Understanding Readability

Applying readability measures to long-form exam questions

Steven LLOYD[1)]

Abstract

Readability formulas are extremely useful for assessing the difficulty of texts for long-form exam questions. However, most are aimed at L1 readers and do not specifically examine the text from an L2 viewpoint. The CEFR project of the Council of Europe has gone a long way to redressing this problem. This paper looks at a range of the different formulas and vocabulary lists, and adds a further list, the level at which the vocabulary is introduced to Japanese school students. The paper outlines the construction of a spreadsheet to apply this list plus other lists and formulas to texts. The results are then compared with those provided by online readability formulas.

Keywords: readability, entrance exam questions, CEFR, NGSL, Google Sheets

1. Introduction

There are many different formulas for measuring the readability of a text, often producing different results. This clearly suggests that what is seen as important in making a text easier or more difficult to read is an issue of contention. Computers have made applying the various formulas much easier, and allowed for more complex formulas to be applied. There are essentially three methods for measuring readability. One is to analyse the various parts of the text: syllables, word and sentence length, and the various ratios. The second is to compare the words to a corpus that has possibly graded the words by use or difficulty. This is often combined with an examination of the ratio of word types, such as verbs or conjunctions per sentence. The third approach is to look at sentence complexity, such as the number of verbs, and verb types, sub-clauses and conjunctions. Each of these methods have their merits and weaknesses.

This paper will examine the application of the various measures of readability, and the strengths and weaknesses of the three patterns, in particular to their usefulness for preparing Japanese university entrance exam long-form questions. While there are several very good (and mostly free) sites available online that will allow the user to calculate readability using a variety of formulas, this paper will look at the efforts made to apply them in a spreadsheet to fully understand their mechanics. In addition to the standard formulas such as Flesch-Kincaid and corpora such as CEFR and NGSL, readability was measured against Japanese school textbook vocabulary. This is seen as a potentially useful measure as it is indicative of when the student was exposed to the vocabulary, and the familiarity they (should) have with it.

1. 1 Structure of this paper

This paper will first look at the readability formulas and then the corpora that have been used in this paper.

[1)] Faculty of International Business Management, Kyoei University

Next, it will look at the structure of the spreadsheet, and discuss some of the difficulties and limitations of the format. Then, it will apply an entrance exam question to the spreadsheet, and compare the various different results. It will also compare these results with those from an online readability service.

2. Measuring Readability

The first scientific attempts at judging a text's readability was word frequency. In the 1920s, Thorndike and Lorge compiled a list of 10,000 words, later expanded to 30,000 words aimed at helping teachers write texts and books for their students as a drive to improve literacy (Thorndike & Lorge, 1944). These lists were seen as essential words a reader should know to be classed as literate.

The idea of measuring readability with a formula came about as a consequence of the USA's experiences of the 1930s and 40s, when it was realised that the material available was often beyond the reading abilities of the adult population, thus holding many of them back. From a military perspective, unreadable manuals and instructions could lead to serious, even fatal, mistakes. The history of readability formulas is littered with cooperation between the military and academia.

2. 1 Readability Formulas

These formulas were taken from the site *Readability Formulas* (https://readabilityformulas.com/), and then confirmed for accuracy using other references.

The most successful of the early formulas was the Flesch Reading Ease formula. It applies average sentence length (ASL) and average syllables per word (ASW) to a formula to provide a score that can then correspond with a reading age, measured by school grade. The higher the score, the lower the reading age; a score of 0 suggests a graduate level text, and that of 100 a 4th grade reading level. The formula is:

$$Flesch\ score = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

This has proven to be a very successful and resilient formula, and continues to be used today. Most other formulas follow a very similar pattern.

The Flesch-Kincaid Grade Level formula takes the same elements but applies different weighting factors to give a result corresponding with a school grade.

$$F\text{-}K\ Grade\ Level = (0.39 \times ASL) + (11.8 \times ASW) - 15.59$$

The Linsear Write Readability Formula takes a somewhat different approach. It takes a 100 word sample from the text and divides words up into easy words, defined as two or less syllables, and assigns them a weight of 1, and hard words, defined as three or more syllables, and assigns them a weight of 3. These are summed together, and if the score is greater than 20, divided by 2. If less than or equal to 20, subtract 2 and divide by 2. The resulting score is the school grade level of the sample.

