

**TEXT DIFFICULTY: A COMPARISON OF  
READABILITY FORMULAE  
AND EXPERTS' JUDGMENT<sup>1</sup>**

**Sermin GÜLERER<sup>2</sup>**

**Abstract**

Teachers of English, librarians, researchers have been interested in finding the right text for the right reader for many years. In teaching Second Language (L2), text writers often try to fulfil the demand by simplifying the texts for the readers. The emerged term "readability" can be defined as "the ease of reading words and sentences" (Hargis, et al. 1998). The aim of this research was to compare the ways to find the right text for the right reader: traditional readability formulae (Flesch Reading Ease, Flesch-Kincaid Grade Level), Coh-Metrix Second Language (L2) Reading Index, which is a readability formula based on psycholinguistic and cognitive models of reading', and teachers' estimation of grade levels by using leveled texts in a web site. In order to do this, a selection of texts from a corpus of intuitively simplified texts was used (N30). Coh-Metrix Readability levels, Flesch Reading Ease, and Flesch-Kincaid Grade Levels of the texts were calculated via Coh-Metrix Web Tool. Three teachers of English were asked to decide the levels of the texts. When the relationship between Coh-metrix Readability Level, traditional formulae and the texts levels in the website was analysed via SPSS, it was found that there was weak negative correlation between Flesch-Kincaid Grade Level and the texts levels in the website (-,39). Additionally, there was weak negative correlation between the texts levels in the website and Flesch Reading Ease scores (-,41). However, there was moderate negative correlation between Coh-metrix Readability levels and the texts levels in the website (-,63), where Teacher1 and Coh-metrix Readability levels had very strong positive correlation (.95). It was identified that readability formulae can help L2 teachers when they select texts for their students for teaching and assessment purposes.

**Keywords:** *Readability, Coh-metrix, Flesch Reading Ease, Flesch-Kincaid Grade Level.*

**Article History:**

**Received**

06.05.2016

**Received in**

**revised form**

12.05.2016

**Accepted**

18.05.2016

**Available online**

27.07.2016

<sup>1</sup> This study is presented in CUELT-2016

<sup>2</sup> Instructor, Yeditepe University, e-mail: [sermingulerer@gmail.com](mailto:sermingulerer@gmail.com)

## 1. INTRODUCTION

### 1.1 Statement of the problem

Reading is regarded as a process which requires the interaction between the text and the reader. It is defined as the “process of receiving and interpreting information encoded in language form via the medium of print” (Urquhart & Weir, 1998, p22). In addition to this definition, Koda (2005) mentions that in order for reading comprehension to occur, “the reader extracts and integrates various information from the text and combines it with what is already known” (p4). However, if the reader’s knowledge is not sufficient to do this combination, s/he can have difficulty in comprehending the text. It is seen that comprehension is related to the readers’ level. In order to find which text is more suitable for which level, readability studies has begun. In other words it can be interpreted that readability emerges from the interaction between the text and the reader. According to Dale and Chall (1949) research in readability comes from the leading people in education and librarians who were interested in finding the right book for the right reader, i.e., finding the “readable” book for the readers. Back in 1949, they defined readability as follows:

The sum total (including all the interactions) of all the elements within a given piece of printed material that affect the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting (Dale & Chall , 1949,p 5).

In 1998 Hargis, Hernandez, Hughes, Ramaker, Rouiller, and Wilde simply defined readability as the “ease of reading words and sentences”. In addition, according to some other researchers “readability is in fact a product of reader, context and text.” (Wray& Janan, 2013).

As it can be understood, readability and how to assess readability in texts have been discussed for a long time. For instance, in the JSTOR academic publication database, readability can be seen as a research topic in almost 2200 papers which are published between 1980 – 2010 (Wray& Janan, 2013), yet no single Readability formula has been an answer to assessing readability. For example, Flesch Reading Ease score might be one of the most classical readability formulas. (Bu cümlelerin çıkarılmasını istiyorum)

Recently, Coh-Metrix Web Tool (<http://tool.cohmetrix.com/>) has been developed and it is claimed to be better than the other traditional readability formulae (Flesch-Kincaid Grade Level and Flesch Reading Ease in predicting the level of the texts) (Crossley, Allen & McNamara, 2011). It is stated that it not only includes a synthesis of the developments in psycholinguistic and cognitive models of reading, but also measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis (Crossley, Allen & McNamara, 2011).

