Journal of Education and Learning (EduLearn)
Vol. 18, No. 4, November 2024, pp. 1422~1432
ISSN: 2089-9823 DOI: 10.11591/edulearn.v18i4.21724

Text readability: its impact on reading comprehension and reading time

Zainurrahman[1,2], Fazri Nur Yusuf[1], Didi Sukyadi[1]
[1]Department of English Education, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Department of English Education, Faculty of Educational Sciences, Institut Sains dan Ilmu Kependidikan Kie Raha, Ternate, Indonesia

Article Info

Article history:

Received Jan 9, 2024
Revised Feb 19, 2024
Accepted Feb 25, 2024

Keywords:

Flesch reading ease
Flesch-kincaid grade level
Reading comprehension
Reading time
Text readability

ABSTRACT

Recently, the readability of texts has become the focus of reading research because it is believed to have implications for reading comprehension, which is of utmost importance in the field of English as a foreign language (EFL), particularly in the teaching, learning and assessment of reading comprehension. Unfortunately, the influence of text readability on reading comprehension (and reading time) has not been well studied in the EFL context. Most text readability studies are conducted in medical contexts, but these studies are often limited in predicting readability scores for sample texts. To address this gap, the current study aimed to evaluate the influence of text readability levels (based on the Flesch-Kincaid grade level (FKGL)) on students' reading comprehension and reading time. Data were collected through reading test and analyzed using SPSS version 22. The Friedman test revealed that the distribution of students' reading comprehension score (X2=197.532, p=0.000) and reading time (X2=215.323, p=0.000) are different in each text, suggesting that the readability of texts has a significant influence on both. This study contributed to the practices of reading instruction and assessment. Limitations and suggestions for further research are briefly discussed.

Corresponding Author:

Zainurrahman
Department of English Education, Faculty of Language and Literature Education
Universitas Pendidikan Indonesia
Sersan Surip Street, Djayaredja 2 Alley, Number 85, Bandung City, West Java, Indonesia
Email: zainurrahman@upi.edu

1. INTRODUCTION

Text readability, the main goal of which is to assist authors in creating easy-to-read texts [1], [2] to convey important and practical information to readers with diverse knowledge and educational backgrounds [3], and allows the design of the best possible match between readers and texts [4], has been a recent focus in reading-related studies. In foreign language learning, the readability of texts plays a crucial role, especially when teaching and assessing reading comprehension [5]. Unfortunately, text readability studies appear to be more ubiquitous in medical contexts than in English as a foreign language (EFL) contexts. The abundance of studies on the readability of texts in a medical context could be attributed to the fact that it is a factor that influences readers' understanding of medical instructions and that accurate understanding of the instructions can have life-threatening consequences. In EFL contexts (or reading contexts in general), the readability of texts plays a dual role. On the one hand, it helps teachers to simplify texts if they are too difficult for students. On the other hand, for a reading comprehension test, teachers must only use texts with an appropriate level of difficulty [6], [7]. In other words: the readability of texts is a

crucial aspect of both teaching and assessing reading comprehension. The paucity of studies on text readability in the EFL context is ironic. In the meantime, this type of study would provide valuable insights into factors influencing reading comprehension and provide important considerations for effective reading instruction and assessment.

Text readability is a term often associated with readability formulas (e.g., Flesch reading ease (FRE) and Flesch-Kincaid grade level (FKGL))-statistical tools developed by experts to objectively assess the relative difficulty of a given text. Text components such as the number of sentences, the number of words per sentence, the number of syllables per word and the number of words with three or more syllables are the basis for the readability of the text [8]; no external factors (e.g., students' ability to understand the text) are taken into account; although these factors are very important for reading comprehension. However, in the area of text readability, the text components that are considered "quantitative parameters of the text" can help or hinder communication [3]. In other words, these are some functions that make the text readable [6].

There are many formulas for text readability, but the most popular are probably FRE and FKGL. One of the reasons for their popularity may be that they are available in popular text editors such as Microsoft Word [9]. Many studies of text readability in medical contexts also used either or both. Both FRE and FKGL were developed by Flesch [1], especially the first one, although some of medical researchers recommended to use other readability formula such as the simple measure of gobbledygook (SMOG) [10]. Other scholars also commented that existing readability formulas are not sophisticated enough to predict the difficulty level of a given text due to its limitation in understanding differences in reader's dialects and cultural backgrounds [11]. Of course, different readability formulas were formulated under different perspectives and when they are compared to each other, inconsistent results will be emerged [12].

The main idea behind the development of FRE (and FKGL as an evolution of FRE) is to adapt the "writing style" to the "grade level" of the intended reader. Flesch himself writes, "Make sure you know who you are writing for," [1] and then provides a rationale for text readability and instructions for writing a readable text. His 1949 book laid the foundation for the FRE. Later, the FKGL was developed when the US Navy adopted the FRE in 1976 [2]. The reason for the FKGL development was that the FRE did not include a "grade level," which is misinformation; the corresponding grade level in FRE already existed [1]. The US Navy has simplified the formula [8]. Consequently, the interpretation of the FRE and FKGL results was different. The following applies to the FRE: the higher the score, the "easier" the text is. With FKGL, on the other hand, the higher the score, the more "difficult" the text is. This happens because the FRE emphasizes "ease" while the FKGL emphasizes "grade level". Flesch [1] introduces us to the "hard way" with the readability formulas, but today the readability of the text can easily be calculated automatically using software. Between the two formulas, we chose to use the FKGL simply because it is inconvenient to use both; There is no particular reason to choose the FKGL over the FRE in this study other than the linearity between the FKGL score and the associated text difficulty.

