A Study on the Impact of Readability on Comprehensibility*

Halil İbrahim Öksüz [i]
Gazi University

Hasan Kağan Keskin [ii]
Duzce University

Abstract

The aim of the study is to examine the effect of readability on comprehensibility. The study is quasi-experimental as it was not possible to assign groups randomly and the groups were partially controllable. Therefore, the "pretest-posttest unequaled control group quasi-experimental design" was used in the study. The study group consists of 121 fourth grade students divided into two experimental and two control groups. The texts taken from the Turkish textbook approved by the Education Board were used to collect the data of the research. In order to measure the comprehensibility of the texts, the comprehension scale of the Mistake Analysis Inventory was used. The analysis of the data of the study was carried out using the SPSS 24 package. In order to ensure reliability in the study, some of the comprehensibility data were scored by two raters. To do so, comprehension data from 50 randomly selected participants were used. As the comprehension data of 50 participants did not show normal distribution, the Spearman Brown Rank Differences Correlation Test was conducted in order to reveal the relationship between the two raters. According to the test results, a strong and positive significant relationship was found between the two raters ($r (48) = .88$, $p = .00$, $p < .05$). The findings showed that the original (more difficult) versions of the texts were more comprehensible than the versions simplified by the researcher. In other words, long sentences are more comprehensible than short sentences. In this direction, the present study revealed that the readable text may not always be comprehensible.

Keywords: Readability, Readability Formulas, Comprehensibility

----

[i] **Halil İbrahim Öksüz,** PhD Student, Gazi Faculty of Education, Gazi University, ORCID: 0000-0003-4338-297X

Correspondence: oksuzhalilibrahim@hotmail.com

[ii] **Hasan Kağan Keskin**, Prof. Dr., Faculty of Education, Duzce University, ORCID: 0000-0002-5495-1747

INTRODUCTION

Why do people prefer reading materials that are below their level? According to Hochhouser (1997), the vast majority of people tend to read written materials that are several levels below their education level (as cited in Burke & Greenberg, 2010). This tendency that people display is an indicator that a problem exists. It stems from two sources. The first is people who choose books below their level while the second is the estimation tools that determine the level of written materials. These tools are the five-finger technique, leveled books, rubrics, checklists, and readability formulas (Ateş, Çetinkaya, & Yıldırım, 2012). Among these, readability formulas may be said to be the most frequently used. Indeed, 3.166 studies were displayed as a result of searching the concept of "readability" in the ERIC database on 16 October 2018 (ERIC, 2018). 21.57% of the existing studies have been done in the last 20 years.

Readability is discovering what is easy and difficult to make the language more comprehensible (Bormuth, 1967). Ateşman (1997) defines it as a text being labelled as easy or difficult to understand by readers. On the other hand, readability formulas are tools for mathematically calculating the difficulty of reading materials (Clarck, 1981; McLaughlin, 1969; Ülper, 2010; Walpole & McKenna, 2007). Used for the first time in institutional communication, readability is a concept frequently used in our day by educators and linguists (Ateşal, 2013).

There is much criticism in the literature for of the frequently used readability concept and readability formulas. These criticisms may be grouped under three headings. The first one is that readability formulas produce different results (Asem, 2012; Ateşman, 1997; Bargate, 2012; Chall, 1988; Çepni, Gökdere, & Küçük, 2002; Gallagher, Fazio, & Gunning, 2012; Geçit, 2010; Köse, 2009; Okur, Arı, Ersoyal, & Okur, 2013; Stokes, 1978; Tekbıyık, 2006; Topkaya, Kalın, & Yılar, 2015; Turan & Geçit, 2010; Ulusoy, 2006). Indeed, Chall (1988) and Gallagher et al. (2012) state that even the same text measured by two independent scorers using the same formula may yield different results.

The second criticism is that authors and publishers shorten the words and sentences in reading materials so as to achieve lower readability scores (Armbruster, Osborn, & Davison, 1985; Bruce, Rubin, & Starr, 1981; Chall, 1988; Marshall, 1979). This results in low quality texts (Chall, 1988).

The final criticism in the literature is that the definitions of readability and the variables used in the formulas do not overlap. The readability definitions of many researchers who have a readability formula (Ateşman, 1997; Bormuth, 1967; Chall, 1988; Dreyer, 1984; Göğüş, 1978; McLaughlin, 1969; Tekbıyık, 2006) imply that texts need to be readable to be comprehensible. However, readability formulas, which seem to have a big influence on comprehensibility, generally use variables that depend on the syntactic structure of a given text such as average word and sentence length (Asem, 2012; Ateşman, 1997; Bezirci & Yılmaz, 2010; Davison, 1988; Fry, 2002; Gallagher et al., 2012; Kong, 2009; Okur & Arı, 2013; Tekbıyık, 2006; Zorbaz, 2007). On the other hand, readability formulas do not consider variables other than syntactic ones, which may actually affect comprehensibility greatly, such as text structure, length, logical pattern and pictures, (Fry, 2002) the reader's interests, motivations and language skills, (Ateşal, 2013; Courtis, 1987; Oakland & Lane, 2004; Stevens, Stevens, & Stevens, 1992; Wissing, Blignaut, & Van den Berg, 2016; Yazıcı & Yeşilbursa, 2007) body of vocabulary, (Yazıcı & Yeşilbursa, 2007) existing information (Marshall, 1979; Oakland & Lane, 2004; Özdemir, 2016; Pishghadam & Abbasnejad, 2016; Stevens et al., 1992; Zakaluk & Samuels, 1996) and context (Armbruster et al., 1985; Harrison & Bakker, 1998; Marshall, 1979). Akyol (2006) states that the text cannot express itself and that it is in constant interaction with the prior knowledge of the reader Based on Akyol's (2006b) view, it may be stated that the reader should be involved in the level identification process (Keskin & Akıllı, 2013); however, readability formulas do not do this and are criticized for it (Compton, Appleton, & Hosp, 2004; Pishghadam & Abbasnejad, 2016; Stevens et al., 1992; Temur, 2003; Wissing et al., 2016).