### **1.2 Justification**

Experts in reading and writing can intuitively grade the reading passages. For instance, teachers select reading passages according to the level of their students and test designers select passages for the test takers according to their levels. According to

Fulcher (1997), readability is important in educational contexts as it helps the educators

to establish the appropriacy of the text. It is stated by Fulcher (1997) that teachers usually use sentence length and vocabulary as the indicators of text difficulty. However, selecting the passages according to the difficulty levels is time consuming for teachers and test designers. Moreover, it may not be a reliable method as the expertise and background of these people may affect their decisions. Thus, this study aims to find the most advantageous way to select the passages according to the level of difficulty, which can help teachers, syllabus designers, and test developers.

Readability formulae have two specific characteristics: One of them is the easiness of the text in understanding it and the other is the focus on quantification (Bailin & Grafstein, 2001). According to Bailin and Grafstein (2001) Flesch-Kincaid Grade Level and Flesch Reading Ease have these two characteristics. They take into consideration vocabulary difficulty and syntactic complexity; however, these concepts are much more complicated than they are thought to be (Bailin & Grafstein, 2001). Reading lists may reflect the common ground as they are made up on account of variety of texts. However, when considering specific texts for a group of people, these formulae may not work properly. In terms of syntactic complexity, these formulae associate sentence length with syntactic complexity; in other words, the longer a sentence is, the more complex it is, yet it is not true for all long sentences. Taking cohesion into consideration can be regarded as an innovation in readability formulae. Cohesion simply accounted for surface indicators which show the relation of sentences to one another in a text that can be measured by Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). It is claimed that there is some correlation between the readability formulae and teachers' perceptions of the texts' difficulty, in some studies (Bailin & Grafstein, 2001).

Consequently, this study compares these formulae with teachers' judgement. Although, the formulae do not take into account the readers' knowledge, the relationship between experts' knowledge and readability formulae is investigated in this study.

### **1.3 Significance**

There has been a debate on comparing the readability formulae to determine a better way to interpret the difficulty level of a text. Some researchers have studied the relationship between traditional readability formulae and text difficulty for L2 learners. Yet, these traditional formulae had mainly been developed for L1 texts (Crossley, Greenfield, McNamara, 2008). Therefore, some contradictory results have been found when these traditional readability formulae were used to explain L2 text difficulty (Crossley, Allen & McNamara, 2011). In one of these studies, Burke and Greenberg

(2010) compared different types of readability tools: formulae calculated by hand, Web tools, and tools which can be used by word processor in order to assess the difficulty level of adult literacy materials. They used extracts from six books only. They found that readability scores vary slightly depending on the readability tool. They recommend using multiple formulae and taking the average in order to provide optimum results (Burke & Greenberg, 2010). However, for the reasons mentioned above, optimum results might not be the best estimate of readability or text difficulty. In another formulae comparison study, Coh-Metrix Second Language (L2) Reading Index is compared to traditional readability formulae (Flesch-Kincaid Grade Level and Flesch Reading Ease) on a large corpus of texts intuitively simplified for language learners (Crossley, Allen & McNamara, 2011). The researchers found that Coh-Metrix L2 Reading Index provides significantly better indications of readability than traditional readability formulae (Crossley, Allen &

McNamara, 2011). Hiebert and Pearson (2010) studied a larger corpus and compared Lexiles and Coh-Metrix in beginner reading texts. They used 444 texts with seven levels of text difficulty. They concluded that although these systems can validate the general order of text levels and show some differences in the text difficulty, the usefulness of this information is not assured and further research should be conducted (Hiebert and Pearson, 2010). Rezaee and Norouzi (2011) evaluated the readability from a different perspective. They examined the relationship between readability of written materials, and learners' comprehension of them. They used Flesch and Fog Index in order to calculate the readability of the texts. They found significant correlation between the readability of written materials and learners' performance (Rezaee and Norouzi, 2011). When comparing the readability assessors and learners' performance, Crossley, Greenfield and McNamara (2008) used data of a former study by Greenfield (1999 as cited in Crossley, Greenfield and McNamara, 2008). The cloze test performances of Japanese learners' were compared to the readability scores calculated by traditional readability formulae and Coh-Metrix. Their results indicated that Coh-Metrix yields a more accurate estimate of text difficulty than traditional readability measures do (Crossley, Greenfield and McNamara, 2008). In the related literature, only Fulcher (1997) compared traditional readability formulae and expert judgment in his study. He concluded that when using the traditional reading formulae some aspects of the texts and features of the reader are ignored. In his study, five experts had difficulty in agreeing rank order of the difficulty of the texts. The correlation between agreed and predicted ranks by the readability formulae was low. It is stated that this is because of the fact that longer sentences do not always mean more difficult for the experts (Fulcher, 1997).

