): 320-328.

Readability Studio. 2015, Oleander Solutions, Vandalia, OH.

Schmitt, Norbert, Tom Cobb, Marlise Horst, and Diane Schmitt. 2017. How much vocabulary is needed to use English? Replication of van Zeeland and Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching* 50(2): 212-226.

Sewell, Anna. 1999 [1877]. *Black beauty*. New York, NY: Dover Publications.

Shin, Jiwon and Tae-Young Kim. 2017. South Korean elementary school students' English learning resilience, motivation, and demotivation. *Linguistic Research* 34: 69-96.

Solnyshkina, Marina, Radif Zamaletdinov, Ludmila Gorodetskaya, and Azat Gabitov. 2017. Evaluating text complexity and Flesch-Kincaid grade level. *Journal of Social Studies Education Research* 8(3): 238-248.

Spache, George. 1953. A new readability formula for primary-grade reading materials. *The Elementary School Journal* 53(7): 410-413.

Stenner, Jackson. 2001. The Lexile Framework: A common metric for matching readers and texts. *California School Library Journal* 25(1): 41-42.

Stine, Robert Lawrence. 2008. *Revenge of the living dummy*. New York, NY: Scholastic.

van Zeeland, Hilda and Norbert Schmitt. 2012. Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*

34(4): 457-479.

Xia, Menglin, Ekaterina Kochmar, and Ted Briscoe. 2016. Text readability assessment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications,* 12-22. San Diego, CA: Association for Computational Linguistics.

Zhang, Wei. 2017. Quality matters: content literacy for English language learners. *TESOL Journal* 8(1): 166-189.

Appendix 1. Variables used in modeling

Variable	Type	Description	Source	Reference
SixLetter	Syntactic	Percentage of words > 6 characters	RiotScan	n/a
Unique_Ratio	Syntactic	The ratio of unique words to total words (%)	RiotScan	n/a
Hapax_Ratio	Syntactic	The ratio of words that appear only once in a body of text to total words (%)	RiotScan	n/a
Unique_Ratio_Sent	Syntactic	Unique_Ratio at the sentence level (%)	RiotScan	n/a
Hapax_Ratio_Sent	Syntactic	Hapax_Ratio at the sentence level (%)	RiotScan	n/a
AvgWordLengthPerSentence	Syntactic	Average word length by sentence.	RiotScan	n/a
AvgWordLengthOverall	Syntactic	Average word length for body of text.	RiotScan	n/a
COVWordLength	Syntactic	Coefficient of variation for word length.	RiotScan	n/a
AvgSentenceLength	Syntactic	Average number of words per sentence.	RiotScan	n/a
COVSentenceLength	Syntactic	Coefficient of variation for number of words per sentence.	RiotScan	n/a
Abstraction	Linguistic	Mergenthaler's measure of linguistic abstraction	RiotScan	Mergenthaler (1996)
GLNorms_DictPercent	Linguistic	% of words in text that are in the G&L word norms	RiotScan	Gilhooly & Logie (1980)
GLNorms_Imagery_M	Linguistic	Mean word imagery based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Imagery_MSD	Linguistic	Mean word imagery based on G&L word norm stdevs	RiotScan	Gilhooly & Logie (1980)
GLNorms_AOA_M	Linguistic	Mean word age of acquisition based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)