There are many studies in the literature that find readability formulas insufficient and inconsistent in measuring the comprehensibility of reading materials (Asem, 2012; Ateşman, 1997;

Bargate, 2012; Chall, 1988; Çepni et al., 2002; Gallagher et al., 2012; Geçit, 2010; Köse, 2009; Okur & Arı, 2013; Okur et al., 2013; Stokes, 1978; Tekbıyık, 2006; Topkaya et al., 2015; Turan & Geçit, 2010; Ulusoy, 2009). In addition, there are a number of studies indicating that reading materials with high readability scores will also have high comprehensibility and that a positive relationship therefore exists between these two concepts (Ateşman, 1997; Bezirci & Yılmaz, 2010; Bormuth, 1967; Chall, 1988; Dreyer, 1984; Fry, 2002; Göğüş, 1978; Güyer, Temur, & Solmaz, 2009; Kalın & Aydemir, 2017; Köse, 2009; McLaughlin, 1969; Mert, 2018; Mirzaoğlu & Akın, 2015; Özbek & Ergül, 2018; Richards, Platt, & Platt, 1992; Tekbıyık, 2006; Yalın, 1996; Zamanian & Heydari, 2012). This difference of opinion in the literature has paved the way for this study. On the other hand, readability definitions (Ateşman, 1997; Bormuth, 1967; Chall, 1988; Dreyer, 1984; Göğüş, 1978; McLaughlin, 1969; Tekbıyık, 2006) imply that high readability is required in order for texts to be understood. However, researcher such as (Geçit, 2010; Köse, 2009; Topkaya et al., 2015; Turan & Geçit, 2010) have calculated the readability of textbook passages via different formulas and revealed that they were appropriate for the target level. When they subsequently applied the same text to the target grade level by using the Cloze Test method, they concluded that the texts were comprehensible with the support of the teacher. It can therefore be claimed that there is a discrepancy between the results of the studies in the literature and the definitions of readability. The method used to measure comprehensibility in these studies was the Cloze Test method. The present study, however, attempted to measure the effect of readability on comprehensibility by using the comprehension scale of the Mistake Analysis Inventory adapted to Turkish by (Akyol, 2006). Another unique feature of the study is that it collects its data by using the original text alongside a simplified version with shorter sentences developed by the researcher. The aim of the study is to examine the effect of readability on comprehensibility. For this purpose, answers to the following questions were sought:

1. Is there a significant difference in comprehensibility scores between the original version of the first informative text and the second one simplified by the researcher?

2. Is there a significant difference in comprehensibility scores between the original version of the second informative text and the first one simplified by the researcher?

3. Is there a significant difference in comprehensibility scores between the first and second informative texts simplified by the researcher?

4. Is there a significant difference in comprehensibility scores between the original versions of the first and second informative texts?

METHOD

Research Model

This is an experimental study which aims to examine the effect of readability on comprehensibility. Experimental studies aim to see how systematic changes in the independent variable affect the dependent variable (Karasar, 2012). In addition, Büyüköztürk (2007) states that the researcher must manipulate the independent variable in order to conduct an experimental study. The manipulated variable whose effect is investigated in this study is *readability,* and the affected variable is *comprehensibility.*

The study is quasi-experimental as it was not possible to assign groups randomly (Büyüköztürk, Çakmak, Akgün, Karadeniz, & Demirel, 2017) and the groups were partially controllable (Singh, 2007, p. 67). Therefore, the "pretest-posttest unequaled control group quasi-experimental design" (Karasar, 2012) was used in the study.

Study Group

The study group consists of 121 fourth grade students divided into two experimental and two control groups. Two experimental and two control sections were created in order to see whether the texts selected for the study were equivalent to each other and to avoid the "same text effect". Two sampling methods were used when determining the study group. The first was convenience sampling when determining the primary school where the study would be conducted. The second method was purposeful sampling used in order to select the 4 sections that did not differ significantly in terms of their comprehension scores by applying a pre-test in the primary school.

Data Collection Tool

The pre-test used "Winter Preparations" text taken from the Turkish textbook approved by the Education Board in the 2013-2014 academic year. The readability scores and levels of the texts used in the study are given in Table 1.

Table 1: Readability Scores and Levels of the Texts Used in the Study

Title of Text	Readability Score	Level
Bicycles (Original)	42.27	Difficult
Bicycles (Simplified)	61.02	Intermediate
Mysterious Organisms (Original)	48.72	Difficult
Mysterious Organisms (Simplified)	61.20	Intermediate

The texts were taken from the fourth grade textbooks approved by the Board of Education to suit the level of the students. In order to measure the comprehensibility of the texts, the comprehension scale of the Mistake Analysis Inventory adapted by Akyol (2006) was used. Consequently, 3 basic and 3 in-depth comprehension questions were written for each text. For basic comprehension questions, 0 points were allocated for unanswered or incorrect questions, 1 point for partially answered questions, and 2 points for full answers. For in-depth comprehension questions, 0 points were allocated for unanswered or incorrect questions, 1 point for partially answered questions, 2 points for expected but incomplete answers, and 3 points for full and effective answers (Akyol, 2006).