From the related literature it can be seen that no study has compared traditional readability formulae, a recent readability assessment method and expert judgment. Therefore, this study fills a gap in the literature with a larger corpus than most of the studies in the literature.

#### **1.4 Research Question**

This study attempts to answer the following question:

- 1) Is there a relationship between the reading formulae and teachers' estimation of grade levels?

### **1.5 Limitations and Delimitations**

In this study it is thought that limitations might occur related to technical problems, e.g., Coh-Metrix web tool may not work properly. In addition, to minimize the limitations, any additional material such as visuals, highlighted words or glossaries are removed from the texts in order to preclude their effects over readability, especially for the expert judgement. As mentioned by Fulcher (1997) font size, type, illustrations

and colours may affect the readability of a text. In this literature review, readability studies on readability levels of text books in different fields are not explained since they do not compare different readability formulae. Although, having interviews with the experts after they decide on the ranking order of the difficulty level of the text would be beneficial, because of the time constraints, a short questionnaire is given in order to find out the features the teachers take into consideration when deciding the difficulty of a text.

### **2. Methods**

The purpose in this paper is to compare the readability scores of traditional readability formulae, CohMetrix L2 Reading Index, and teachers' estimation of grade levels over intuitively graded texts. It is stated by Crossley, Allen, and McNamara (2011, p.89) that

the readability scores might be affected if the text simplification is done by manipulating "the factors related to cohesion, decoding, parsing, and meaning construction". In this research it is aimed to see if teachers' estimation of the text levels is also affected by text simplification. In addition, the aim is to see the correlation between these formulae and teachers' estimation.

#### **2.1 Corpus Selection**

In order to see the classification potential of the Readability formulae and Coh-Metrix L2 Reading Index, a selection of texts from a corpus of simplified news texts is used (N:30). The texts are taken from an English teaching website ([www.onestopenglish.com](http://www.onestopenglish.com)) which offers a number of graded texts. These news texts are based on articles from the Guardian Weekly, which is a British-based publication (Crossley, Allen, McNamara, 2011, News Lessons, nd). The texts were simplified under an intuitive approach by a team of authors, into three levels: advanced, intermediate and beginning (Crossley, Allen, McNamara, 2011). Ten topics were chosen randomly from general topics offered by the website. There are advanced, intermediate and beginning level texts for each subject.

#### **2.2 Selected Readability Formulas**

##### **2.2.1 CohMetrix Index**

Coh-Metrix is a computer tool which enables an analysis of text over 100 measures of cohesion, language, and readability (Graesser, McNamara, Louwerse, & Cai, Z., 2004). These measures are on *Descriptive, Text Easability Principle Component Scores, Referential Cohesion, Latent Semantic Analysis, Lexical Diversity, Connectives, Situation Model, Syntactic Complexity, Syntactic Pattern Density, Word Information, and Readability* (Graesser, McNamara, Louwerse, & Cai, Z., 2004). In this study

Readability Scores are used to be compared with other formulae.

### **2.2.2. Flesh-Kincaid Grade Level and Flesch Reading Ease**

Coh-Metrix calculates the formula based on the number of words sentence (sentence length) and the number of syllables per word (word length) which is the Flesch-Kincaid Grade Level (Crossley, Allen, and McNamara, 2011). In addition to this, Coh-Metrix calculates Flesch Reading Ease, which is based on the number of words per sentence (sentence length) and the number of syllables per word (word length) (Crossley, Allen, and McNamara, 2011). In this study both the Readability levels calculated by Coh-Metrix, Flesch Reading Ease, and Flesch-Kincaid Grade Level are used.