GLNorms_AOA_MSD	Linguistic	Mean word age of acquisition based on G&L word norm stdevs	RiotScan	Gilhooly & Logie (1980)
GLNorms_Familiar_M	Linguistic	Mean word familiarity based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Familiar_MSD	Linguistic	Mean word familiarity based on G&L word norm stdevs	RiotScan	Gilhooly & Logie (1980)
GLNorms_Concrete_M	Linguistic	Mean word concreteness based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Concrete_MSD	Linguistic	Mean word concreteness based on G&L word norm stdevs	RiotScan	Gilhooly & Logie (1980)
GLNorms_Uncertain_M	Linguistic	Mean word uncertainty based on G&L word norm stdevs	RiotScan	Gilhooly & Logie (1980)
GLNorms_Imagery_SDM	Linguistic	Stdev of word imagery based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_AOA_SDM	Linguistic	Stdev of word age of acquisition based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Familiar_SDM	Linguistic	Stdev of word familiarity based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Concrete_SDM	Linguistic	Stdev of word concreteness based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
GLNorms_Uncertain_SDM	Linguistic	Stdev of word uncertainty based on G&L word norm means	RiotScan	Gilhooly & Logie (1980)
% of long (6+ characters) words	Syntactic	Percentage of words >= 6 characters	Readability Studio	n/a
% of SMOG hard words	Syntactic	Percentage of words >= 3 syllables	Readability Studio	McLaughlin (1969)
% of Fog hard words	Syntactic	% words >=3 syllables excluding suffixes and compound words	Readability Studio	Gunning (1968)
% of Dale-Chall unfamiliar words	Word List	% words from the Dale-Chall unfamiliar word list	Readability Studio	Dale and Chall (1995)
% of Spache unfamiliar words	Word list	% words from the Spache unfamiliar word list	Readability Studio	Spache (1953)
% of HJ unfamiliar words	Word List	% words from the Harris-Jacobson unfamiliar word list	Readability Studio	Harris and Jacobson (1982)
Passive	Grammatical	Instances of the passive voice divided by word count	Readability Studio	n/a
Bormuth Cloze Mean	Readability	Bormuth Cloze Mean readability formula	Readability Studio	Bormuth (1966)
Bormuth Grade Placement	Readability	Bormuth Grade Placement readability formula	Readability Studio	Bormuth (1966)
Coleman-Liau (grade levels)	Readability	Coleman-Liau (grade levels) readability formula	Readability Studio	Coleman and Liau (1975)
Coleman-Liau (pred. Cloze scores)	Readability	Coleman-Liau (pred. Cloze scores) readability formula	Readability Studio	Coleman and Liau (1975)

Danielson-Bryan 1	Readability	Danielson-Bryan 1 readability formula	Readability Studio	Danielson' and Bryan (1963)
Danielson-Bryan 2	Readability	Danielson-Bryan 2 readability formula	Readability Studio	Danielson' and Bryan (1963)
Degrees of Reading Power	Readability	Degrees of Reading Power readability formula	Readability Studio	Carver (1985)
Degrees of Reading Power (GE)	Readability	Degrees of Reading Power (GE) readability formula	Readability Studio	Carver (1985)
EFLAW	Readability	EFLAW readability formula	Readability Studio	McAlpine (2005)
Farr, Jenkins, Paterson	Readability	Farr, Jenkins, Paterson readability formula	Readability Studio	Farr, Jenkins and Paterson (1951)
Flesch Reading Ease	Readability	Flesch Reading Ease readability formula	Readability Studio	Flesch (1979)
FORCAST	Readability	FORCAST readability formula	Readability Studio	Caylor (1973)
Gunning Fog	Readability	Gunning Fog readability formula	Readability Studio	Gunning (1968)
LIX (index values)	Readability	LIX (index values) readability formula	Readability Studio	Bjornsson (1983)
Modified SMOG	Readability	Modified SMOG readability formula	Readability Studio	McLaughlin (1969)
New Fog Count	Readability	New Fog Count readability formula	Readability Studio	Kincaid et al (1975)
PSK Dale-Chall	Readability	PSK Dale-Chall readability formula	Readability Studio	Powers, Sumner and Kearl (1958)
PSK Farr, Jenkins, Paterson	Readability	PSK Farr, Jenkins, Paterson readability formula	Readability Studio	Powers, Sumner and Kearl (1958)
PSK Gunning Fog	Readability	PSK Gunning Fog readability formula	Readability Studio	Powers, Sumner and Kearl (1958)
PSK Flesch	Readability	PSK Flesch readability formula	Readability Studio	Powers, Sumner and Kearl (1958)
RIX (index values)	Readability	RIX (index values) readability formula	Readability Studio	Anderson (1983)
SMOG	Readability	SMOG readability formula	Readability Studio	McLaughlin (1969)
SMOG (simplified)	Readability	SMOG (simplified) readability formula	Readability Studio	McLaughlin (1969)

Spache Revised	Readability	Spache Revised readability formula	Readability Studio	Spache (1953)
Mean	Readability	Mean grade level derived from applicable readability formulas	Readability Studio	n/a