It is difficult to establish whether text readability has a direct influence on reading comprehension without continuous empirical studies in the second and/or foreign language context. However, studies on the readability of texts in a medical context have shown that the effects are obvious and an accurate understanding of the (medical) texts is crucial, especially after the COVID-19 pandemic when most patients turn to websites for medical information [13]. There is now a disconnect between the wealth of medical and health information available on the internet and patient literacy [14], and sourcing this information from the Internet means there is less advice from medical experts. To avoid unwanted risks, continuous evaluations of the readability of medical texts are carried out. Surprisingly, studies of the text readability of electronic (and printed) medical and health instructions have, from time to time, resulted in "warnings" that those texts needed to be revised [15]–[19]. Due to the fact that most medical texts are outside the expected readability level, some medical scholars recommended to reduce the use of complex words and sentence length in the medical texts [20]. This phenomenon is not confined in the medical context; in the academic context broadly, the readability of scientific texts has also been found decreasing over time [21].

For our information, the American Medical Association (AMA) has published a guideline that health education or medical teaching materials should have a readability level equivalent to 0^{th} to 5^{th} grade. The scale used can be either the FRE or the FKGL. In the EFL context, readability of texts is viewed as a variable that can influence comprehension. Native English speakers may have less difficulty reading English texts, but things are different for non-native speakers. Apart from the fact that non-native speakers have comparatively lower cognitive and metacognitive skills in reading compared to native speakers [22], non-native speakers tend to mentally translate the text by relying only on their foreign language knowledge, which turns out to be a problem in foreign language reading [6], [23]. Studies have suggested that metacognition and metacomprehension knowledge correspond to the information level of the text [24], and texts with low readability have a direct impact on the cognitive load of reading and can hinder comprehension. For example, to understand a sentence, theoretically, one must visually process the individual words, identify and access their phonological, orthographic, and semantic representations, and

connect these representations to develop an understanding of the underlying meaning of the sentence. For these processes to be successful, many factors come into play, and one of them is text properties [25], in which the level of readability plays a role.

Given the possibility that text readability has an impact on comprehension, it is difficult to accept that studies of text readability in the EFL context are extremely scarce. Several studies have been conducted focusing on matching the readability level of reading materials to the language proficiency levels of intended readers [26]–[28], text readability and readers' perceived readability [29], and the interaction between text readability and the level of task difficulty [22]. We found only one study, which was slightly related to the aim of this study, but the study was more focused on assessing the relationship between students' challenge in understanding inference text-based questions and the readability of the reading text [30]. The study concluded that there is a significant relationship between the readability of the text, the difficulty of the questions and the students' understanding. However, the context does not necessarily mean that the readability of texts has an influence on students' reading comprehension. In summary, there is a need to conduct further studies on text readability in the EFL context.

The current study was conducted to address the serious gap mentioned above. The main aim of this study is to examine the effects of readability and text comprehension in an EFL reading context. In addition, we further extended the study to examine its effects on reading time (the time spent understanding the text). The reason for this expansion was that time is a crucial factor in standardized tests such as test of English as a foreign language (TOEFL) or International English Language Testing System (IELTS). In general, reading time is based on text length, measured in words per minute (WPM), which is called "reading rate". A comprehensive meta-analysis conducted by Brysbaert [31] concluded that the normal reading rate for silent reading differs between non-fiction and fiction, with the former being 238 WPM and the latter 260 WPM. However, since the level of readability (as one of the text properties) can have a significant impact on the cognitive load of reading, the concept of reading rate seems to be irrelevant, especially when reading foreign languages. Therefore, it is quite important to extend this study to reading time. In summary, this study addressed the following research questions: i) RQ1: does the readability of texts have an impact on students' reading comprehension? and ii) RQ2: does the readability of texts have an impact on students' reading time? The results of this study would provide valuable insights for both the teaching and assessment of reading skills in various situations.

2. METHOD

To answer the research questions, the current study analyzed the variance between students' reading comprehension scores and reading time for texts with different readability levels. The participants were 24 EFL students from a private university in Indonesia, aged 22 to 35 years (M=28, SD=4.47) with English proficiency, as measured by paper-based TOEFL, between 425 and 499 (M=463, SD=4.17). We invited 35 students, but 11 did not respond to the invitation. Participants read 10 nonfiction texts (consisting of research summaries and short essays) of varying length (M=212, SD=42.46) and readability levels, measured using the FKGL range from easy (FKGL=5.17) to extremely difficult (FKGL=18.84) (M=11.66, SD=4.87), as shown in the Table 1.

Table 1. Texts, lengths, FKGL scores, and categories used in the study

Text *n*	Length (word)	FKGL score	Category
T1	242	18.84	Extremely difficult
T2	261	17.66	Extremely difficult
T3	224	16.55	Very difficult
T4	211	14	Fairly difficult
T5	190	12.17	Difficult
T6	185	9.29	Fairly easy
T7	159	8.51	Fairly easy
T8	259	7.57	Fairly easy
T9	140	6.91	Fairly easy
T10	249	5.17	Easy

All texts were printed in black with Times new roman font and 12pt font size. The reason for this specification is to prevent reading problems related to text representation, as some scholars have suggested that text representation could affect readability [32]. Each text was followed by six text-based conclusions that students must identify as either true (T) or false (F) statements based on the accompanying text. This technique for assessing reading comprehension is known as a "dichotomous item" [6]. To address the issue

that this test might give students a chance to be correct by simply guessing, the statements were inferential in nature, with three true statements and three false statements randomly ordered. They had to read and identify the statements only after they had read and understood the accompanying text. Their work was evaluated based on percentage of accomplishment. To record their reading time, they had to use their stopwatches available on their smartphones. They had to write the length of their reading directly on the text they were reading. The reading time was further converted into seconds. There was no time limit for reading the text, so students took as much time as they needed.