### **2.3 Teachers' Estimation of Grade Levels**

As mentioned before, the texts were simplified by a group of authors intuitively. In this study, English Language Teachers' estimation of text difficulty levels is investigated, too.

Three teachers of English, working at a University in Turkey participated in this study. They are asked to read the texts and decide whether the text is Level 1 – Beginner, Level 2 – Intermediate, or Level 3 – Advanced. Then, the teachers are asked to fill in an online questionnaire, which is based on Fulcher (1997) in order to find out the features taken into consideration when the teachers decide the difficulty of a text.

### **2.5 Procedures**

In this study, first, the texts are copied from the website which are chosen randomly. They are pasted in Microsoft Word Processor (MWP) and the irrelevant information like the glossary and the name of the author are deleted. Then, the texts are put into Coh-Metrix L2 Readability Index web tool (<http://tool.cohmetrix.com/>) and the Readability levels are calculated. In addition to these, the texts are sent to the teachers via email with a grading sheet which the teachers fill in. They choose which texts are at Level 1, Level 2 and Level 3. The teachers are not given any criterion for grading. Thus, these teachers are asked to fill in an online questionnaire which is based on Fulcher (1997) to reveal their ideas while deciding on the level of the text. This questionnaire is created online via google forms (<https://docs.google.com/forms/d/1veANemkVMcXrRw44HYjf3sdugWXmCm6zbbon6SnzJSI/viewform>) (See Appendix A).

### **2.4 Data Analysis**

In order to compare the Readability levels obtained from CohMetrix web tool, and teachers, Statistical Package for the Social Sciences (SPSS) is used. Correlational Analysis is done via SPSS with the aim of finding out the go togetherness. In addition, teachers' responses for the questionnaire are analysed.

## **3. RESULTS**

In this study, 30 texts are used. 21.815 words are analysed in these texts. These texts have minimum 435, maximum 979 words with a mean of 727 (See Table 1).

**Table 1. Descriptive Statistics**

	N	Minimum	Maximum	Mean
WordCount	30	435	979	727
Valid N (listwise)	30			

In addition, the levels defined in the website can be seen in Table 2. There were 10 Elementary (1), 10 Intermediate (2), and 10 Advanced (3) level texts (See Table 2).

**Table 2. Website Level**

	Frequency
1	10
2	10
3	10
Total	30

All of the teachers assigned the texts as 1, 2, or 3, Elementary, Intermediate, and Advanced respectively (See Table 3, Table 4, and Table 5).

**Table 1. Descriptive Statistics**

	N	Minimum	Maximum	Mean
WordCount	30	435	979	727
Valid N (listwise)	30			

In addition, the levels defined in the website can be seen in Table 2. There were 10 Elementary (1), 10 Intermediate (2), and 10 Advanced (3) level texts (See Table 2).

**Table 2. Website Level**

	Frequency
1	10
2	10
3	10
Total	30

All of the teachers assigned the texts as 1, 2, or 3, Elementary, Intermediate, and Advanced respectively (See Table 3, Table 4, and Table 5).

**Table 3. Teacher1**

	Frequency
1	10
2	10
Valid 3	10
Total	30

**Table 4. Teacher2**

	Frequency
1	10
2	10
Valid 3	10
Total	30

**Table 5. Teacher 3**

	Frequency
1	10
2	10
Valid 3	10
Total	30

When the level scores assigned by Flesch Reading Ease by Cohmetrix Web Tool are considered it can be seen that among the thirty texts no texts have the same score (See Table 6).

**Table 6. Flesch Reading Ease**

	Frequency	Percent
50,50	1	3,3
52,60	1	3,3
56,20	1	3,3
57,20	1	3,3
57,40	1	3,3
57,50	1	3,3
Valid 59,30	1	3,3
60,20	1	3,3
60,90	1	3,3
61,00	1	3,3
61,50	1	3,3
61,70	1	3,3
62,00	1	3,3
62,10	1	3,3



62,20	1	3,3
63,40	1	3,3
63,70	1	3,3
64,30	1	3,3
65,20	1	3,3
67,50	1	3,3
68,00	1	3,3
68,30	1	3,3
70,00	1	3,3
72,20	1	3,3
72,80	1	3,3
73,30	1	3,3
75,00	1	3,3
76,00	1	3,3
77,00	1	3,3
80,30	1	3,3
Total	30	100,0