The Automated Readability Index (ARI) takes an even simpler approach by looking at characters per word and ASL, thus not requiring the need for understanding syllables, and making it much easier for computers to calculate. The result is a score from 1 for kindergarten up to 14 for professor. The formula is:

$$ARI = 4.71\left(\frac{characters}{words}\right) + \left(\frac{words}{sentences}\right) - 21.43$$

The Coleman-Liau Index takes a similar approach, but only using 100 word samples. Where L is the number of words in the 100 word sample, and S is the number of sentences, the formula results in a grade level score:

$$CLI = 0.0588L - 0296S - 15.8$$

The *Läsbarhetsindex* (or LIX for short) incorporates longer words into the formula. It includes the total numbers of words (A), the number of periods and colons (or capital letters, as a means of counting sentences) (B), and the number of words longer than 6 letters (C). It produces a number that corresponds with a difficulty level from very easy to very difficult.

$$LIX = \frac{A}{B} + \left(\frac{C \times 100}{A} \right)$$

Although there is a lot of overlap in the criteria used, they can often produce quite different levels of reading difficulty, as will be shown below.

Although not a formula as such, word frequency is also seen as an important measure of readability. There are a wide range of studies looking at the frequency of words in a text, in part popularised by George R. Klare (1963).

2. 2 Corpora

The main corpus used in this analysis is the Common European Framework of Reference (CEFR) word list, or vocabulary profile, available at (https://www.englishprofile.org/wordlists/evp). It was set up under the direction of the Council of Europe in the 1990s (COE, 2020). CEFR ranks language and vocabulary by learner ability into six broad levels, from A1 to A2 and up to C2. It measures what tasks a speaker is able to do, so is language neutral (thus making it possible to compare a student of German with a student of Greek, by assessing what they 'can do'). The vocabulary profile includes a list of over 15,600 items; many words are repeated as they can have different meanings, thus a separate entry and a different CEFR level score. For example, "or" has four entries, denoting possibilities, alternatives, a warning, and change. Each is a different level (from A1 to B2), but that is not always the case. It has become immensely popular and useful, with other countries adapting it to their own circumstances (such as Japan) and it is widely used in language textbooks. Japan, for example, has an extensive CEFR-J project (http://www.cefr-j.org/) that has gone a long way towards making the work done by the original CEFR even more relevant for Japanese language learners. For the purpose of this analysis, only A1 to B2 words are used.

The second word list applied is the New General Service List (NGSL), which has been around for about 70 years. It is a series of lists of several million words, available at (http://www.newgeneralservicelist.org/). It is widely used in testing (such as TOEIC) and textbook production. The NGSL 1.01 word list of 3,000 words was used for this analysis.

A third list is the words used in Japanese school textbooks. While I was not able to find a definitive list (which may not exist as publishers - although following a core vocabulary - will add their own vocabulary too), there are quite a few lists compiled by cram schools online that are freely available. I used the lists from Eigo Zuke (https://www.eigo-duke.com/) which provides a variety of lists grouped by school year and test type (such as the Japanese University Entrance Exam Centre Test). For the purpose of this study, words are grouped into Elementary, Junior High 1, 2 and 3, and High School 1, 2 and 3.

There is yet another list, the New Dale-Chall list of 3,000 words (all of which occur within the other two)

that applies both a corpora and a formula. This calculates the percentage of difficult words (PDW), defined as those not appearing on the list (made up of words that should be familiar to most US fourth grade students) and combines that with average sentence length (ASL):

Raw Score = 0.1579 x (PDW) + 0.0986 x ASL

Adjusted Score = PDW > 5%: Raw Score + 3.6365, otherwise Adjusted Score = Raw Score

The adjusted score is then applied to a table matching school grade level (so for example , 4.9 or lower is grade 4 or lower, 6.0~6.9 is equal to grades 7 to 8; a score of 10 or more is equal to college graduate reading level).

3. Word Analysis Spreadsheet

3. 1 A brief history of the Word Analysis Spreadsheet

The original Word Analysis spreadsheet was a very simple attempt to count the words and their frequency, using the Apple *Numbers* spreadsheet application. If I had known where this project was leading, I should have taken a course in Python and started from there (however, later attempts led me to believe this was beyond my computing skills, especially given the complexity of the program I was then trying to compile). I then tried to add a wordlist from a high school dictionary (a very laborious task; and now out of date so, although still in the current word list, not used in any way). I then realised that the CEFR list would be the most useful as it was the basis for a growing number of English language textbooks aimed at university students, and even school textbooks. While a valuable resource, the main short-coming of this list is that it is not easily downloaded and the site only has the stems and is not lemmatized in a form that can be easily broken up by a computer. Therefore, it was necessary to add all the inflections. While for a lot of words this was not a problem, English is cursed with a plethora of irregular nouns, verbs, spellings and so on. A lot of the work could be done through the application of formulas, but it was necessary to add a lot by hand. It was also necessary to add the syllable count for every word by hand too. Fortunately, the NGSL provides lemmatized lists, so this was much easier to add. The CEFR did, however, provide an extremely useful item, the grammar marker for each word.