Appendix 2. Texts ranked by reading complexity score

Author	Text	Score	Word Count	Primary Text = 1	Flesch-Kincaid Grade Level
Stine, RL	Revenge of the Living Dummy	-2.11	595	1	2.80
Fargo, Jimmy	Tales of a Fourth Grade Nothing	-2.10	822	1	3.00
Cooper, Susan	The Magician's Boy	-1.96	529	1	5.40
Pearce, Philippa	A Finder's Magic	-1.56	861	1	2.90
Wilder, Laura Ingalls	Little House on the Prairie	-1.51	531	1	3.10
Gertsein, Mordicai	The Old Country	-1.46	633	1	3.60
DiCamillo, Kate	The Tiger Rising	-1.44	695	1	6.30
Blyton, Enid	Five of a Treasure Island	-1.42	1899	1	3.80
Sewell, Anna	Black Beauty	-1.37	782	1	6.60
Holm, Jennifer L	Turtle in Paradise	-1.37	730	1	4.20
Rawls, Wilson	Summer of the Monkeys	-1.35	1023	1	6.50
Gardner, Sally	Coriander	-1.35	760	1	8.80
Yoo, Paula	Sixteen Years in Sixteen Seconds	-1.34	455	1	6.40
Sachar, Louis	Holes	-1.34	1163	1	5.80
Alexander, William	Ghoulish Song	-1.29	1223	1	4.50
Cleary, Beverley	Ramona and Her Father	-1.28	737	1	6.50
Lawrence, Iain	Lord of the Nutcracker Men	-1.27	719	1	5.40
Baum, Frank	The Wonderful Wizard of Oz	-1.17	1147	1	7.10
Godden, Rumer	Doll's House	-1.15	1673	1	7.70
de Saint-Exupery, Antoine	The Little Prince	-1.13	1407	1	3.00
Brothers Grimm	Hansel and Gretel	-1.12	2710	1	6.00
Lofting, Hugh	The Voyages of Doctor Doolittle	-1.12	896	1	7.90
Dahl, Roald	George's Marvellous Machine	-1.12	1674	1	3.30
Lewis, CS	The Chronicles of Narnia	-1.10	1613	1	6.20
Stead, Rebecca	When You reach Me	-1.09	1008	1	4.90
DuPrau, Jeanne	In the City of Ember	-1.07	540	1	6.20
Clements, Andrew	The School Story	-1.06	1025	1	3.60
Palacio, RJ	Wonder	-1.05	1543	1	4.90
Yolen, Jane	Centaur Rising	-1.04	1419	1	6.00
Nix, Garth	Sabriel	-1.04	702	1	5.60