**Table 7.** Flesch-Kincaid Grade Level

	Frequency	Percent
5,00	1	3,3
5,70	1	3,3
5,90	1	3,3
6,00	1	3,3
6,30	2	6,7
6,80	1	3,3
6,90	1	3,3
7,20	2	6,7

7,30	1	3,3
7,50	1	3,3
7,80	1	3,3
7,90	1	3,3
8,00	2	6,7
8,10	1	3,3
8,20	2	6,7
8,40	1	3,3
8,60	1	3,3
8,70	1	3,3
8,80	1	3,3
9,00	1	3,3
9,10	1	3,3
9,20	1	3,3
9,30	2	6,7
11,80	1	3,3
12,60	1	3,3
Total	30	100,0

In Table 7, it can be seen that among 30 texts, 2 texts have Grade level of 6,3, 2 texts have Grade level 7,2 and 2 texts have Grade level 8,2. Other than these 6 texts, no texts have the same grade level.

**Table 8. Cohmetrix Readability Frequency**

	Frequency	Percent
5,90	1	3,3
7,10	1	3,3
8,10	1	3,3
8,70	1	3,3
8,80	1	3,3
9,50	1	3,3
9,70	1	3,3
10,30	1	3,3

11,00	1	3,3
11,60	1	3,3
12,00	1	3,3
12,80	1	3,3
12,90	2	6,7
13,00	1	3,3
13,60	1	3,3
13,70	1	3,3
14,20	1	3,3
14,60	2	6,7
15,00	1	3,3
15,06	1	3,3
15,80	1	3,3
16,00	1	3,3
16,20	1	3,3
16,70	1	3,3
16,90	1	3,3
18,30	2	6,7
19,00	1	3,3
Total	30	100,0

In Table 8, it can be seen that among 30 texts, 2 texts have Cohmetrix Readability score of 12,9, 2 texts have Cohmetrix Readability score of 14,6 and 2 texts have Cohmetrix Readability score of 18,3. Other than these 6 texts, no texts have the same Cohmetrix Readability score.

**Table 9. Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
FRE	30	50,50	80,30	64,64	7,39
FKGL	30	5,00	12,60	7,97	1,63
Cohmetrix	30	5,90	19,00	13,07	3,45
Valid N (listwise)	30				

It can be seen in Table 9 that, the minimum score obtained by Flesch Reading Ease formula is 5,5, the maximum is 80,3, and the mean is 64,64 with a standard deviation of 7,39. The minimum level obtained by Flesch-Kincaid Grade Level formula is 5, the maximum is 12,6 and the mean is 7,9 with a standard deviation of 1,63. Moreover, the minimum score obtained by Cohmetrix Readability Level is 5,9, the maximum is 19 and the mean is 13 with a standard deviation of 3,45.

In order to answer the research question in this study, which was: *Is there a relationship between the reading formulae and teachers' estimation of grade levels?* Spearman's Rho rank order correlation is run in SPSS.

Table 10 summarizes the results:

Table 10. Correlations

			WebsiteLevel	Teacher1	Teacher2	Teacher3	Cohmetrix	FREC	FKGLC
Spearman's rho	WebsiteLevel	Correlation Coefficient	1,000						
		Sig. (2-tailed)							
		N	30						
	Teacher1	Correlation Coefficient	,95**	1,000					
		Sig. (2-tailed)	,000						
		N	30	30					
	Teacher2	Correlation Coefficient	,55**	,60**	1,000				
		Sig. (2-tailed)	,002	,000					
		N	30	30	30				
	Teacher3	Correlation Coefficient	,80**	,85**	,75**	1,000			
		Sig. (2-tailed)	,000	,000	,000				
		N	30	30	30	30			

	Cohmetrix	Correlation Coefficient	-,63**	-,63**	-,38*	-,54**	1,000		
		Sig. (2-tailed)	,000	,000	,036	,002			
		N	30	30	30	30	30		
	FRE	Correlation Coefficient	-,41*	-,39*	-,208	-,274	,81**	1,000	
		Sig. (2-tailed)	,026	,032	,271	,144	,000		
		N	30	30	30	30	30	30	
	FKGL	Correlation Coefficient	,39*	,39*	,172	,257	-,74**	-,97**	1,000
		Sig. (2-tailed)	,031	,035	,363	,170	,000	,000	
		N	30	30	30	30	30	30	30