After collecting data, students' reading comprehension scores and reading time were tested for normal distribution before analysis of variance (ANOVA) was conducted. We used ANOVA because the data came from a single group in different situations [33]. The result of the normality test showed that the data was not normally distributed as shown in Figure 1. Therefore, a nonparametric test was performed. The Friedman two-way ANOVA test by ranks was chosen because it is an ideal statistic for repeated measures experiments to determine whether a particular factor has an effect [34]. For a more targeted analysis, we separately analyzed students' reading comprehension scores and reading time using the Friedman test.

The data analyses performed in this study, from normality test to repeated measures ANOVA, were performed using SPSS version 22. For each analysis, a test hypothesis was formulated with a significance level of 0.05. In the following discussion, we coded the text's readability levels as X, students' reading comprehension as Y_1, and students' reading time as Y_2. Due to limited space, we could only present a brief pairwise comparison for each analysis result.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Reading Texts	Statistic	df	Sig.	Statistic	df	Sig.
Reading Time	Text 10	.134	24	.200*	.933	24	.115
	Text 9	.140	24	.200*	.937	24	.142
	Text 8	.174	24	.058	.914	24	.044
	Text 7	.112	24	.200*	.952	24	.301
	Text 6	.132	24	.200*	.956	24	.368
	Text 5	.202	24	.013	.886	24	.011
	Text 4	.135	24	.200*	.946	24	.217
	Text 3	.190	24	.025	.916	24	.047
	Text 2	.187	24	.030	.901	24	.023
	Text 1	.201	24	.013	.913	24	.041
Reading Comprehension	Text 10	.519	24	.000	.393	24	.000
	Text 9	.401	24	.000	.616	24	.000
	Text 8	.427	24	.000	.647	24	.000
	Text 7	.399	24	.000	.679	24	.000
	Text 6	.304	24	.000	.733	24	.000
	Text 5	.443	24	.000	.608	24	.000
	Text 4	.443	24	.000	.573	24	.000
	Text 3	.312	24	.000	.782	24	.000
	Text 2	.283	24	.000	.768	24	.000
	Text 1	.431	24	.000	.621	24	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 1. The result of the normality test

3. RESULTS AND DISCUSSION

This section summarizes and discusses the key findings of the current study, divided based on the research questions posed earlier in this article. In the first subsection, we examine the influence of text readability on the students' reading comprehension. Meanwhile, in the second subsection, we examine the influence of text readability on the students' reading time.

3.1. Text readability and students' reading comprehension

It was found that a specific study examining the influence of the readability level of texts on students' reading comprehension is extremely rare, especially in the EFL context. From 1994 to 2023, rigorous studies in EFL reading instruction focused more on technology-based instruction, critical thinking, and literacy skills [35]. Some studies have attempted to determine the suitability of EFL reading texts for intended EFL readers. In the medical context, numerous studies have been conducted to assess the suitability of medical texts for patients worldwide. However, whether text readability has an impact on students' reading comprehension has not been assessed.

To assess whether text readability has an impact on students' reading comprehension, especially in a quantitative study, we need to examine whether there is a significant difference in the distribution of students' reading comprehension among texts with different readability levels. If the distribution shows a significant difference, the readability of the text has an impact on students' reading comprehension. Otherwise, there is no connection between the readability of the text and the students' reading comprehension. This simple logic led us to formulate and test the following hypothesis:

H_0: The distribution of students' reading comprehension of text 1 to text 10 are the same

Due to the normality of the data, we used a nonparametric statistic, the Friedman test, to test this hypothesis under the significance level of 0.05. To draw the conclusion, we use this logic: if the p-value is less than the alpha value ($p<0.05$), the null hypothesis is rejected. Otherwise, it is accepted.

The following figures are the results of the statistical analyses that we carried out from which we can draw a conclusion. Figure 2 displays the descriptive statistic analysis, Figure 3 shows the mean rank, while Figure 4 is the Friedman test result. An important reminder that there were 10 texts with different readability levels calculated by using the FKGL formula, which were then classified into extremely difficult to easy (text 1 is extremely difficult, text 10 is easy). Additionally, we have coded text readability as X and students' reading comprehension as Y_1.

Figure 2 demonstrates that the mean distance across the texts is visible, with the exception of "neighboring texts" like texts 1 and 2 or 9 and 10. The students' reading comprehension scores are represented by these means, and the ascending order of the ranks indicates that the students performed better on the later texts than the earlier ones. The chart of the mean rank below as shown in Figure 3 is consistent with descriptive statistics as shown in Figure 2 in displaying the pattern. While Figures 2 and 3 have described that the scores of students' reading comprehension are different among the texts, whether the real difference in the distribution is present should be referred to the Friedman test as shown in Figure 4.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Text 1 Reading Comprehension Score	24	38.8889	9.41100	33.33	66.67	33.3333	33.3333	50.0000
Text 2 Reading Comprehension Score	24	44.4444	10.61703	33.33	66.67	33.3333	50.0000	50.0000
Text 3 Reading Comprehension Score	24	52.7778	10.61703	33.33	66.67	50.0000	50.0000	66.6667
Text 4 Reading Comprehension Score	24	61.8056	7.73843	50.00	66.67	50.0000	66.6667	66.6667
Text 5 Reading Comprehension Score	24	68.7500	7.47379	50.00	83.33	66.6667	66.6667	66.6667
Text 6 Reading Comprehension Score	24	76.3889	9.72653	66.67	100.00	66.6667	83.3333	83.3333
Text 7 Reading Comprehension Score	24	84.7222	8.39350	66.67	100.00	83.3333	83.3333	83.3333
Text 8 Reading Comprehension Score	24	86.1111	8.02572	66.67	100.00	83.3333	83.3333	83.3333
Text 9 Reading Comprehension Score	24	93.7500	8.24226	83.33	100.00	83.3333	100.0000	100.0000
Text 10 Reading Comprehension Score	24	97.9167	5.63053	83.33	100.00	100.0000	100.0000	100.0000