While a very elegant and easy to use application, Apple *Numbers* is not at all very suitable for heavy data analysis. As the spreadsheet, and the ambition of the project grew, the spreadsheet got increasingly slow. Even breaking it up and copy-pasting between sheets proved ineffective. I then decided to move over to Google *Sheets* (very similar to Microsoft *Excel*, but without all the bells and whistles, plus a few bells not available in *Excel*). The vastly greater variety of functions made the spreadsheet much lighter and faster. For example, in *Numbers* it was necessary to use a series of FIND(), MID() and LEN() functions to separate the sentences into individual words. With Google *Sheets*, this was done using SPLIT(). Similarly, listing all the words used in the text was a simple matter of using UNIQUE(), a function not available in *Numbers*, which required several steps to get around it. What previously took five or six separate documents, is now done in one, plus a lot more.

3. 2 The word list

The word list is currently made up of 13,626 words. This includes plurals, past tense, alternative spellings (for example, both British and American spellings of words) and commonly occuring proper knowns and related words (such as country names and their adjectival forms). Each word includes its CEFR level, a

grammar marker, the number of syllables, its length, its school level, whether it is on the NGSL, and additional markers that will be discussed below. See Fig 1 for a screenshot of a section of the word list.

8	Word	CEFR	Grammar	School	Syllables	character	Irregular	NGSL	Conjunction
9	13,626	13,080		4,459			498	6,064	20
358	already	A2	adv	JH1	3	7		NGSL	
359	alright	A2	adj		2	7		NGSL	
360	also	A1	adv	JH1	2	4		NGSL	
361	altar	B2	noun		2	5			
362	altars	B2	noun		2	6			
363	alter	B2	verb	HS2	2	5		NGSL	
364	altered	B2	verb	HS2	2	7		NGSL	
365	altering	B2	verb	HS2	3	8		NGSL	
366	alternate	B2	verb		3	9			
367	alternated	B2	verb		4	10			
368	alternates	B2	verb		3	9			
369	alternating	B2	verb		4	11			
370	alternative	B2	noun	HS2	4	11		NGSL	
371	alternatively	B2	adv		5	13			
372	alternatives	B2	noun	HS2	4	12		NGSL	
373	alters	B2	verb	HS2	2	5		NGSL	
374	although	B1	func	JH3	2	8		NGSL	conj
375	altitude	B2	noun		3	8			
376	altitudes	B2	noun		3	9			
377	altogether	B1	adv	HS3	4	10		NGSL	
378	aluminium	B2	noun		4	9			
379	aluminum	B2	noun		4	8			
380	always	A1	adv	JH1	2	6		NGSL	
381	am	A1	verb	EL	1	2	irr	NGSL	
382	amaze	B2	verb		2	5		NGSL	
383	amazed	B1	adj		2	6		NGSL	
384	amazement	B2	noun		3	9			
385	amazes	B2	verb		2	6		NGSL	
386	amazing	A2	adj	HS3	3	7		NGSL	
387	amazingly	B1	adv		4	9			

Fig 1　Screenshot of word list

There are several issues with the list that are very hard to address unless a lot of the analysis is done manually. As mentioned above, many words have more than one meaning, and have different CEFR levels. However, the spreadsheet is unable to determine which meaning is being used. Duplication of words is of no use as the spreadsheet will only take the first occurrence of it in the list. Furthermore, this would also make the list even longer and slower. So, it was decided to only add each word once, and use the meaning that was determined to be most useful, and the corresponding CEFR level. Similarly, many words have more than one grammar function - a word can be a verb, noun or adjective, depending on the context. Again, one grammar marker was given per word. This does occasionally result in errors, but unavoidable without manual inputting. However, this can in part be overcome by temporarily adding words, and forcing the spreadsheet to accept your preferences, as described below.

The grammar markers used are noun, verb, adjective (adj), adverb (adv) and function words (func). As can be seen in Fig 1, there are also markers for irregular verb forms and conjunctions. The word list has made a lot of progress from its original version of words manually copied from *The Grand Century English-Japanese Dictionary* (Sanseido, 2010).