Data Collection and Procedures

In the preparation phase of the study, the researcher scanned fourth grade Turkish textbooks. A text pool was thus created with selected texts from these books. The selected texts were then scanned and computed. Fourteen of the texts in the pool were removed as their readability score was above 70. On the other hand, 27 other texts were also removed from the pool as their score increased by less than 10 points during the simplification process. Among the remaining texts, the two closest were selected based on their readability scores.

The texts have been simplified in terms of sentence length by paying attention to context. The simplification process was based on the principle of shortening sequential sentences by using full stops. Following this, the readability of the texts was calculated via Flesch's Ease of Reading Formula adapted into Turkish by Ateşman (1997). The two variables in the formula are average word and sentence lengths. These averages are calculated by counting all syllables, words and sentences in the text. Then, the number of syllables is divided by the number of words to find average word length, and the average number of words is divided by the average number of sentences to find average sentence length. Online applications were used to count syllables, words and sentences in texts. The accuracy of different syllable, word and sentence calculation pages was tested as follows:

The researcher counted the syllables, words and sentences of a selected text, then the computed text was uploaded to different syllable, word and sentence calculation pages and the data were recorded. The webpages overlapping with the data obtained by the researcher were used to count the syllables, words and sentences in the texts. Internet sites were used to calculate the number of syllables (HesaplamaOnline, 2019), words (Hesaplamalar, 2019) and sentences (eHesaplama, 2019) in the texts.

The opinions of two experts from Düzce University's Faculty of Education, Elementary Education Department and one from the Department of Turkish Education were obtained about the texts. Other than these experts, the views of a linguist were also obtained. The simplified texts were revised in line with expert views. The resulting texts were finalized with the help of the Turkish Education expert again. As a last step, expert opinion was gathered once again on the final simplified texts to rule out any contextual errors.

After deciding on the elementary school where the study was to be conducted, a pre-test was applied to all fourth grades in the school. After the two-day pre-test, four fourth-grade classes that did not vary significantly in their comprehension scores were randomly assigned to experimental and control groups.

The implementation phase of the study was completed in four school days. It was conducted by the researcher with the thought that there might be teacher competition in the school. Before the implementation, the researcher introduced himself to the students and informed them that the scores they receive would not be entered into the e-School system. During the implementation, care was taken for each student to read the text once and the readers were asked to turn the paper over upon completion.

There are also foreign students whose mother tongue is not Turkish in the classrooms. In order not to make these students feel their differences, the students were included in the implementation process. However, the data of these students were not included in the analysis of the research. Therefore, research data were collected from students whose mother tongue was Turkish, who did not receive any special education, and who showed normal development. A total of seven students, two inclusion students in experimental group 1, one foreign student in experimental group 2, two foreign students in control group 1, and one foreign student and one inclusion student in control group 2 were removed from the study.

On the other hand, the texts used in the study were applied crosswise to the experimental and control groups. The reason for this situation is to prevent the participants from gaining familiarity with the texts by encountering two different versions of the same text.

Data Analysis

For normality analysis, the study made use of the Lilliefors-corrected Kolmogorov-Smirnov Test (K-S). According to the results obtained from the normality test, Paired Samples T Test was used to analyze the data with normal distribution, and Wilcoxon Signed Ranks Test was used to analyze the data that did not show normal distribution. As the scores from the comprehension test given to determine group equivalence were normally distributed, One Way ANOVA Test was used to compare the 4 groups. In these tests, significance value was set as 0.05. These tests were conducted using the SPSS 24 package (IBMCorp., 2016)

In order to ensure reliability in the study, some of the comprehensibility data were scored by two raters. To do so, comprehension data from 50 randomly selected participants were used. This procedure was not applied to all the data for time and funding reasons. After this process, the correlation method was used to ensure inter-rater reliability. As the comprehension data of 50 participants did not show normal distribution, the Spearman Brown Rank Differences Correlation Test was conducted in order to reveal the relationship between the two raters. According to the test results,

a strong and positive significant relationship was found between the two raters (r (48) = .88, p = .00, p <.05)

FINDINGS AND INTERPRETATIONS

The K-S Test was used to measure the normality of the equivalence test data. The results showed that the comprehension scores of experimental group 1 D (30) = 0.200, p> .05, experimental group 2 D (32) = 0.200, p> .05, control group 1 D (29) = 0.200, p> .05 and control group 2 D (32) = 0.200, p> .05 showed normal distribution. Based on the K-S Test results, One Way ANOVA was used to compare the comprehension scores of the groups.

Table 2: One Way ANOVA Test Results According to the Comprehension Scores of the Groups

Groups	n	x̄	S
Experimental Group 1	30	8.27	3.151
Experimental Group 2	32	8.00	2.383
Control Group 1	29	8.00	3.128
Control Group 2	32	8.56	2.514
Total	123	8.21	2.777

Source of variance	Sum of Squares	Sd	Mean of Squares	F	p	
Gruplar Arası	6.762	3	2.254	.287	.835	
Gruplar İçi	933.742	119	7.847			
Total	940.504	122				

As presented in Table 2, the ANOVA test (F (3, 119) =. 287; p> .05) which was conducted to see if there was a statistically significant difference between the groups revealed no significant difference in terms of comprehension scores. It may therefore be stated that the experimental and control groups were equivalent to each other at the onset of the study.