Figure 2. Descriptive statistics of Y_1 across X

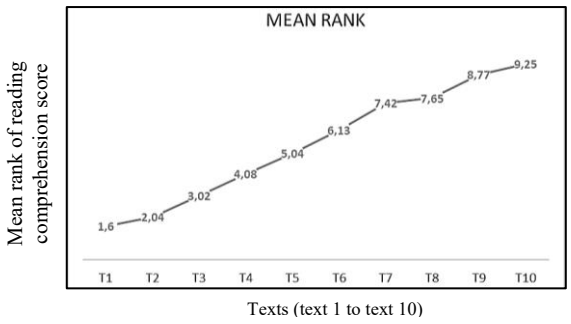
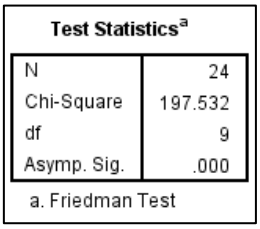


Figure 3. Mean rank of Y_1 across X



Test Statistics[a]

N	24
Chi-Square	197.532
df	9
Asymp. Sig.	.000

a. Friedman Test

Figure 4. Friedman test result of Y_1

Evaluation: a Friedman test was conducted on 24 students to examine the influence of 10 texts with different readability levels (X) on students' reading comprehension (Y_1). The results showed that the readability of the texts read led to statistically significant differences in students' reading comprehension (X^2=197.532, p=0.000). The p-value is lower than the alpha (p<0.05), indicating strong evidence against the null hypothesis that students' reading comprehension distributions from text 1 to text 10 are the same. Therefore, the null hypothesis is rejected. The pairwise comparison is presented in Table 2.

Table 2. Pairwise comparison of Y_1 across X

Texts	1 (F)	2 (α)	3 (α)	4 (α)	5 (α)	6 (α)	7 (α)	8 (α)	9 (α)	10 (α)
1 (α)		1.000	1.000	.205	.004*	.000*	.000*	.000*	.000*	.000*
2 (F)	-.438		1.000	.877	.027*	.000*	.000*	.000*	000*	.000*
3 (F)	-1.417	-.979		1.000	.935	.017*	.000*	.000*	.000*	.000*
4 (F)	-2.479	-2.042	-1.062		1.000	.877	.006*	.002*	.000*	.000*
5 (F)	-3.438*	-3.000*	-2.021	-.958		1.000	.296	.130	.001*	.000*
6 (F)	-4.521*	-4.083*	-3.104*	-2.042	-1.083		1.000	1.000	.111	.016*
7 (F)	-5.812*	-5.375*	-4.369*	-3.333*	-2.375	-1.292		1.000	1.000	1.000
8 (F)	-6.042*	-5.604*	-4.625*	-3.562*	-2.604	-1.521	-.229		1.000	1.000
9 (F)	-7.167*	-6.729*	-5.750*	-4.688*	-3.729*	-2.646	-1.354	-1.125		1.000
10 (F)	-7.646*	-7.208*	-6.229*	-5.167*	-4.208*	-3.125*	-1.833	-1.604	-.479	

Table 2 is a modified table for pairwise comparison of students' reading comprehension (Y_1) across texts with different readability levels (X). The table shows the Friedman test statistics (F, vertical column) and the adjustment significance (α, horizontal column). Of the pairwise comparisons, the * sign indicates a significant difference between the compared texts. It turns out that there are significant differences, particularly at distances of 3 to 5 levels of text readability (e.g., texts 1 and 5, texts 2 and 5, texts 3 and 6). Although not all comparisons produce a significant result, the minus sign (when read vertically) indicates that students found second sample to be easier to understand than first sample. For example, while the difference between texts 6 and 7 is not significant (α=1,000), the students still found text 7 easier than text 6 (-1,292).

As mentioned above, there are few studies on the influence of text readability on students' reading comprehension. Despite the few relevant studies, this finding confirms the results of the study by Dirgantari and Susantiningdyah [30] that there is a significant relationship between the readability and comprehension of texts. While their study did not cover the trend of the correlation or the effects, we now have a clear picture: the higher the FKGL score, the lower the readability and the more difficult the text. The possible reason why students found texts with a higher FKGL score more difficult to understand could be the discrepancy between the text difficulty and their language skills since they are not native speakers. It has been found that non-native speakers tend to have lower cognitive and metacognitive skills when reading compared to native speakers [22]. Non-native speakers tend to translate the texts they read in their heads and rely only on their knowledge of the target language, which has proven to be a problem when reading foreign languages [6], [23]. The problem could be worse when they are also poor readers who the comprehension strategies have not attained the degree of automaticity like fluent readers [36]. The way the information is presented in the texts with higher FKGL scores may also have increased the cognitive load on students and impaired comprehension. It can be assumed that difficult texts affect students' willingness to read, thereby affecting their comprehension. This is debatable.

One study found that there is no significant relationship between students' reading attitudes and reading performance [37] while another study revealed that foreign language readers are unlikely to absorb what they read due to their lower appreciation of the texts [38]. We hypothesized that difficult texts (represented by higher FKGL scores) affect EFL students' reading comprehension due to the lack of automaticity in reading, linguistic inadequacy, and the high cognitive load associated with their bottom-up

reading strategy. Congruent with this hypothesis is the observation made by some scholars that text comprehension is an active and dynamic process that enables the reader to construct coherent mental models and draw conclusions from the text while staying within the bounds of the amount of information that can be stored in working memory [39]. This implies that difficult texts relative to the reader's skill level can tax students' cognitive abilities as they attempt to understand the meaning.