3. 3 The steps

3. 3. 1 The user interface

The first sheet is the input page. Here, the text to be analysed is pasted into the appropriate cell. The user is presented with a list of missings words that are in the text but not on the word list, that need to be manually assigned a CEFR level and the number of syllables in the word. They are able to look up words that are similar that may be on the list (for example, *younger* is not on the list, but *young* is, and the word look-up gives you its CEFR level). As the spreadsheet looks at this list first, the user is also able to override the wordlist, thereby forcing a word to be seen as, for example, a noun rather than a verb. There is also the option to add a list of keywords the user wishes to highlight.

The final sheet is the output page. Here, the user is given a summary of all the statistics, plus a drop down menu of options that allow the full text to be colour-coded by particular markers. So, by choosing CEFR, all the A1 words are light blue, the A2 words darker blue, the B1 words light green and the B2 words darker green. Similarly, by choosing grammar, all the different markers (noun, verb, adjective, adverb and function) are colour coded individually. The options available are CEFR (A1 to B2), school (elementary to high school), Dale-Chall, grammar, NGSL, keywords (as defined by the user on the input page) frequency, content frequency (words other than function words), and conjunctions (the sheet marks the most common eight conjunctions[1] plus a few more). Although the keywords marker is not all that important, it is useful to see how these words are distributed through the text. In addition, a short note summarising the information is given, such as how many of each marker is present and the plurality marker as a percentage.

Because each word is in its own cell, it is not quite as easy to read as a regular paragraph, and each sentence is on its own line. Longer words will often not be completely visible because the cell does not resize, while smaller words can appear somewhat isolated. This is a drawback of using a spreadsheet, but not too serious.

As can be seen in Figure 2, the output page gives the results of all six formulas outlined above. It also gives the vital statistics of the text: sentence, word, character and syllable count, the average sentence and word

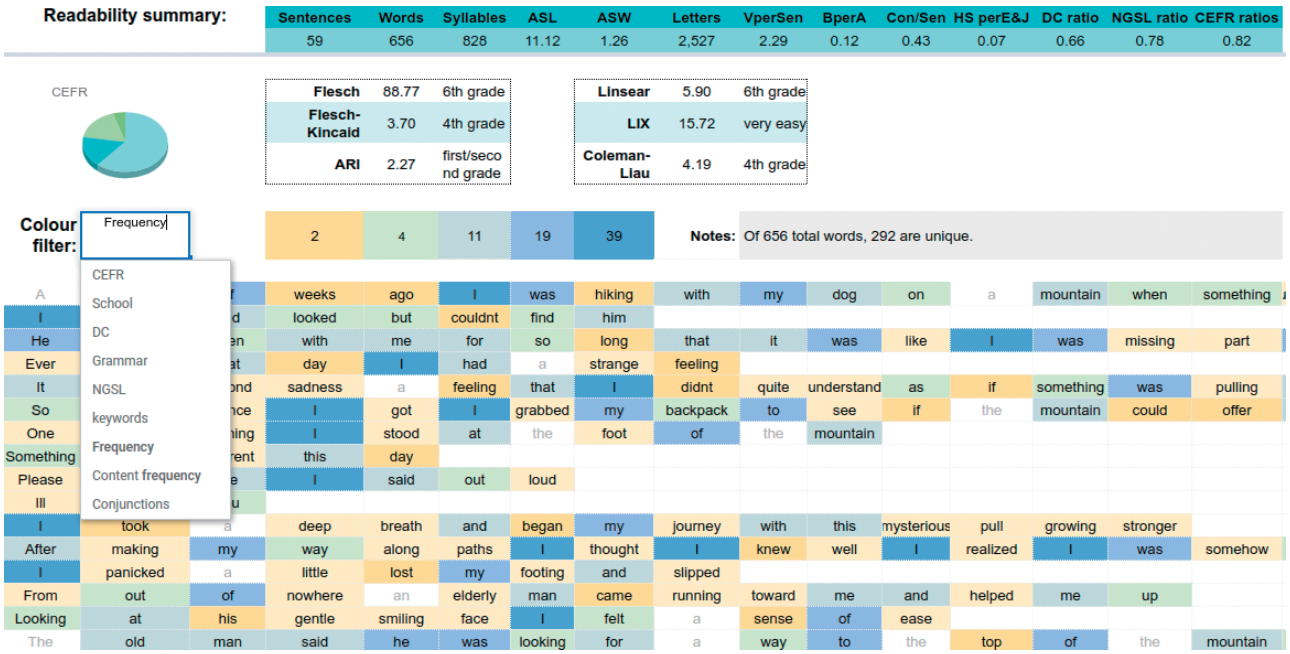

Fig 2　Screenshot of output page

lengths, the average verbs per sentence (VperS), the ratio of B per A (CEFR), the ratio of content words per sentence, the ratio of high school words per elementary plus junior high school words, ratio of NGSL words and the ratio of CEFR words. It also gives a simple pie chart of CEFR words.