Findings and Comments on the First Sub-Problem

The comprehensibility of the texts was measured by first applying the original version of the *Bicycles* text to experimental group 1, followed by the simplified version of the *Mysterious Creatures* text. While the *Bicycles* text D (30) =. 010, p <.05 did not show normal distribution, the *Mysterious Creatures* text D (30) =. 200, p> .05 did. For this reason, Wilcoxon Signed Ranks Test was used to compare the two related measurements.

Table 3: Comprehensibility of Informative Pretest-Posttest Texts Applied in Experimental Group 1

Bicycles (Original)/Mysterious Creatures (Simplified)		n	Mean of Ranks	Sum of Ranks	z	p
Comprehensibility	Negative Ranks	20	14.78	295.50	-3.60	.000*
	Positive Ranks	5	5.90	29.50		
	Equal	5	-	-		

*p<.001

As shown in Table 3, experimental group 1 read the original version of the *Bicycles* text, which ranks as difficult with a readability score of 42.27. Following this, they read the *Mysterious Creatures* text that was simplified by the researcher to a readability score of 61.20. The

comprehensibility test revealed a significant difference between the comprehensibility of the high and moderate difficulty texts (z = -3.60, p <.05). The mean score of students whose comprehension of decreased between the readings was 14.78, while the mean score of students whose comprehension increased was 5.90. Based on these results, it may be inferred that the text with a low readability score is more comprehensible than the text with a higher score.

Findings and Comments on the Second Sub-Problem

In experimental group 2, the original version of the *Mysterious Creatures* text and then the simplified version of the *Bicycles* text were implemented to measure comprehensibility. Neither *Mysterious Creatures* D (32) =. 015, p <.05 nor *Bicycles* D (32) =. 036, p <.05 showed normal distribution. For this reason, the Wilcoxon Signed Ranks Test was used to compare the two related measurements.

Table 4: Comprehensibility of Informative Pretest-Posttest Texts Applied in Experimental Group 2

Mysterious Creatures (Original)/Bicycles (Simplified)		n	Mean of Ranks	Sum of Ranks	z	p
Comprehensibility	Negative Ranks	23	16.57	381.00	-4.08	.000*
	Positive Ranks	5	5.00	25.00		
	Equal	4	-	-		

*p<.001

According to Table 4, experimental group 2 read the original version of the *Mysterious Creatures* text, which ranks as difficult with a readability score of 48.72. Following this, they read the *Bicycles* text that was simplified by the researcher to a moderate level with a readability score of 61.02. According to the test results shown in the table, a significant difference occurred between the comprehensibility of the text with moderate difficulty and that with high difficulty (z = -4.08, p <.05). While the mean rank of the students whose comprehension decreased from the first text to the second was 16.57, that of students whose comprehension increased was 5.00. Based on these results, it may be argued that the text with a low readability score is more comprehensible than the text with a higher score.

Findings and Comments on the Third Sub-Problem

The comprehensibility and consistency of the texts were tested by first applying the *Bicycles* text, followed by the *Mysterious Creatures* text in control group 1. Both the *Bicycles* text D (29) =. 081, p> .05 and the *Mysterious Creatures* text D (29) =. 200, p> .05 showed normal distribution. Therefore, the Associated Samples T Test was applied to compare the two related measurements.

Table 5: Comprehensibility of Informative Pretest-Posttest Texts Applied in Control Group 1

	n	Mean	ss	sd	t	p
Bicycles (Simplified)	29	7.38	2.25	28	-1.28	.211
Mysterious Creatures (Simplified)	29	8.10	2.67			

Table 5 shows that the comprehensibility and consistency of the texts were tested in control group 1 by applying the simplified versions of *Bicycles* whose readability score was 61.02 and *Mysterious Creatures* whose readability score was 61.20, both showing a moderate difficulty level. According to the test results presented in the table, there is no significant difference between the comprehensibility of the moderately difficult texts *Bicycles* (M = 7.38, SS = 2.25) and *Mysterious Creatures* (M = 8.10, SS = 2.67) (t (28) = - 1.28, p> .05). These results showed that the texts

simplified by the researcher were consistent. It may therefore be claimed that the changes in the experimental groups were not random but caused by manipulating the readability of the texts.

Findings and Comments on the Fourth Sub-Problem

The comprehensibility of the texts was measured by applying the the original versions of the *Bicycles* and *Mysterious Creatures* texts in control group 2. While the *Bicycles* text D (30) =. 057, p> .05 showed normal distribution, the *Mysterious Creatures* text D (30) =. 026, p <.05 did not. For this reason, the Wilcoxon Signed Ranks Test was utilized to compare the two related measurements.

Table 6: Comprehensibility of Informative Pretest-Posttest Texts Applied in Control Group 2

Bicycles (Original)/Mysterious Creatures (Original)		n	Mean rank	Sum of ranks	z	p
Comprehensibility	Negative Rank	14	14.93	209.00	-.086	.388
	Positive Rank	12	11.83	142.00		
	Equal	4	-	-		

According to Table 6, the comprehensibility of the texts was tested in control group 2 by applying the original versions of *Bicycles* whose readability score was 42.27 and *Mysterious Creatures* whose readability score was 48.72, both showing a high difficulty level. Based on the results given in the table, there was no significant difference between the comprehensibility of the *Bicycles* and *Mysterious Creatures* texts, both of which had a difficulty level (z = -. 086, p> .05). While the mean rank of the students whose comprehension decreased from the *Bicycles* text to the *Mysterious Creatures* text was 14.93, that of students whose comprehension increased was 11.83. This revealed that the original texts were consistent. Here, too, it may be argued that the changes in the experimental groups were not random but caused by manipulating the readability of the texts.