This result shows us that the correct selection of reading materials for teaching or assessing students' reading comprehension is crucial in the EFL context [6], [7]. Reading should be seen as an interaction between students and the texts [40]. And for good reason, this interaction would be more effective if both parties were "on the same level." The participants in this study were undergraduate students, while most of the texts that fall into the FKGL category were "extremely difficult" texts related to higher education degrees. Of course, training students to read higher difficulty texts can motivate them to improve their reading skills. However, the use of texts with incompatible levels in the assessment would make it difficult to justify the assessment results. Finally, we noticed that studies of text readability in a medical context rarely assessed patients' understanding of the medical instructions they tested. However, the current study provided strong evidence that text readability has a significant impact on readers' comprehension, informing them that their studies were relevant and important.

3.2. Text readability and students' reading time

Time is an important factor in reading foreign languages because students must take time to grasp the correct meaning of the text they are reading. We briefly discussed about how native and non-native English speakers differ in reading strategy and how poor readers (EFL students with linguistic inadequacies) struggle with reading comprehension. When it comes to reading assessments, whether in the classroom or on standardized tests, time is usually limited. Unfortunately, since no research has been conducted to date on whether the level of text readability influences reading time, experiences taught us that reading teachers limit reading assessment time by referring to the time allocation specified in the lesson plan.

In this section, we examine the impact of text readability levels on students' reading time. As with the previous analysis, we examine whether there is a significant difference in the distribution of students' reading time (Y_2) among texts with different readability levels (X). If the distribution shows a significant difference, the readability of the text has an impact on students' reading time. Otherwise, there is no connection between the readability of the text and the students' reading time. Based on this logic, we formulated a hypothesis as follows:

H_0: The distributions of students' reading time from text 1 to text 10 are the same

Due to the normality of the data, we used a nonparametric statistic, the Friedman test, to test this hypothesis under the significance level of 0.05. To draw the conclusion, we use this logic: If the p-value is less than the alpha value ($p<0.05$), the null hypothesis is rejected. Otherwise, it is accepted. Figure 5 shows that, apart from "neighboring texts" such as texts 1 and 2 or 9 and 10, the mean distance between the texts is visible. These means show the students' reading times, and the rankings' descending order suggests that the students required more time to read the earlier texts than the later ones. The pattern is displayed in a way that is consistent with the descriptive statistics in the mean rank chart in Figure 6. Figures 5 and 6 have indicated that the reading times of the students vary depending on the text; however, Figure 7 shows the Friedman test, which should be used to determine whether a true difference in the distribution is present.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Text 1 Reading Time	24	122.2083	6.07188	113.00	132.00	117.0000	119.5000	127.0000
Text 2 Reading Time	24	115.2083	2.94853	110.00	119.00	113.0000	114.5000	118.0000
Text 3 Reading Time	24	111.5000	3.33623	106.00	116.00	108.2500	111.5000	114.7500
Text 4 Reading Time	24	106.3750	3.97615	101.00	115.00	102.5000	105.5000	109.7500
Text 5 Reading Time	24	94.8333	2.47890	90.00	102.00	94.0000	94.5000	96.0000
Text 6 Reading Time	24	89.3750	3.38555	83.00	95.00	86.0000	90.0000	92.7500
Text 7 Reading Time	24	86.4583	4.01063	80.00	93.00	83.0000	86.0000	89.7500
Text 8 Reading Time	24	83.0000	4.36388	77.00	92.00	80.0000	82.0000	86.0000
Text 9 Reading Time	24	73.0000	6.36567	62.00	82.00	68.0000	73.5000	79.2500
Text 10 Reading Time	24	63.4167	6.26874	55.00	75.00	58.2500	62.5000	69.5000

Figure 5. Descriptive statistics of Y_2 across X

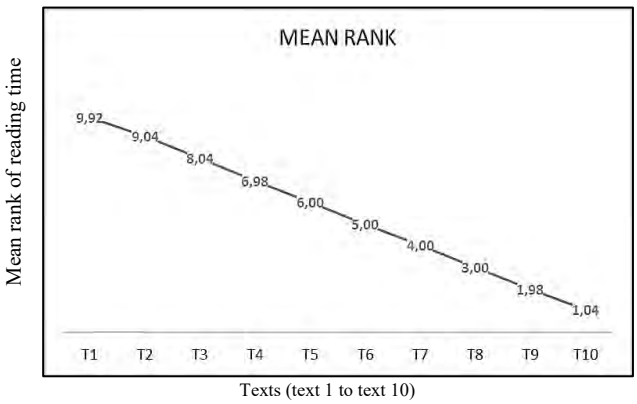


Figure 6. Mean ranks of Y_2 across X

Test Statistics[a]

N	24
Chi-Square	215.323
df	9
Asymp. Sig.	.000

a. Friedman Test

Figure 7. Friedman test result of Y_2

Evaluation: a Friedman test was conducted on the reading time of 24 students to examine the influence of 10 texts with different readability levels (X) on the students' reading time (Y_2). The results showed that the readability of the texts read led to statistically significant differences in students' reading time (X2=215.323, p=0.000). The p-value is lower than the alpha ($p<0.05$), indicating strong evidence against the null hypothesis that the distributions of students' reading time from text 1 to text 10 are equal. Therefore, the null hypothesis is rejected. The pairwise comparison is presented in Table 3.