The colour coding is all done using a series of conditional formatting rules. However, they do not apply directly to the words that are visible, but to a group of markers hidden lower down the page. So, if School is chosen, the cell corresponding with cell D10 and the word *weeks* has the marker JH1 to indicate that it is a junior high school first year word. The conditional formatting rule notes the *JH* and colour codes the corresponding cell appropriately using an OFFSET() function. A change of option to CEFR, and the markers lower down the page also change, thus triggering a different set of conditional formatting rules.

3. 3. 2 Analysing the text

There are four pages involved in analysing the text the user pastes into the input page. The first page, S1, (Fig 3) simply breaks the paragraphs into individual sentences and cleans it up (essentially removing punctuation) to make it easier to analyse. The second page, S2, (Fig 4) breaks the sentences into separate words and attaches the various markers to each word. It also summarises the data for each sentence, such as the count of each marker, the number of easy and difficult words (more or less than three syllables), and the number of long words (greater than 6 letters long). This is all used by the third page, stats, (Fig 5) where the statistical analysis is performed. The data summaries for each sentence are listed and various calculations are made for each and collated, such as the B per A ratios for each sentence. The amount of detail given is much greater than on the output page. A fourth page calculates word frequency, including by various markers, such

Sentence		1	2	3	4	5	6	7	8	9
1	Words	A	couple	of	weeks	ago	I	was	hiking	with
	CEFR	A1	B1	A1	A1	A2	A1	A1	A2	A1
	Grammar	func	noun	func	noun	adv	func	verb	noun	func
	School	EL	JH3		JH1	JH1	EL	EL		
	DC	DC		DC		DC		DC		DC
	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL		NGSL
	Syllable	1	2	1	1	2	1	1	2	1
	Length	1	6	2	5	3	1	3	6	4
	Frequency	art	1	12	2	2	39	16	2	5
Content Frequency			1		2	2		16	2	
2	Words	I	looked	and	looked	but	couldnt	find	him	
	CEFR	A1	A1	B1	A1	A1	B2	A1	A1	
	Grammar	func	verb	func	verb	func	func	verb	func	
	School	EL	EL	EL	EL	JH3		JH2	JH1	
	DC			DC		DC		DC	DC	
	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL		NGSL	NGSL	
	Syllable	1	1	1	1	1	2	1	1	
	Length	1	6	3	6	3	7	4	3	
	Frequency	39	4	11	4	4	2	3	5	
Content Frequency			4		4			3		
3	Words	He	had	been	with	me	for	so	long	that
	CEFR	A1	A1	A2	A1	A1	B1	A1	A1	B1
	Grammar	func	verb	verb	func	func	func	func	adj	func
	School	JH1				JH1	EL	JH1	EL	EL
	DC	DC	DC	DC	DC	DC	DC	DC	DC	DC
	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL
	Syllable	1	1	1	1	1	1	1	1	1
	Length	2	3	4	4	2	3	2	4	4
	Frequency	14	8	3	5	10	5	4	2	8
Content Frequency			8	3					2	

Fig 3 Screenshots of S1 page

Sentence		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Words	A	couple	of	weeks	ago	I	was	hiking	with	my	dog	on	a	mountain
	CEFR	A1	B1	A1	A1	A2	A1	A1	A2	A1	A1	A1	A2	A1	A2
	Grammar	func	noun	func	noun	adv	func	verb	noun	func	func	noun	func	func	noun
	School	EL	JH3		JH1	JH1	EL	EL			EL	EL	EL	EL	JH2
	DC	DC		DC		DC		DC		DC	DC	DC	DC	DC	DC
	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL	NGSL		NGSL	NGSL	NGSL	NGSL	NGSL	NGSL
	Syllable	1	2	1	1	2	1	1	2	1	1	1	1	1	2
	Length	1	6	2	5	3	1	3	6	4	2	3	2	1	8
	Frequency	art	1	12	2	2	39	16	2	5	19	8	3	art	6
	Content Frequency		1		2	2		16	2			8			6