RESULTS AND DISCUSSION

The main purpose of this study is to determine the effect of readability of texts on comprehensibility. The findings show that the readability levels of the texts are not as effective as thought on the comprehensibility levels of the texts. Sentences in the texts used in the study were shortened while paying attention to context. The readability of the selected and shortened texts was calculated with the readability formula adapted to Turkish by Ateşman (1997). In this way, an original text and an easier version as depicted by its readability score were obtained. The results showed that the original (more difficult) versions of the texts were more comprehensible than the versions simplified by the researcher. In other words, long sentences are more comprehensible than short sentences. These results contradict those of previous studies in the literature which claim that short sentences are more comprehensible than long sentences (Benjamin, 2012; Bezirci & Yılmaz, 2010; Clifton, Staub, & Rayner, 2007; Coke, 1973; Çiftçi, Çeçen & Melanlıoğlu, 2007; Durukan, 2014; Elli, 2011; Güven, 2010; Güyer et al., 2009; İskender, 2013; Jongsma, 1971, p.28; Karatay, Bolat and Güngör, 2013; Küçükahmet, 2011; McLaughlin, 1969; Mirzaoğlu & Akın, 2015; Okur & Arı, 2013; Özbek & Ergül, 2018; Tekbıyık, 2006; Temur, 2002, 2003; Tosunoğlu & Özlük, 2011; Ulusoy, 2009; Yazıcı & Yeşilbursa, 2007; Zorbaz, 2007). According to Budak (2005), it is hard to comment on the difficulty level of texts whose readability is calculated with the variables of mean sentence and word length. In addition, Stevens et al. (1992) report that readability formulas based on sentence length assume that a longer sentence will be more difficult to comprehension than a short one, but that this may not always be the case.

There are opinions in the literature that reading materials with high readability will also have high comprehensibility (Ateşman, 1997; Bezirci & Yılmaz, 2010; Bormuth, 1967; Chall, 1988; Dreyer, 1984; Fry, 2002; Güyer et al., 2009; Kalın & Aydemir, 2017; Köse, 2009; McLaughlin, 1969;

Mert, 2018; Mirzaoğlu & Akın, 2015; Özbek & Ergül, 2018; Richards et al., 1992; Tekbıyık, 2006; Yalın, 1996; Zamanian & Heydari, 2012). The results of the present study do not parallel this common view in the literature. This may be attributed to two different reasons. The first is the difference in the focus points of readability and comprehensibility. A big majority of the readability formulas in the literature focus on the word or sentence means when calculating text readability (Asem, 2012; Ateşman, 1997; Bezirci & Yılmaz, 2010; Davison, 1988; Fry, 2002; Gallagher et al., 2012; Kong, 2009; Okur and Arı, 2013; Stevens et al., 1992; Tekbıyık, 2006; Zorbaz, 2007). comprehensibility, on the other hand, focuses on semantics (Baş & Yıldız, 2015; Fry, 2002; Kasule, 2011; Puurtinen, 1998), reader and text characteristics (Dale & Chall, 1949; Jones, 1997; Kasule, 2011), and cohesion (Coşkun, 2005).

The second reason involves certain factors that highly affect comprehensibility but are ignored by readability formula. These factors are the logical pattern of the text and the pictures used (Fry, 2002), the reader's interest, motivation and language skills (Ateşal, 2013; Courtis, 1987; Oakland & Lane, 2004; Stevens et al., 1992; Wissing et al., 2016; Yazıcı & Yeşilbursa, 2007), vocabulary (Yazıcı & Yeşilbursa, 2007), background information (Marshall, 1979; Oakland & Lane, 2004; Özdemir, 2016; Pishghadam & Abbasnejad, 2016; Stevens et al., 1992; Zakaluk & Samuels, 1996), and context (Armbruster, Osborn, & Davison, 1985; Harrison & Bakker, 1998; Marshall, 1979).

The results of the present study show that the readability measured by formulas is not as effective as it is believed to be in determining whether a given text or reading material is comprehensible. Asem (2012) also concluded that readability formulas yield inconsistent results about comprehensibility. In addition, many researchers criticize readability formulas for yielding inconclusive results (Asem, 2012; Ateşman, 1997; Bargate, 2012; Chall, 1988; Çepni et al., 2002; Gallagher et al., 2012; Geçit, 2010; Köse, 2009; Okur et al., 2013; Stokes, 1978; Tekbıyık, 2006; Topkaya et al., 2015; Turan & Geçit, 2010; Ulusoy, 2009).

Contrary to the results of previous studies and definitions of readability in the literature that reading materials shown to be readable by formulas are also comprehensible (Ateşman, 1997; Bezirci & Yılmaz, 2010; Bormuth, 1967; Chall, 1988; Dreyer, 1984; Fry, 2002; Kalın & Aydemir, 2017; Köse, 2009; McLaughlin, 1969; Mert, 2018; Mirzaoğlu & Akın, 2015; Özbek & Ergül, 2018; Richards, Platt & Platt, 1992; Tekbıyık, 2006; Yalın, 1996; Zamanian & Heydari, 2012), the present study has revealed that this may not always be the case. As documented by Puurtinen (1998), readability depends on the abilities of the reader and the characteristics of the reading materials.