Table 3. Pairwise comparison of Y_2 across X

Texts	1 (F)	2 (α)	3 (α)	4 (α)	5 (α)	6 (α)	7 (α)	8 (α)	9 (α)	10 (α)
1 (α)		1.000	1.000	.035*	.000*	.000*	.000*	.000*	.000*	.000*
2 (F)	.875		1.000	.832	.023*	.000*	.000*	.000*	.000*	.000*
3 (F)	1.875	1.000		1.000	.877	.023*	.000*	.000*	.000*	.000*
4 (F)	2.938*	2.062	1.062		1.000	1.000	.029*	.000*	.000*	.000*
5 (F)	3.917*	3.917*	2.042	.979		1.000	.995	.027*	.000*	.000*
6 (F)	4.917*	4.042*	3.024*	1.979	1.000		1.000	.995	.025	.000*
7 (F)	5.917*	5.042*	4.042*	2.979*	2.000	1.000		1.000	.935	.032*
8 (F)	6.917*	6.042*	5.042*	3.979*	3.000*	2.000	1.000		1.000	1.000
9 (F)	7.938*	7.062*	6.062*	5.000*	4.021*	3.021*	2.021	1.021		1.000
10 (F)	8.875*	8.000*	7.000*	5.938*	4.958*	3.958*	2.958*	1.958	.938	

Figures 5 and 6 show that there are differences in students' reading time between texts, with the exception of neighboring texts such as texts 1 and 2 or texts 9 and 10. The trend is different from the previous analysis. If the analysis of student reading comprehension across texts shows average ranks in ascending order, the average ranks of student reading time across texts are shown in descending order. The descending order shows that students spent more time reading and understanding texts with higher FKGL scores, suggesting that the more difficult the text, the more time students needed to understand. This linear trend between the FKGL score of the texts and the students' reading time did not count for the reading rate, as Table 1 shows that the FKGL scores do not correspond to the length of the text.

Table 3 also shows that the comparisons between texts indicate that there are significant differences in students' reading times, particularly at intervals of 2 to 3 levels of text readability (e.g., texts 1 and 4, texts 2 and 5, texts 3 and 6). For example, a comparison between text 1 and text 4 results in an F-score of 2.938 (vertical column) and an α of 0.035* (horizontal column), indicating that text 1 required more time to

understand than text 4. However, not represents a significant result in all comparisons, the positive F value indicates that sample 1 requires more time to comprehend than sample 2.

Time spent reading is typically measured in WPM. Brysbaert's [31] meta-analysis on reading rate studies revealed that normal silent reading rates differ between nonfiction and fiction, with nonfiction having a rate of 238 WPM and fiction at 260 WPM. While our study does not focus on this specific distinction, we have found that when it comes to foreign language reading, the reading rate is not relevant. Our findings indicate that text readability level, rather than length, impacts reading time, with shorter texts requiring more time to read compared to longer texts. For example, a 224-word text (text 3) took an average of 115.50 seconds to read, while a 259-word text (text 8) only took 83 seconds.

The famous reading rate, which calculators are ubiquitous on the internet, may be relevant to measuring the time it takes native speakers to read a given text because they have different reading strategies compared to non-native speakers. Due to their familiarity with the language used in the text, native speakers tend to connect the reading material with their existing knowledge (known as top-down strategy), while non-native speakers tend to read and understand the words first (known as bottom-up strategy) [41]. These different strategies may have influenced students' reading time. Because the non-native English speakers used the bottom-up strategy when reading the texts, they appeared to have a high cognitive load when reading texts with higher FKGL scores, so they needed more time to understand. Since the readability of a text is not just about the number of words, but also the overall internal complexity of the text, they may have had to read the text several times to ensure that they understood the information, especially when they encountered infrequently occurring words and highly complex structures.

When it comes to reading assessments, whether in the classroom or on standardized tests, it is common for students' reading time to be restricted. Although this study does not provide a framework for measuring reading time, relying on reading rate should be unwise. Establishing a framework for measuring reading time in the EFL context in conjunction with text readability is a complex effort. Since the generally accepted reading rate has been shown to be irrelevant for this purpose, there is a need for a framework that addresses this issue; this provides an opportunity for future studies.

In summary, we have strong evidence that text readability has an impact on students' reading time. We found that, regardless of text length, a higher FKGL score is significantly associated with "more reading time". These findings are consistent with previous analysis and suggest that appropriate selection of texts for reading assessment appropriate to students' language proficiency is critical. Time constraints should not be based solely on reading rate, while waiting for future studies to develop how reading time should be managed in an EFL reading context.

4. CONCLUSION

The current study aims to answer the following clear questions: i) RQ1: does the readability of texts have an impact on students' reading comprehension? and ii) RQ2: does the readability of texts have an impact on students' reading time? However, due to several limitations, the conclusions should be approached with caution. We acknowledge that a notable limitation of this study is related to the sampling technique. Due to the number of available participants and the need for statistical analysis, purposive stratified sampling was preferred over random sampling. Furthermore, the number of participants in this study was relatively small. We also did not interview the participants to collect their experiences during the reading test, although the information could be valuable in giving us more information from their side. Another limitation to consider is that the inferential statements that follow each text in this study have not been checked for readability. It could be that students found the statements as difficult as the accompanying text, so their answers were influenced in one way or another. Further studies should address these limitations by including more participants with random sampling, conducting an interview with participants, and checking statements or questions for readability following each reading material. Additionally, since the current study did not examine the relationship between students' reading comprehension and reading time, further research could be expanded for this purpose.

The results of this study led us to conclude that text readability has an impact on both students' reading comprehension and reading time. The statistical analyses revealed differences in the distribution of students' reading comprehension and reading time for 10 texts with different readability levels based on FKGL. This study provided insights into EFL reading instruction by suggesting appropriate reading materials for both instruction and assessment based on students' language skills and grade level. In addition, because the level of readability of texts influences the time it takes for students to read and understand the texts, teachers should not rely solely on the generally accepted reading rate when assessing reading. We need a new framework to measure required reading time based on the level of text readability, and future studies are likely to provide us with one.

ACKNOWLEDGEMENTS

The authors thank to all parties that have supported this study, specifically the Indonesia Endowment Fund for Education Agency (*Lembaga Pengelola Dana Pendidikan*-LPDP), the Ministry of Finance, Republic of Indonesia, for fully supporting the study of an author of this paper including the accomplishment of this study.