Fig 4　Screenshots of S2 page

Basic Stats		min	max	average
total sentences	59			
total words	656	3.0	27.0	11.02
total syllables	828	2.0	33.0	14.03
ASL	11.12			
ASW	1.26			
total characters	2527	10.0	113.0	42.83
average words per sentence	11.0			
unique words	292			
Unique to total ratio	0.45			
Long words by syllable	27	0.0	2.0	0.46
short words by syllable	615	2.0	24.0	10.42
Elem	239	0.0	11.0	4.05
JH	208	0.0	11.0	3.53
HS	31	0.0	3.0	0.53
Dale-Chall	430.00	1.00	22.00	8.96
New General Service List	510.00	3.00	24.00	10.63
CEFR	537.00	3.00	26.00	11.19
A1	391.00	0.00	7.00	0.51
A2	106.00	0.00	2.00	4.19
B1	113.00	0.00	11.00	0.42
B2	30.00	0.00	2.00	0.12

Readability Stats		
Flesch	88.77	6th grade
Flesch-Kincaid	3.70	4th grade
ARI	2.27	first/second grade
Linsear	5.90	6th grade
LIX	15.72	very easy
Coleman-Liau	4.19	4th grade
VperSen	2.29	
BperA	0.12	
Con/Sen	0.43	
HS perE&J	0.07	
DC ratio	0.66	
NGSL ratio	0.78	
CEFR ratios	0.82	
A1	0.60	
A2	0.16	
B1	0.17	
B2	0.05	
Conjuctions	23.00	
CVS	2.75	
Content words	272.00	

Sentence	Words	Syllables	<3	>2	characters	Average length	short words	long words (6~)	verb
Averages:	11.02	14.03	10.42	0.46	42.83	3.90	8.66	2.36	2.29
1	24	33	23	1	96	4.00	18	6	2
2	8	9	8	0	33	4.13	5	3	3
3	19	20	19	0	60	3.16	18	1	5
4	9	11	9	0	35	3.89	7	2	2
5	23	32	22	1	97	4.22	16	7	7

Fig 5　Screenshots of stats page

as content or keyword, and the number of unique words in the text.

It should be noted that a shortcut was taken in those formulas that required using a 100 word sample from the text, such as the Linsear Write Readability Formula. As this would require a whole separate page to perform, the formula was applied to the whole text and then, essentially, divided by 100. What effect this had on the final results will be discussed below.

4. Effectiveness of the spreadsheet

It is clear that the sheet has various short-comings, but the objective is to produce a useful measure of a text's readability. There are a plethora of websites offering readability measures, so the same text was applied to them as to the Word Analysis spreadsheet for comparison. The text used was a long-form question from the Japanese 2020 Centre Exam for high school graduates. Question 5 was chosen as there were no numbers in the text. The full text is provided in the appendix.

The sites used for the comparison are: Readability Formulas (https://readabilityformulas.com/), Online Utility (https://www.online-utility.org/), Readable (https://app.readable.com/) and WebFX (https://www.webfx.com/tools/ read-able/). These were chosen as the top Google hits, and so most likely to be used.

First, what is interesting to note, is that the sites do not entirely agree with each other, in particular on the Flesch calculation. Of all the scores, that assigned by the Word Analysis spreadsheet was the highest, by as much as 8 points and as little as 1.17. However, all the scores fall within 6th grade reading level. Similarly, the Flesch-Kincaid scores calculated by the sites varied by quite a degree, from 3.90 to 4.72. Rounding would separate them by a whole school grade. The Word Analysis spreadsheet gave an even lower score of 3.70, still within the same reading grade as the lowest assigned by the websites.

Only one site, *Readability* provided a Linsear score, perhaps because it measures syllables, and as such requires a word list or a formula able to define syllables.[2] The spreadsheet assigned the text as a whole grade higher than the website. This may be because, as mentioned above, the formula was applied to the whole text and then rounded, whereas the *Readability* site (presumably) applied it to a 100 word sample.

One of the biggest differences between the spreadsheet and the sites was on the ARI measure. The spreadsheet gave 2.27, whereas the websites all gave 2.40 or 2.41. All the results, however, assign the text as second grade level. The Coleman-Liau Index gave a wide variety of results, in part because it is applied to a 100 word sample, and so the results very much depend on the sample chosen. Although the Word Analysis spreadsheet is the lowest, the outlier is *WebFX* with a score of 7.10, two whole grades higher than the next highest. Again, the spreadsheet applied the formula to the whole text and averaged it out, to get a score that underestimated the level compared to the websites.