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

It can be seen in Table 10 that, there is a very strong positive correlation between Teacher 1 and Levels stated in the website (0,95) and it is statistically significant ( $p>0,01$ ). There is moderate positive correlation between Teacher 2 and the website grade levels (0,55) which is statistically significant ( $p>0,01$ ). In addition, there is a strong positive correlation between Teacher 3 and the website grade levels (0,8) which is statistically significant ( $p>0,01$ ). It can also be seen in Table 10 that, there is a moderate negative correlation (-0,63) between the Cohmetrix Readability level and the Levels of the texts in the website which is statistically significant ( $p>0,01$ ). When the website levels and the Flesch Reading Ease scores are analysed, it can be seen that there is a weak negative correlation (-0,41) which is statistically significant ( $p>0,05$ ). Moreover, it can be seen that there is a weak negative correlation (-0,39) between website levels and Flesch-Kincaid Grade Level which is statistically significant ( $p>0,05$ ). Here, negative correlation means that while website levels increase, Flesch-Kincaid Grade Level decreases or vice versa.

There is a moderate positive correlation (0,60) between Teacher 1 and Teacher 2, and also there is a strong positive correlation (0,85) between Teacher 1 and Teacher 3 which are statistically significant ( $p>0,01$ ). When Teacher 1 and Cohmetrix Readability Levels are analyzed, it can be seen that, there is a moderate negative correlation (-0,63) between the Cohmetrix Readability levels and the Levels of the text assigned by Teacher 1 which is statistically significant ( $p>0,01$ ). Moreover, it can be understood that there is a weak negative correlation (-0,39) between the Levels of the text assigned by Teacher 1 and Flesch Reading Ease scores which is statistically significant ( $p>0,05$ ). Likewise, there is a weak positive correlation (0,39) between the Levels of the text assigned by Teacher 1 and Flesch-Kincaid Grade Level which is statistically significant ( $p>0,05$ ).

It can be said that, there is a moderate negative correlation (-0,54) between the Cohmetrix Readability levels and the Levels of the texts assigned by Teacher 1 which is statistically significant ( $p > 0,01$ ).

When the Cohmetrix Readability levels and the Flesch Reading Ease scores are analysed, there is a strong positive correlation between Cohmetrix Readability levels and the Flesch Reading Ease scores (0,81) which is statistically significant ( $p > 0,01$ ). In addition, it can be seen that, there is a strong negative correlation between Cohmetrix Readability levels and the Flesch-Kincaid Grade Level which is statistically significant ( $p > 0,01$ ). There is also very strong negative correlation (0,97) between the Flesch-Kincaid Grade Level and Flesch Reading Ease scores which is statistically significant ( $p > 0,01$ ).

The Levels of the texts assigned by Teacher 2 and 3, do not show any significant relationship with Flesch-Kincaid Grade Level and Flesch Reading Ease scores.

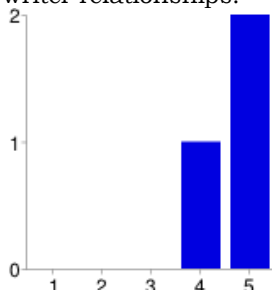
**3.1. Teachers' Estimation of Grade Levels**

Three teachers were asked to fill in a Microsoft Excel Sheet and send it to the researcher about the levels of the texts. Then, they were asked to participate an online questionnaire, which is based on Fulcher (1997). They choose what they take into consideration when deciding on the level of a text on a five level Likert scale (Strongly Agree – Strongly Disagree). In addition, they answered a question about their year of experience in teaching English. Table 11 shows the teachers' experience in teaching English.

Table 11. Teachers' Experience in Teaching English

Teacher:	1	2	3
Year of experience in Teaching	10	2	8

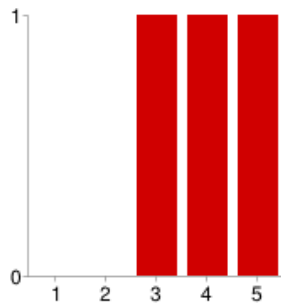
According to the results of the questionnaire, in order to decide the text difficulty of a text, they take into consideration linguistic structure, conceptual structure and reader writer relationships.