Travers, PL	Mary Poppins	-1.03	1530	1	6.40
Rowling, JK	Harry Potter and the Sorcerer's Stone	-1.03	3452	1	6.20
Cohn, Edith	Spirit's Key	-1.03	1324	1	3.00
Hill, Stuart	The Cry of the Icemark	-1.02	701	1	7.80
Carlo Lorenzini	The Adventures of Pinocchio	-1.01	2428	1	4.60
Pedley, Ethel C	Dot and the Kangaroo	-1.00	2227	1	7.90
O'Dell, Scott	Island of the Blue Dolphins	-0.96	1661	1	4.20
White, EB	Charlotte's Web	-0.94	2082	1	3.50
Wilson, Jacqueline	Tracy Backer	-0.90	685	1	4.70
Carroll, Lewis	Alice's Adventures in Wonderland	-0.90	2160	1	10.50
Atwater, Richard & Florence	Mr Popper's Penguins	-0.88	3126	1	5.20
Farmer, Nancy	The Sea of trolls	-0.87	667	1	1.90
Selden, George	A Cricket in Times Square	-0.87	3188	1	4.40
Turner, Pamela	Hachicko, the true Story of aLoyal Dog	-0.86	379	1	5.10
Shulevitz, Uri	The travels of Benjamin of Tudela	-0.83	636	1	4.10
Gleitzman, Morris	Once	-0.81	1789	1	4.30
Dhami, Narinder	Bindi Babes	-0.79	1699	1	3.90
Castro, Adam-Troy	Gustav Gloom and the People Taker	-0.79	1806	1	7.70
L'Engle, Madeleine	The Twnty-Four Days before Christmas	-0.78	2824	1	5.80
Valente, Catherynne	The Girl Who Fell Beneath Fairyland and Led the Revels There	-0.78	2880	1	7.20
Kipling, Rudyard	The Jungle Book	-0.77	6059	1	5.70
McCloskey, Robert	Sensational Scent	-0.75	1876	1	8.20
Griff, Patricia Reilly	The House of Tailors	-0.74	1196	1	5.40
Ransom, Arthur	A Child's Book of the Seasons	-0.73	2967	1	7.80
Curtis, Christopher Paul	Bud not Buddy	-0.73	1960	1	6.20
Andersen, Hans Christian	The Ugly Duckling	-0.72	3800	1	7.40
Coraline	Gaiman, Neil	-0.71	1936	1	5.00
Hughes, Ted	The Iron Man	-0.70	1345	1	3.10
McKay, Hilary	Saffy's Angel	-0.69	3123	1	4.20
Horvarth, Polly	MyOne Hundred Adventures	-0.69	1048	1	6.20
Nesbit, E	Five Children and It	-0.68	1017	1	7.10
Pope, Elizabeth Marie	The Sherwood Ring	-0.64	1636	1	7.00
Armstrong, KL	Loki's Wolves	-0.61	3380	1	2.90
Faulkner, William	As I Lay Dying	-0.60	2731	0	5.70
Pullman, Philip	Book of Dust	-0.58	2770	1	4.30
Jacobs, Joseph	Jack and the Beanstalk	-0.57	4022	1	6.40
Funke, Cornelia	Inkheart	-0.56	2997	1	4.10
Riley, James	Story Thieves	-0.53	1767	1	5.80
Cowell, Cressida	How to Train your Dragon	-0.52	1443	1	6.10
Wilde, Oscar	The Happy Prince	-0.52	3516	1	5.90
Stewart, Trenton Lee	The Mysterious Benedict	-0.51	836	1	9.10

Spyri, Johanna	Heidi	-0.51	1009	1	8.20
Juster, Norton	The Phantom Tollbooth	-0.51	1008	1	8.80
Hemingway, Ernest	The Old Man and the Sea	-0.51	1145	0	3.30
Norton, Mary	The Borrowers	-0.49	2006	1	5.10
Steinbeck, John	East of Eden	-0.47	2103	0	7.90
Levy, Andrea	Small Island	-0.43	3538	0	3.70
Wiggin, Kate Douglas	Rebecca of Sunnybrook Farm	-0.43	1006	1	11.30
Burnett, Frances Hodgson	The Secret Garden	-0.42	1916	1	6.10
Le Guin, Ursula	The Wizard of Earthsea	-0.40	2680	1	8.90
Spinelli, Jerry	Hokey Pokey	-0.39	519	1	3.00
Perrault, Charles	Beauty and the Beast	-0.37	5723	1	9.80
Adams, Richard	Tales from Watership Down	-0.37	1406	1	6.20
Bond, Michael	A Bear Called Paddington	-0.36	2994	1	4.70
Ibbotson, Eva	Journey to the River Sea	-0.33	3268	1	5.00
Potter, Beatrix	The Tailor of Gloucester	-0.33	2888	1	7.50
Cather, Willa	O Pioneers!	-0.32	3180	0	6.30
Morpurgo, Michael	Private Peaceful	-0.31	503	1	3.50
Winterson, Jeanette	Oranges Are Not the Only Fruit	-0.30	1745	0	6.10
Barrie, JM	Peter Pan	-0.27	2551	1	8.80
Ferris, Jean	Once Upon a Marigold	-0.26	2143	1	4.30
Aesop	Fables	-0.26	2291	1	7.50
Dickens, Charles	A Christmas Carol	-0.23	2367	1	3.10
Alcott, Louise May	Little Women	-0.23	4118	1	7.30
MacDonald, George	The Princess and Curdie	-0.22	1849	1	13.40
Atwood, Margaret	The Handmaid's Tale	-0.22	2834	0	5.50
Salinger, JD	Catcher in the Rye	-0.21	1932	0	4.90
Mass, Wendy	The Candymakers	-0.20	2984	1	6.80
Riddell, Chris & Stewart, Paul	Corby Flood	-0.19	612	1	5.10
Chandler, Raymond	The Big Sleep	-0.19	2452	0	4.80
Porter, Elanor H	Pollyanna	-0.18	904	1	6.70
van Eekhout, Greg	The Boy at the End of the World	-0.17	1669	1	3.80
Gruen, Sara	At the Water's Edge	-0.16	2537	0	4.40
Malamud, Bernard	Armistice	-0.15	2243	0	5.50
Gratz, Alan	The League of Seven	-0.11	3908	1	5.60
Montgomery,Lucy Maud	Anne of Green Gables	-0.07	2958	1	8.90
Homer	The Odyssey	-0.07	3225	0	12.60
Sedgwick, Marcus	Saint Death	-0.06	1966	1	6.80
Smith, Dodie	I Capture the Castle	-0.05	3522	1	7.20
McDermott, Alice	Charming Billy	-0.05	2497	0	6.80
Wyss, Johann David	The Swiss Family Robinson	-0.03	3113	1	8.20
Parker, Dorothy	The Waltz	0.01	2008	0	3.30
Lessing, Doris	Wine	0.02	1903	0	4.60
Flaubert, Gustave	Madame Bovary	0.07	6509	0	8.90