Only *Readable* provided an LIX measure. Although almost ten points separates the two scores, the measure is "very easy" for the spreadsheet and "easy" for *Readable*. It also gave a CEFR measure, of A1, which corresponded with the score assigned by the Word Analysis spreadsheet, which found that 60% of all the words were A1 words.

Chart 1　Comparison of Word Analysis spreadsheet with online readability services

Measures	Word Analysis	Readability	Online Utility	Readable	WebFX
Flesch	88.77	86.60	80.73	85.20	86.00
Flesch-Kincaid	3.70	3.90	4.72	4.10	4.00
Linsear	5.90	4.90	-	-	-
ARI	2.27	2.40	2.41	2.40	2.40
Coleman-Liau	4.19	5.00	4.43	-	7.10
LIX	15.72	-	-	25.00	-
CEFR	A1	-	-	A1	-

Some of the sites give an overall readability score. *Readability Formulas* gave it a grade 4 (ages 8 to 9) and *WebFX* assigned it a grade level of 5 (ages 10 to 11), probably dragged up by the high Coleman-Liau score. This

corresponds well with the Word Analysis spreadsheet that, while not giving an overall rating, gives two 4th grade scores and two 6th grade scores, and a single 1st/2nd grade score, averaging out to an overall grade 4.

The *Readable* site also provides more data, some of which mirrors that measured in the spreadsheet. The number of unique words (292 out of 656 total words) is in agreement, but the syllable count is out by a significant 29 (857 in the website, 828 in the spreadsheet). Counts for nouns, verbs, adjectives and adverbs also disagree by varying amounts. This is probably because of the difference in how the words are defined in the word list, and how they are calculated in the website's formulas. The syllable count mismatch may be due to a variety of factors, including human error in one or both of the word lists used (or in the formula if used in *Readable*),

Lacking from the sites is the Japanese school grade measure. The spreadsheet found that a majority of words were elementary or junior high school, and the ratio of high school to lower words was 0.07, suggesting a fairly easy-to-read text. This would suggest that the standard formulas and word lists are accurate measures of readability for Japanese entrance exams, and also that the exam writers have access to a similar vocabulary list to that used in this analysis. However, further study is needed to refine this part of the analysis.

5. Conclusions

The spreadsheet is a far from ideal platform for such an exercise, but, when compared with online services, it performs adequately, although it does tend to under-estimate the reading level. It also allows for much greater scope to be applied. For example, none of the services gave useful data that hint at the text's complexity such as verbs per sentence or content words per sentence. Furthermore, as all the formulas were developed by US or Western-centric academics, the reading grade is for a native, L1, reader; only one site includes a CEFR score, which is specifically aimed at language learners. The Word Analysis spreadsheet, in addition to CEFR, adds the Japanese school vocabulary list.

Generally speaking, formulas have the advantage over lists in that they are light-weight, as no long lists of words are needed. However, they ignore the meaning and grammar function of the word, and the complexity of the word itself. Lists on the other hand, are cumbersome and have the problem of misapplying the meaning or level of a word, unless artificial intelligence is also incorporated into your application.

The main shortcoming of this spreadsheet is the limitations of the word list. As each word can only have a single item for each marker, words with multiple meanings and levels are misassigned in the analysis. However, as the comparison shows, this is not too much of a drawback in the basic formulas that ignore grammar and meaning. If there is a need to correct this issue, a word can be added manually in the input page as a 'missing word', and given the correct CEFR level, although this will also be applied to all incidences of this word in the text, regardless of meaning.

The next step in this research (which had to be delayed due to the COVID-19 outbreak) is to compile a set of short texts using graded language, and then assessing the readability for L2 learners. In addition, the role of conjunctions, word frequency and sentence complexity (verb per sentence, content words per sentence, etc) will also be accessed. Of interest is to see what grade is easier for the students to recall - the vocabulary learnt during junior high school or that learnt more recently, during senior high school.

Endnote

[1] *also, and, because, but, however, therefore, though, so.* In addition, there is *although*, as and others that are

not so commonly used. There are of course many others, but they can have alternative meanings, which are given preference, or are phrasal, such as *in order to*, adding a level of complexity to the spreadsheet not warranted by its usefulness.

[2] *Readable* also gave a "Lensear [sic] Write" score, but the result was a somewhat perplexing 97.4, so it was excluded from the table.