1	<b>0</b>	0%
2	<b>0</b>	0%
3	<b>0</b>	0%
4	<b>1</b>	33%
5	<b>2</b>	67%

Figure 1. Linguistic Structure

It can be understood from Figure 1, teachers agree that they take into consideration linguistic structure when they decide on the level of the text.



1	<b>0</b>	0%
2	<b>0</b>	0%
3	<b>1</b>	33%
4	<b>1</b>	33%
5	<b>1</b>	33%

Figure 2. Conceptual structure

It can be seen in Figure 2 that, 66% of the teachers took part in this study agree that they pay attention to conceptual structure when they decide on the text difficulty.

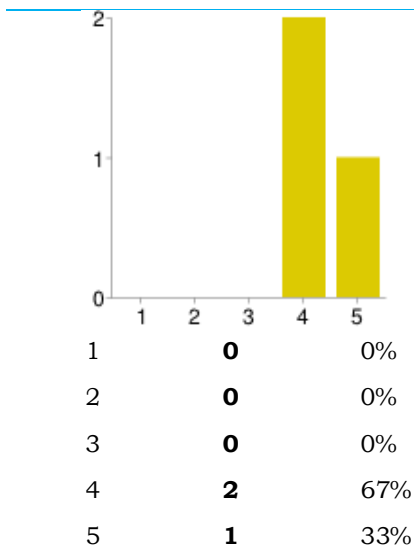


Figure 3. Reader-writer relationship

It can be seen in Figure 3 that, reader-writer relationship is taken into consideration by all of the teachers participated in this study while deciding the text difficulty.

Moreover, the teachers wrote in the space provided for additional comments that, they also look at the length of the paragraphs and length of the text when they decide on the difficulty of a text.

**4. DISCUSSION**

In this study, the relationship between intuitive text simplification by the website editors, teachers' estimation of grade levels, traditional readability formulae (Flesch-Kincaid Grade Level and Flesch Reading Ease) and Cohmetrix Readability Level are compared.

Since there is an ongoing debate in the literature on the comparison of teachers' estimation of grade levels and the Readability formulae, null hypothesis is formulated for the research question. The results indicate that the hypothesis is partially supported.

To start with the relationship of the teachers' estimation of grade levels, it is seen that there is strong correlation between the texts levels in the website and Teacher 1 and 3. However, there is a moderate correlation between Teacher 2 and Teacher 1 and 3 and the texts levels in the website. It can be seen that experience in teaching English might be a factor in deciding the difficulty of the texts. When teachers' experience in teaching is taken into consideration, it is seen that Teacher 1 has 10 years of experience while Teacher 3 has 8 years of experience. However, Teacher 2 has the minimum experience in teaching with 2 years (See Table 11).

When Flesch-Kincaid Grade Level and Flesch Reading Ease are compared, it is seen that there is a strong correlation between these two formulae. This can be because of the fact that they take into consideration same characteristics of the texts in terms of vocabulary difficulty and syntactic complexity (Bailin and Grafstein, 2001).

When the relationship between Cohmetrix Readability Level and these formulae is



analysed, it is seen that there is a strong correlation. However, there is weak correlation between the texts levels in the website and Flesch-Kincaid Grade Level and Flesch Reading Ease. In addition, there is a weak correlation between Teacher 1 and Readability Formulae. No correlation can be seen between Teacher 2 and 3 and Readability Formulae. This finding is different from the findings of Bailin and Grafstein (2001) who claimed that there is some correlation between the readability formulae and teachers' perceptions of the texts' difficulty. Nevertheless, this finding is in line with Fulcher (1997) who claimed that the correlation between agreed and predicted ranks by the readability formulae was low (Fulcher, 1997).