Plath, Slyvia	The Bell Jar	0.09	1723	0	10.00
Camus, Albert	The Plague	0.09	5094	0	5.50
Stedman, ML	The Light Between Oceans	0.10	1488	0	5.90
Grahame, Kenneth	Wind in the Willows	0.13	4411	1	7.30
Gaskell, Elizabeth	Mary Barton	0.13	3535	0	12.50
Golding, William	Lord of the Flies	0.14	5450	0	4.00
Ballantyne, RM	The Coral Island	0.14	3109	1	10.90
Naji, Ahmed	Using Life	0.16	3470	0	5.90
Roth, Philip	Indignation	0.16	3197	0	7.30
Saunders, George	The Tenth of December	0.22	8945	0	3.00
Stevenson, Robert Louis	The Strange Case of Dr Jekyll and Mr Hyde	0.27	2436	0	8.10
Tartt, Donna	The Secret History	0.27	1570	0	11.10
Grisham, John	The Runaway Jury	0.32	3069	0	7.40
Bram Stoker	Dracula	0.32	5534	0	7.90
Tsiolkas, Christos	The Slap	0.33	1159	0	6.20
Berendt, John	Midnight in the Garden of Good and Evil	0.35	7612	0	6.00
Sheckley, Robert	Ask a Foolish Question	0.39	2701	0	4.70
Doctorow, EL	Ragtime	0.39	2202	0	6.10
Bulgakov, Mikhail	The Master and Margarita	0.43	1399	0	6.40
Bellow, Saul	The Adventures of Augie March	0.43	5153	0	8.20
Woolfe, Virginnia	To the Lighthouse	0.50	1612	0	13.60
Salter, James	All That Is	0.50	2650	0	5.80
Murdoch, Iris	An Accidental Man	0.51	4603	0	5.30
McCulloch, Colleen	Let the Dice Fly	0.51	9875	0	6.50
Morrison, Toni	Tar Baby	0.52	2628	0	8.50
Dostoyevsky, Fyodor	Crime and Punishment	0.53	3389	0	7.80
Amis, Martin	Night train	0.56	5214	0	4.50
Burgess, Anthony	A Clockwork Orange	0.58	3450	0	8.60
Esquivel, Laura	Pierced by the Sun	0.59	2066	0	7.70
Hesse, Herman	Siddhartha	0.60	9072	0	9.10
Conrad, Joseph	Heart of Darkness	0.61	2689	0	9.10
Fitzgerald, F Scott	The Great Gatsby	0.62	2649	0	11.90
Proulx, Annie	The Shipping News	0.64	3494	0	5.80
Cavendish, Margaret	The Blazing_World	0.64	7717	0	19+
Vonnegut, Kurt	Slaughterhouse 5	0.65	5405	0	5.70
Austen, Jane	Northanger Abbey	0.70	1403	0	13.10
James, Henry	The Portrait of a Lady	0.70	9326	0	7.60
Beckett, Samuel	Dante and the Lobster	0.72	4323	0	5.40
Wilder, Thornton	The Eighth Day	0.72	537	0	10.10
McCarthy, Cormac	All the Pretty Horses	0.74	3811	0	3.50
Marquez,Gabriel Garcia	One Hundred Years of Solitude	0.74	997	0	12.70
Joyce, James	Ulysses	0.76	7342	0	3.70
Wharton, Edith	Age of Innocence	0.79	1993	0	12.80
Dick, Philip K	Do Androids Dream of Electric	0.86	886	0	4.90