References and Resources

Asahi Newspaper, *Center Entrance Exam*, accessed 12 July 2020, https://www.asahi.com/edu/center-exam/shiken2020/mondai01day/english_05.html

CERF-J accessed on 5 February 2018, http://www.cefr-j.org/

COE, *Council of Europe Language Policy Portal,* accessed 21 July, 2020, https://www.coe.int/en/web/language-policy/home

Eigo Zuke [Pickled English] accessed 5 February 2018, https://www.eigo-duke.com/

English Profile *The CEFR for English,* accessed 3 October 2017, https://www.englishprofile.org/wordlists/evp

Klare, G.R. 1963, The measure of readability accessed 3 June 2020 https://babel.hathitrust.org/cgi/pt?id=mdp.39015004229558&view=1up&seq=7

New General Service List accessed 18 June 2018, http://www.newgeneralservicelist.org

Online Utility accessed 13 July 2020, https://www.online-utility.org/

Readability Formulas accessed 16 October 2017, https://readabilityformulas.com/

Readable accessed 13 July 2020, https://app.readable.com/

Sanseido, 2010, *The Grand Century English-Japanese Dictionary* Third Edition, Sanseido, Tokyo

Thorndike E L & Lorge I, 1944, *The teacher's word book of 30,000 words* accessed 15 June 2020, https://babel.hathitrust.org/cgi/pt?id=mdp.39015004965110&view=1up&seq=20

WebFX accessed 13 July 2020, https://www.webfx.com/tools/read-able/

Appendix

Text used in analysis, taken from the 2020 National Centre for University Entrance written English paper, question 5, as taken from the Asahi Newspaper, accessed at: (https://www.asahi.com/edu/center-exam/shiken2020/mondai01day/english_05.html):

A couple of weeks ago, I was hiking with my dog on a mountain when something unexpected happened and I lost sight of him. I looked and looked but couldn't find him. He had been with me for so long that it was like I was missing part of my soul.

Ever since that day, I had a strange feeling. It was beyond sadness - a feeling that I didn't quite understand, as if something was pulling me to go back to the mountain. So every chance I got, I grabbed my backpack to see if the mountain could offer me some sense of relief.

One Sunny morning, I stood at the foot of the mountain. Something felt different this day. "Please forgive me," I said out loud. "I'll find you!" I took a deep breath and began my journey with this mysterious pull growing stronger. After making my way along paths I thought I knew well, I realized I was somehow in an unfamiliar place. I panicked a little, lost my footing, and slipped. From out of nowhere, an elderly man came running toward me and helped me up.

Looking at his gentle, smiling face, I felt a sense of ease. The old man said he was looking for a way to

the top of the mountain, so we decided to climb together.

Soon the path began to feel familiar again. We talked about many things, including my dog. I told him that he was a German shepherd. When he was younger, he served briefly as a police dog but had to stop due to an injury. The man let out a lot of saying he had been a police officer for a short time, but he quit. He didn't say why. Later, he spent a long time as a bodyguard. He also had German roots. We laughed at these similarities.

Before we knew it, we reached a large open area and took a break. I told the man what had happened to my dog. "He had a tiny bell on his collar to scare away bears. We came to this very spot and saw a bear. It was looking back at us. I should have held my dog because, sensing danger, he chased after the bear. I couldn't find him after that. I should have been more careful."

As I was telling the story, the man's expression changed. "It wasn't your fault. Your dog just wanted to keep you safe," he said. "I'm sure Tomo would want to tell you this. Also, thank you for not giving up."

Tomo is my dog's name. Did I tell him this? The old man's comment rang in the air.

Before I could ask anything, the man proposed we hurry to get to the top of the mountain. I was planning to do this with my dog a few weeks ago. After two more hours of hiking, we reached the peak. I set down my backpack and we sat taking in the magnificent view. The old man looked at me and said, "Mountains offer truly magical experiences."

I looked around for a place to rest. I guess I was pretty tired, because I fell asleep right away. When I woke up, I noticed that the old man had disappeared. I waited, but he never returned.

Suddenly, in the sunlight, something caught my eye. I walked over and saw a small metal tag beside my backpack. It was the same silver name tag that my parents originally gave to my dog. Tomo it said.

It was then that I heard a familiar noise behind me. It was the ringing of a tiny bell. I turned around. What I saw caused so many emotions to rush over me.

After a while on the mountaintop, I attached the name tag to my old friend and carefully made my way home with the mountain's gift beside me. My soul felt very much complete.