REFERENCES

[1] R. Flesch, *The art of readable writing*. New York, 1949.
[2] W. DuBay, *The principles of readability impact information*. 2004.
[3] A. Bailin and A. Grafstein, *Readability: Text and context*. London: Palgrave Macmillan UK, 2016.
[4] E. Matricciani, "Readability indices do not say it all on a text readability," *Analytics*, vol. 2, no. 2, pp. 296–314, Mar. 2023, doi: 10.3390/analytics2020016.
[5] N. Nassiri, A. Lakhouaja, and V. Cavalli-Sforza, "Arabic L2 readability assessment: dimensionality reduction study," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3789–3799, Jun. 2022, doi: 10.1016/j.jksuci.2020.12.021.
[6] J. C. Alderson, *Assessing reading*. Cambridge University Press, 2005.
[7] J. Oakhill, K. Cain, and C. Elbro, *Understanding and teaching reading comprehension: A handbook*. London ; New York: Routledge, 2015.
[8] A. Kiselnikov, D. Vakhitova, and T. Kazymova, "Fundamentals of materials (text readability evaluation)," *E3S Web of Conferences*, vol. 274, p. 12006, Jun. 2021, doi: 10.1051/e3sconf/202127412006.
[9] S. Nahatame, "Text readability and processing effort in second language reading: a computational and eye-tracking investigation," *Language Learning*, vol. 71, no. 4, pp. 1004–1043, Dec. 2021, doi: 10.1111/lang.12455.
[10] C. D. Meade and C. F. Smith, "Readability formulas: cautions and criteria," *Patient Education and Counseling*, vol. 17, no. 2, pp. 153–158, Apr. 1991, doi: 10.1016/0738-3991(91)90017-Y.
[11] M. Sarin and M. Garraffa, "Can readability formulae adapt to the changing demographics of the UK school-aged population? A study on reading materials for school-age bilingual readers," *Ampersand*, vol. 11, Dec. 2023, doi: 10.1016/j.amper.2023.100141.
[12] U. Izgi and B. S. Seker, "Comparing different readability formulas on the examples of science-technology and social science textbooks," *Procedia - Social and Behavioral Sciences*, vol. 46, pp. 178–182, 2012, doi: 10.1016/j.sbspro.2012.05.089.
[13] P. Joseph, N. A. Silva, A. Nanda, and G. Gupta, "Evaluating the readability of online patient education materials for trigeminal neuralgia," *World Neurosurgery*, vol. 144, pp. e934–e938, Dec. 2020, doi: 10.1016/j.wneu.2020.09.123.
[14] S. Y. Sudah *et al.*, "Most patient education materials on shoulder conditions from the American Academy of Orthopaedic Surgeons exceed recommended readability levels," *JSES International*, vol. 7, no. 1, pp. 126–131, Jan. 2023, doi: 10.1016/j.jseint.2022.09.004.
[15] T. Szmuda, C. Özdemir, S. Ali, A. Singh, M. T. Syed, and P. Słoniewski, "Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study," *Public Health*, vol. 185, pp. 21–25, Aug. 2020, doi: 10.1016/j.puhe.2020.05.041.
[16] S. Ladhar, S. L. Koshman, F. Yang, and R. Turgeon, "Evaluation of online written medication educational resources for people living with heart failure," *CJC Open*, vol. 4, no. 10, pp. 858–865, Oct. 2022, doi: 10.1016/j.cjco.2022.07.004.
[17] Y. Abdullah, A. Alokozai, S. O'Connell, and M. K. Mulcahey, "Online patient education materials for common sports injuries are written at too-high of a reading level: a systematic review," *Arthroscopy, Sports Medicine, and Rehabilitation*, vol. 4, no. 3, pp. e861–e875, Jun. 2022, doi: 10.1016/j.asmr.2021.12.017.
[18] C. M. Schreidah, L. M. Fahmy, B. A. Lapolla, E. R. Gordon, B. D. Kwinta, and L. J. Geskin, "Accessibility and readability of online patient education on cutaneous lymphomas," *JAAD International*, vol. 13, pp. 83–90, Dec. 2023, doi: 10.1016/j.jdin.2023.07.010.
[19] H. Uzunçıbuk *et al.*, "Assessment of the readability of online orthodontic educational resources related to 'orthodontic treatment': A cross-sectional study," *Seminars in Orthodontics*, Feb. 2024, doi: 10.1053/j.sodo.2024.01.015.
[20] S. J. Hanish, N. Cherian, J. Baumann, S. D. Gieg, and S. DeFroda, "Reducing the use of complex words and reducing sentence length to <15 words improves readability of patient education materials regarding sports medicine knee injuries," *Arthroscopy, Sports Medicine, and Rehabilitation*, vol. 5, no. 1, pp. e1–e9, Feb. 2023, doi: 10.1016/j.asmr.2022.10.004.
[21] P. Plavén-Sigray, G. J. Matheson, B. C. Schiffler, and W. H. Thompson, "The readability of scientific texts is decreasing over time," *BioRxiv*, 2017, doi: 10.1101/119370.
[22] H. Anggia and A. Habók, "Textual complexity adjustments to the English reading comprehension test for undergraduate EFL students," *Heliyon*, vol. 9, no. 1, p. e12891, Jan. 2023, doi: 10.1016/j.heliyon.2023.e12891.
[23] J. C. Alderson, A. Huhta, and L. Nieminen, "Characteristics of weak and strong readers in a foreign language," *The Modern Language Journal*, vol. 100, no. 4, pp. 853–879, Dec. 2016, doi: 10.1111/modl.12367.
[24] C. Soto, A. P. G. de Blume, M. Jacovina, D. McNamara, N. Benson, and B. Riffo, "Reading comprehension and metacognition: The importance of inferential skills," *Cogent Education*, vol. 6, no. 1, Jan. 2019, doi: 10.1080/2331186X.2019.1565067.
[25] P. Kendeou, P. van den Broek, A. Helder, and J. Karlsson, "A cognitive view of reading comprehension: Implications for reading difficulties," *Learning Disabilities Research & Practice*, vol. 29, no. 1, pp. 10–16, Feb. 2014, doi: 10.1111/ldrp.12025.
[26] A. A. Hakim, E. Setyaningsih, and D. Cahyaningrum, "Examining the readability level of reading texts in English textbook for Indonesian senior high school," *Journal of English Language Studies*, vol. 6, no. 1, Mar. 2021, doi: 10.30870/jels.v6i1.8898.
[27] D. Yetti, "An analysis of readability level of reading material in English textbook for first grade of senior high school," *Journal of Education and Teaching*, vol. 2, no. 1, p. 1, May 2021, doi: 10.24014/jete.v2i1.8211.
[28] I. A. Fata, E. Komariah, and A. R. Alya, "Assessment of readability level of reading materials in Indonesia EFL Textbooks," *Lingua Cultura*, vol. 16, no. 1, pp. 97–104, May 2022, doi: 10.21512/lc.v16i1.8277.
[29] A. Widyantoro, J. Jamilah, and A. Purnawan, "Text difficulty vs text readability: students voices," *EduLite: Journal of English Education, Literature and Culture*, vol. 7, no. 1, p. 125, Feb. 2022, doi: 10.30659/e.7.1.125-136.
[30] A. S. Dirgantari and H. Susantiningdyah, "Students' reading comprehension skill: the roles of text readability and question difficulty," *PEOPLE: International Journal of Social Sciences*, vol. 6, no. 1, pp. 533–544, May 2020, doi: 10.20319/pijss.2020.61.533544.
[31] M. Brysbaert, "How many words do we read per minute? a review and meta-analysis of reading rate," *Journal of Memory and Language*, vol. 109, p. 104047, Dec. 2019, doi: 10.1016/j.jml.2019.104047.