In line with the results of the study and the literature discussed, several recommendations were made for implementation and research. The first of these recommendations is to consider the perception level of readers, instead of using methods that ignore reader characteristics, when choosing reading materials. At the same time, publishers and authors are recommended to focus on the appropriateness of reading materials to the level of the reader rather than the score obtained from readability formulas. Finally, readability formulas focus on variables such as the ratio of mean word and sentence length in the syntactic structure of reading materials to difficult words. As this leads to inconsistency in matching the reader with the text, researchers are recommended to focus on the reader in the formulas they develop.

LIMITATIONS

The most serious limitation of the study is that it only makes inferences about the comprehensibility of texts whose readability were calculated by means of formulas. Therefore, the conclusion in the study that "texts that are readable may not always be comprehensible" does not include readability estimates made by using a cloze test, the five finger technique, or various rubrics.

The study was also limited to fourth grade texts and students. Although it could have covered all the levels included in the formula adapted to Turkish by Ateşman (1997), the fourth grade was chosen due to potential problems with time and costs.

The final limitation of the study is that the groups were formed earlier in the school system due to the class structure at schools. As a result of biased grouping, experimental and control groups may have matured differently during the study, consequently posing a disadvantage to internal validity (Karasar, 2012). Therefore, caution needs to be exercised when generalizing the data.

REFERENCES

Akyol, H. (2006). *Türkçe öğretim yöntemleri*. Ankara: Kök Yayıncılık.

Armbruster, B. B., Osborn, J. H., & Davison, A. (1985). Readability Formulas May Be Dangerous to Your Textbooks. *Educational Leadership, 42*(7), 18-20.

Asem, B. (2012). Readability assessment of printed materials: going beyond readability formulas. *International Journal of Environment, Ecology, Family and Urban Studies, 2*(4), 45-56.

Ateş, S., Çetinkaya, Ç., & Yıldırım, K. (2012). Okunabilir bir kitap nasıl seçilir? *İlköğretim Online, 11*(2), 319-331.

Ateşal, Z. (2013). *Didaktik metinlerin anlaşılabilirlik düzeyi (9. sınıf örneği)*. (Yayımlanmamış yüksek lisans tezi). Sakarya Üniversitesi Eğitim Bilimleri Enstitüsü. Sakarya, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 06.10.2018 tarihinde erişildi.

Ateşman, E. (1997). Türkçede okunabilirliğin ölçülmesi. *Ankara Üniversitesi TÖMER Dil Dergisi, 58*, 171-174.

Bargate, K. (2012). The readability of managerial accounting and financial management textbooks. *Meditari Accountancy Research, 20*(1), 4-20. doi:org/10.1108/10222521211234192

Baş, B., & Yıldız, F. İ. (2015). 2. sınıf Türkçe ders kitabındaki metinlerin okunabilirlik açısından incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 11*(1), 52-61.

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review, 24*(1), 63-88. doi:10.1007/s10648-011-9181-8

Bezirci, B., & Yılmaz, A. E. (2010). Metinlerin okunabilirliğinin ölçülmesi üzerine bir yazılım kütüphanesi ve Türkçe için yeni bir okunabilirlik ölçütü. *DEÜ Mühendislik Fakültesi Fen Bilimleri Dergisi, 12*(3), 49-62.

Bormuth, J. R. (1967). *New data on readability*. Minneapolis, Minnesota: University of Minnesota

Bruce, B. C., Rubin, A., & Starr, K. S. (1981). Why readability formulas fail. *IEEE Transactions on Professional Communication, 24*(1), 1-17. doi:10.1109/TPC.1981.6447826

Budak, Y. (2005). Metinlerin okunabilirlik düzeyinin saptanmasına yönelik eleştirel bir bakış *Eurasian Journal of Educational Research, 21*, 76-87.

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.

Büyüköztürk, Ş. (2007). *Deneysel desenler: Öntest-sontest kontrol gruplu desen ve SPSS uygulamalı veri analizi* (Geliştirilmiş 2 ed.). Ankara: Pegem Yayınları.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri* (23 ed.). Ankara: Pegem Akademi.

Chall, J. S. (1988). The beginning years. In B. L. Zakaluk & S. J. Samuels (Eds.), *Readability: Its past, present ve future*. Newark, Delaware: International Reading Association.

Clarck, C. H. (1981). Assessing comprehensibility: The PHAN system. *The Reading Teacher, 34*(6), 670-675.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. v. Gompel, M. H. Fisher, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. Amsterdam: Elsevier.

Coke, E. U. (1973). Readability and its effects on reading rate, subjective judgments of comprehensibility. and comprehension. *American Educational Research Association, New Orleans, Louisiana, February 25-March*.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research and Practice, 19*(3), 176-184.

Coşkun, E. (2005). *İlköğretim öğrencilerinin öyküleyici anlatımlarında bağdaşıklık, tutarlılık ve metin elementleri*. (Yayımlanmamış doktora tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 17.10.2018 tarihinde erişildi.

Courtis, J. K. (1987). Fry, smog, lix and rix: Insinuations about corporate business communications. *The Journal of Business Communication, 24*(2), 19-27.

Çepni, S., Gökdere, M., & Küçük, M. (2002). Adaptation of the readability formulas into the Turkish science textbooks. *Energy Education Science and Technology, 10*(1), 49-58.

Çiftçi, Ö., Çeçen, M. A., & Melanlıoğlu, D. (2007). Altıncı sınıf Türkçe ders kitaplarındaki metinlerin okunabilirlik açısından değerlendirilmesi. *Elektronik Sosyal Bilimler Dergisi, 6*(22), 206-219.

Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English, 26*(1), 19-28.