When the relationship between Cohmetrix Readability Level, traditional formulae and the texts levels in the website is analysed, it is seen that there is weak correlation between Flesch-Kincaid Grade Level and the texts levels in the website. In addition, there is also weak correlation between the texts levels in the website and Flesch Reading Ease scores. However, there is moderate correlation between Cohmetrix Readability and the texts levels in the website. This finding can be regarded as in line with the findings of Crossley, Allen and McNamara (2011) who found that Coh-Metrix L2 Reading Index performs significantly better than traditional readability formulas in predicting the level of the text and Crossley, Greenfield and McNamara (2008) who claim that Coh-Metrix yields a more accurate estimate of reading difficulty than traditional readability measures.

## 5. CONCLUSION

In this study, readability scores of traditional readability formulae, CohMetrix L2 Reading Index, and teachers' estimation of grade levels over 30 intuitively levelled texts in a website are compared. It is seen that there was a strong relationship between one of the teachers' estimation of the grade level and the level of the texts defined in the website.

It can be said that this study fills a gap in the literature with a larger corpus than most of the studies in the literature because no study had compared traditional readability formulae, a recent readability assessment method and expert judgment.

In addition, English language teachers who want to select texts for their students can use the information provided in this study. They can select the formulae accordingly in case they have difficulty in deciding the level of the text for the students. Moreover, the researchers who want to select corpus for their studies can get help from these formulae in line with the information given in this study.


## REFERENCES

- Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication, 21*(3), 285-301.
- Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review, 24*(1), 63-88.

- Burke, V., & Greenberg, D. (2010). Determining Readability: How to Select and Apply Easy-to-Use Readability Formulas to Assess the Difficulty of Adult Literacy Materials. *Adult Basic Education and Literacy Journal*, 4(1), 34-42.
- Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text Readability and Intuitive Simplification: A Comparison of Readability Formulae. *Reading in a foreign language*, 23(1), 84-101.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *Tesol Quarterly*, 42(3), 475-493.
- Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26(1), 19-26.
- Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25(4), 497-513.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Hargis, G., A. K. Hernandez, P. Hughes, J. Ramaker, S. Rouiller, and E. Wilde. 1998. *Developing quality technical information: A handbook for writers and editors*. Upper Saddle River, NJ: Prentice Hall.
- Hiebert, E. H., & Pearson, P. D. (2010). An Examination of Current Text Difficulty Indices with Early Reading Texts. Reading Research Report# 10-01. *Online Submission*.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.
- Rezaee, A. A., & Norouzi, M. H. (2011). Readability Formulas and Cohesive Markers in Reading Comprehension. *Theory and Practice in Language Studies*, 1(8), 1005-1010.
- Urquhart, S., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. New York: Longman.
- Wray, D., & Janan, D. (2013). Readability revisited? The implications of text complexity. *The Curriculum Journal*, 24(4), 553-562.

Appendix A

<https://docs.google.com/forms/d/1veANemkVMcXrRw44HYjf3sdugWXmCm6zbbon6SnzJSI/viewform>



*Teachers' Estimation of Readability Levels*

Dear Participant,

This questionnaire is a part of a research which investigates Teachers' Estimation of Readability Levels of Texts.

Thank you very much for agreeing to participate.

The information provided by you in this questionnaire will be used for research purposes only. It will not be used in a manner which would allow identification of your individual responses.

Your consent to participate in this project is assumed once you have completed the questionnaire.

If you have any questions or require further information about the project you may contact

Sermin Gulerer: [serming.yeditepe@gmail.com](mailto:serming.yeditepe@gmail.com)

\* Required

**Personal Information (will remain confidential) \***

Name

**Email Address \***

**Years of Experience in Teaching \***

**In order to decide the text difficulty of a text, I take into consideration :**

(Please choose 1 - Strongly Disagree to 5- Strongly Agree for the items below.)

1 2 3 4 5

Strongly Disagree ● ● ● ● ● Strongly Agree

**Linguistic structure \***

1 2 3 4 5

Strongly Disagree ● ● ● ● ● Strongly Agree

**Contextual structure \***

purpose and audience, context of use, information gaps, layout and visual support

1 2 3 4 5

Strongly Disagree      Strongly Agree**Conceptual structure \***

Degree of familiar and unfamiliar text content, informational relationships, degree of abstractness, and temporal structure

1 2 3 4 5

Strongly Disagree      Strongly Agree**Reader-writer relationship \***

Use of pronouns, tense and voice

1 2 3 4 5

Strongly Disagree      Strongly Agree**Other \***

Please mention