	Sheep?				
Christie, Agatha	The Mysterious Affair at Styles	0.88	2524	0	6.20
Heller, Joseph	Catch 22	0.92	3652	0	6.40
Singer, Isaac Bashevis	Shadows on the Hudson	0.93	3151	0	7.90
Waugh, Evelyn	The Loved One	0.94	3810	0	5.50
Drabble, Margaret	The Witch of Exmoor	0.96	9019	0	6.00
Melville, Herman	Moby Dick	0.98	2244	0	9.40
Orwell, George	Nineteen Eighty-Four	0.99	9353	0	8.70
Tolstoy, Leo	War and Peace	0.99	2087	0	7.60
Chabon, Michael	Telegraph Road	0.99	7102	0	7.20
Carey, Peter	Amnesia	1.00	3436	0	7.90
Mohsin, Hamid	Exit West	1.04	1731	0	14.30
Updike, John	Seek my Face	1.04	5436	0	12.20
Capote, Truman	In Cold Blood	1.05	2749	0	14.30
Hardy, Thomas	The Mayor of Casterbridge	1.15	4635	0	8.20
Pynchon, Thomas	Gravity's Rainbow	1.15	2193	0	7.90
Eliot, George	Middlemarch	1.19	3304	0	11.40
Sienkiewicz, Henryk	Quo Vadis?	1.20	4689	0	8.10
Perrotta, Tom	Heroes' Day	1.22	6183	0	10.10
Cervantes, Miguel de	Don Quixote	1.24	26500	0	18.60
Mishima, Yukio	The Temple of the Golden Pavilion	1.24	3930	0	10.00
Mann, Thomas	The Magic Mountain	1.25	2785	0	7.70
Goethe, Johann, Wolfgang	The Sorrows of Young Werther	1.36	16023	0	9.70
Eco, Umberto	Foucault's Pendulum	1.38	1598	0	11.30
Rushdie, Salman	The Satanic Verses	1.40	2643	0	9.40
Barnes, Djuna	Nightwood	1.47	1146	0	14.30
Stendhal	The Charterhouse of Palma	1.48	4726	0	14.80
Proust, Marcel	Swan's Way	1.48	20927	0	17.20
Fielding, Henry	The Adventures of Tom Jones	1.51	6137	0	17.90
Franzen, Jonathan	The Corrections	1.52	3612	0	10.90
Rand, Ayn	Atlas Shrugged	1.56	2663	0	8.30
Bronte, Emily	Wuthering Heights	1.67	5385	0	9.10
Borges, Jorge Luis	The Aleph	1.68	3831	0	10.50
Abbott, Edwin	Flatland	2.02	1328	0	14.00
Rabelais, François	Gargantua and Pantagruel	2.07	5827	0	8.90
Shelley, Mary	Frankenstein	2.07	17120	0	11.60
Hugo, Victor	The Hunchback of Notre Dame	2.11	5363	0	9.20
Asimov, Isaac	Second Foundation	2.12	3701	0	8.60
Hawthorne, Nathaniel	The Scarlet Letter	2.25	9376	0	15.90
Huxley, Aldous	Brave New World	2.45	1005	0	10.30
Wallace, David Foster	Backbone	2.58	4894	0	14.90
Boccaccio, Giovanni	The Decameron	2.67	5564	0	16.40
Poe, Edgar Allan	The Fall of the House of Usher	2.80	7180	0	14

Michael Dalvean
Teacher
Humanities Department
Woden Campus, Canberra College
Launceston Street, Phillip ACT 2606
Australia
Email: michael.dalvean@ed.act.edu.au

Galbadrakh Enkhbayar
Undergraduate Student
Department of International Business
College of Business
Southern Taiwan University of Science and Technology
No 1 Nan-Tai Street, Yungkang District, Tainan City 71005, Taiwan
Email: ggbegalbadrakh@gmail.com

Received: 2018. 01. 11.
Revised: 2018. 08. 19.
Accepted: 2018. 08. 19.