Davison, A. (1988). Assigning grade levels without formulas: Some case studies. In B. L. Zakaluk & S. J. Samuels (Eds.), *Readability: Its past, present, and future*. Newark, Delaware International Reading Association.

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

Durukan, E. (2014). Metinlerin okunabilirlik düzeyleri ile öğrencilerin okuma becerileri arsındaki ilişki. *Ana Dili Eğitimi Dergisi, 2*(3), 68-76.

eHesaplama. (2019). Cümle sayısı hesaplama. http://www.ehesaplama.com/d/dil-bilgisi-hesaplamalari/cumle-sayisi-hesaplama adresinden 11.02.2019 tarihinde erişildi.

Elli, T. (2011). *İlköğretim 8. sınıf Türkçe ders kitaplarındaki metinlerin anlaşılabilirlik düzeyleri*. (Yayımlanmamış yüksek lisans tezi). Niğde Üniversitesi Sosyal Bilimler Enstitüsü. Niğde, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 08.10.2018 tarihinde erişildi.

ERIC. (2018). https://eric.ed.gov/? adresinden 16.10.2018 tarihinde erişildi.

Fry, E. B. (2002). Readability versus leveling. *The Reading Teacher, 56*(3), 286-292.

Gallagher, T. L., Fazio, X., & Gunning, T. G. (2012). Varying readability of science-based text in elementary readers: Challenges for teachers. *Reading Improvement, 49*(3), 93-112.

Geçit, Y. (2010). Lise coğrafya 9 ve lise coğrafya 11 ders kitaplarının bazı okunabilirlik formülleri ile değerlendirilmesi. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(4), 2177-2220.

Göğüş, B. (1978). *Orta dereceli okullarımızda Türkçe ve yazın öğretimi*. Ankara: Gül Yayınevi.

Güven, S. (2010). *İlköğretim 6. sınıf fen ve teknoloji ders kitabının okunabilirliği ve hedef yaş düzeyine uygunluğu*. (Yayımlanmamış yüksek lisans tezi). Selçuk Üniversitesi Fen Bilimleri Enstitüsü. Konya, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 11.10.2018 tarihinde erişildi.

Güyer, T., Temur, T., & Solmaz, E. (2009). Bilgisayar destekli metin okunabilirliği analizi. *Türk Eğitim Bilimleri Dergisi, 7*(4), 751-766.

Harrison, S., & Bakker, P. (1998). Two new readability predictors for the professional writer: Pilot trials. *Journal of Research in Reading, 21*(2), 121-138.

Hesaplamalar. (2019). Kelime, karakter sayısı hesaplama. http://www.m5bilisim.com/tr/hesaplamalar/kelime-sayisi-hesaplama/ adresinden 11.02.2019 tarihinde erişildi.

HesaplamaOnline. (2019). Hece hesaplama. https://www.hesapla.online/hece-hesaplama.html adresinden 11.02.2019 tarhinde erişildi.

IBMCorp. (2016). IBM SPSS Statistics for Windows (Version 24.0). IBM Corp: Armonk, NY.

İskender, E. (2013). *Türkçe ders kitaplarındaki metinlerin kelime ve cümle yapılarıyla okunabilirlik düzeyleri arasındaki ilişki*. (Yayımlanmamış yüksek lisans tezi). Karadeniz Teknik Üniversitesi Eğitim Bilimleri Enstitüsü. Trabzon, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 11.10.2018 tarihinde erişildi.

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

Jongsma, E. A. (1971). *The cloze procedure as a teaching technique*. Newark, Delaware: International Reading Association.

Kalın, Ö. U., & Aydemir, A. (2017). 4. sınıf sosyal bilgiler ders kitabının farklı okunabilirlik formüllerine göre incelenmesi. *Studies in Educational Research and Development, 1*(1), 83-108.

Karasar, N. (2012). *Bilimsel araştırma yöntemi* (24 ed.). Ankara: Nobel Akademik Yayıncılık.

Karatay, H., Bolat, K. K., & Güngör, H. (2013). Türkçe ders kitaplarindaki metinlerin okunabilirlik ve Anlaşılabirlirliği. *The Journal of Academic Social Science Studies, 6*(6), 603-623. doi:10.9761/jasss1770

Kasule, D. (2011). Textbook readability and ESL learners. *Reading and Writing, 2*(1), 63-76.

Keskin, H. K., & Akıllı, M. (2013). Fen ve teknoloji ders kitaplarının okunabilirliğinin farklılaştırılmış boşluk doldurma testleri ile ölçülmesi *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 13*(27), 47-66.

Kong, K. (2009). A comparison of the linguistic and interactional features of language learning websites and textbooks. *Computer Assisted Language Learning, 22*(1), 31-55.

Köse, E. Ö. (2009). Biyoloji 9 ders kitabında hücre ile ilgili metinlerin okunabilirlik düzeyleri. *Çankaya Üniversitesi Fen-Edebiyat Fakültesi Journal of Arts and Sciences, 12*, 141-150.

Küçükahmet, L. (2011). *Konu alanı ders kitabı inceleme kılavuzu* (3 ed.). Ankara: Nobel Akademik Yayıncılık.

Marshall, N. (1979). Readability and comprehensibility *Journal of Reading, 22*(6), 542-544.

McLaughlin, G. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12*, 639-646.

Mert, E. L. (2018). Türkçe ders kitaplarındaki metinlerin okunabilirlikleri. *International Journal of Language Academy, 6*(1), 184-198. doi:http://dx.doi.org/10.18033/ijla.3866

Mirzaoğlu, V., & Akın, E. (2015). 5. sınıf Türkçe ders kitabındaki metinlerin okunabilirliği üzerine bir inceleme *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 4*, 167-176.