[32] A. Koornneef and A. Kraal, "Does beeline reader's gradient-coloured font improve the readability of digital texts for beginning readers?," *Computers in Human Behavior Reports*, vol. 6, p. 100197, May 2022, doi: 10.1016/j.chbr.2022.100197.
[33] T. C. Urdan, *Statistics in Plain English*, Fourth. New York London: Routledge, Taylor & Francis Group, 2017.
[34] S. W. Scheff, "Nonparametric statistics," in *Fundamental Statistical Principles for the Neurobiologist*, 2016, pp. 157–182.
[35] F. Chen and R. Abdullah, "Scientific mapping of research on English as a foreign language (EFL) reading instruction: A thirty-year perspective," *Heliyon*, vol. 10, no. 3, p. e25217, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25217.
[36] E. Block, "The comprehension strategies of second language readers," *TESOL Quarterly*, vol. 20, no. 3, p. 463, Sep. 1986, doi: 10.2307/3586295.
[37] A. Akhmetova, G. Imambayeva, and B. Csapó, "A study of reading attitude and reading achievement among young learners in middle school," *Heliyon*, vol. 8, no. 7, p. e09946, Jul. 2022, doi: 10.1016/j.heliyon.2022.e09946.
[38] S. T. M. R. Ramachers and R. M. Willems, "Reading short stories in a foreign language: Is it as pleasurable as in the L1? Absorption and appreciation of short stories in Dutch secondary school students learning German as a foreign language," *System*, vol. 120, p. 103211, Feb. 2024, doi: 10.1016/j.system.2023.103211.
[39] C. Ruffini, C. Tarchi, and C. Pecini, "Which executive functions affect text comprehension and writing in paper and digital mode? An investigation in primary school children," *Computers and Education*, vol. 207, p. 104936, Dec. 2023, doi: 10.1016/j.compedu.2023.104936.
[40] M. Habib, "Assessment of reading comprehension," *Revista Romaneasca pentru Educatie Multidimensionala*, vol. VIII, no. I, pp. 125–147, Jun. 2016, doi: 10.18662/rrem/2016.0801.08.
[41] J. N. Davis and L. Bistodeau, "How do L1 and L2 reading differ? Evidence from think aloud protocols," *The Modern Language Journal*, vol. 77, no. 4, pp. 459–472, Dec. 1993, doi: 10.1111/j.1540-4781.1993.tb01993.x.

BIOGRAPHIES OF AUTHORS

Zainurrahman received his bachelor's degree in English literature from Universitas Khairun Ternate and his master's degree in English education from Universitas Pendidikan Indonesia. At the time of writing, he is pursuing a Ph.D. in English education from Universitas Pendidikan Indonesia. He is also currently an EFL lecturer at a private institute (Institut Sains dan Ilmu Kependidikan Kie Raha) in Ternate, Indonesia. He has published several papers related to EFL and is actively conducting relevant studies in the field of TEFL. He can be contacted at email: either zainurrahman@upi.edu or zainurrahmankalero@gmail.com.

Fazri Nur Yusuf is a lecturer at Universitas Pendidikan Indonesia with expertise in EFL pedagogy, with particular emphasis on the areas of reading, writing and EFL lesson planning. His academic career began at IKIP Bandung and led to a master's degree from Universitas Pendidikan Indonesia and a doctorate degree from Universitas Negeri Malang. He can be contacted at email: fazrinuryusuf@upi.edu.

Didi Sukyadi is a professor at Universitas Pendidikan Indonesia with expertise in applied linguistics and semiotics. His academic career began at IKIP Bandung, where he completed his undergraduate studies. He expanded his knowledge through a master's degree at the University of London, gaining a global perspective on linguistics. His commitment to advancing knowledge led him to obtain his doctorate from Universitas Indonesia. He can be contacted at email: dsukyadi@upi.edu.