Oakland, T., & Lane, H. B. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing, 4*(3), 239-252. doi:org/10.1207/s15327574ijt0403_3

Okur, A., & Arı, G. (2013). 6, 7, 8. sınıf Türkçe ders kitaplarındaki metinlerin okunabilirliği *İlköğretim Online, 12*(1), 202-226.

Okur, A., Arı, G., Ersoyal, F., & Okur, E. K. (2013). 5. sınıf Türkçe ders kitaplarındaki metinlerin okunabilirliği. *Sakarya University Journal of Education, 3*(2), 65-79.

Özbek, A. B., & Ergül, C. (2018). İlkokul 4. sınıf ders kitaplarının okunabilirliklerinin değerlendirilmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 14*(2), 653-668. doi:10.17860/mersinefd.398401

Özdemir, S. (2016). Beşinci sınıf Türkçe ders kitabındaki öyküleyici ve bilgilendirici metinlerin okunabilirlik durumu. *Sınırsız Eğitim ve Araştırma Dergisi, 1*(1), 33-46.

Pishghadam, R., & Abbasnejad, H. (2016). Emotioncy: A potential measure of readability. *International Electronic Journal of Elementary Education, 9*(1), 109-123.

Puurtinen, T. (1998). Syntax, readability and ideology in children's literature. *Meta: Translators' Journal, 43*(4), 524-533. doi:10.7202/003879ar

Richards, J. C., Platt, J., & Platt, H. (1992). *Longman dictionary of language teaching and applied linguistics*. London: Longman.

Singh, K. (2007). *Quantitative social research methods*. New Delhi: Sage Publications.

Stevens, K. T., Stevens, K. C., & Stevens, W. P. (1992). Measuring the readability of business writing: the cloze procedure versus readability formulas. *The Journal of Business Communication, 29*(4), 367-382.

Stokes, A. (1978). The reliability of readability formulae. *Journal of Research in Reading, 1*, 21-34. doi:org/10.1111/j.1467-9817.1978.tb00170.x

Tekbıyık, A. (2006). Lise fizik 1 ders kitabının okunabilirliği ve hedef yaş düzeyine uygunluğu. *Kastamonu Eğitim Dergisi, 14*(2), 441-446.

Temur, T. (2002). *İlköğretim 5. sınıf Türkçe ders kitaplarında bulunan metinler ile öğrenci kompozisyonlarının okunabilirlik düzeyleri açısından karşılaştırılması*. (Yayımlanmamış

yüksek lisans tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara, https://tez.yok.gov.tr/UlusalTezMerkezi/ adresinden 11.10.2018 tarihinde erişildi.

Temur, T. (2003). Okunabilirlik (readability) kavramı. *TÜBAR, 13*, 169-180.

Topkaya, Y., Kalın, Ö. U., & Yılar, B. (2015). Readability of the primary education 8th grade citizenshıp and democracy education textbook and its suitability for target age level. *International Journal of Languages' Education and Teaching, 3*(1), 480-491.

Tosunoğlu, M., & Özlük, Y. Ö. (2011). Okunabilirlik ve ilköğretim 1. sınıf Türkçe ders kitabındaki düz yazı metinlerinin okunabilirlik açısından değerlendirilmesi. *Milli Eğitim Dergisi, 189*, 219-230.

Turan, İ., & Geçit, Y. (2010). Lise coğrafya 10. sınıf ders kitabının farklı okullara göre okunabilirliği ve hedef yaş düzeyine uygunluğu. *Kastamonu Eğitim Dergisi, 18*(2), 613-622.

Ulusoy, M. (2006). Readability approaches: Implications for Turkey *International Education Journal, 7*(3), 323-332.

Ulusoy, M. (2009). Boşluk tamamlama testinin okuma düzeyini ve okunabilirliği ölçmede kullanılması. *Türk Eğitim Bilimleri Dergisi, 7*(1), 105-126.

Ülper, H. (2010). *Okuma ve anlamlandırma becerilerinin kazandırılması*. Ankara: Nobel Yayınları.

Walpole, S., & McKenna, M. C. (2007). *Differentiated reading instruction: Strategies for the primary grades*. The Guildford Press: New York.

Wissing, G.-J., Blignaut, A. S., & Van den Berg, K. (2016). Using readability, comprehensibility and lexical coverage to evaluate the suitability of an introductory accountancy textbook to Its readership. *Stellenbosch Papers in Linguistics, 46*, 155-179. doi:10.5774/46-0-205

Yalın, H. İ. (1996). Ders kitapları tasarım. *Milli Eğitim Dergisi*, 61-65.

Yazıcı, K., & Yeşilbursa, C. C. (2007). Sosyal bilgiler derslerinde yazılı materyallerin öğrencilerin okuma seviyelerine uygunluğunun belirlenmesinde kullanılan ölçme araçları. *TSA, 11*(1), 83-98.

Zakaluk, B. L., & Samuels, S. J. (1996). Issues related to text comprehensibility: The future of readability. *Revue Québécoise de Linguistique, 25*(1), 41-59. doi:10.7202/603126ar

Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies, 2*(1), 43-53.

Zorbaz, K. Z. (2007). Türkçe ders kitaplarındaki masalların kelime – cümle uzunlukları ve okunabilirlik düzeyleri üzerine bir değerlendirme. *Eğitimde Kuram ve Uygulama, 3*(1), 87-101.