



## EFL Learners' Perception of Text Difficulty: What Lies Beyond Traditional Text Readability Formulas?

Fatemeh Rezaei<sup>1\*</sup>

Mazandaran University<sup>1</sup>

Correspondence e-mail : fatemehrezaei@gmail.com

**Abstract:** The present study, adopting a mixed methods design, is in line with these recent studies which examines the relationship between text difficulty level and reading comprehension of a group of 112 pre-intermediate EFL learners. Moreover, in addition to two reading tests with less and more difficult passages, a semi-structured interview with 11 volunteer learners were devised to probe the perceptions of the learners about the elements contributing to text difficulty in reading comprehension tests. The results of the quantitative analysis indicated that there was a significant direct strong correlation between the test scores of the two reading tests with less and more difficult passages. In addition, the difference between the observed means of the two tests was significant. This implied that although the learners approached both tests in a similar manner, text difficulty could play a significant role in their achievement in reading test. The results of qualitative analysis also indicated that the role of background knowledge and vocabulary knowledge of the learners, as perceived by the participants of the study, have to be taken into account. Moreover, the text features such as the existence of cohesive devices and other inferential cues are of significance. The results confirm the significance of readability formula as a considerable means of measuring reading difficulty for EFL learners, and introduces the key roles of learners' background knowledge and vocabulary knowledge as sources of perceived difficulty among learners. This is specifically important when determining the texts for different learners various ages and educational levels. An important implication of this study is that teaching and reading assessment need to consider factors such as students' vocabulary knowledge, knowledge background, coherence of the text, as well as the interaction between reader characteristics and text features. It is recommended that English as a foreign language (EFL) educators develop a more holistic approach to selecting and using reading texts taking into account not only traditional readability formulas, but also other important factors such as students' vocabulary knowledge, knowledge background, and level of text coherence.

**Keyword :** L2 reading, Readability, Text difficulty, Learners' Perceptions

**Article info:** Submitted : 2025-03-31 | Accepted : 2025-04-22 | Published : 2025-04-24

Copyright © 2025, Author.

This is an open-access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



How to Cite :

### Introduction

Even though many readability formulas have a variety of weaknesses, they have been adopted by a number of institutions including primary and secondary schools, universities, the military and testing agencies (to name but a few) where they are used to select reading material presumed to be appropriate for a variety of reading skills. Unfortunately, in many cases, readability formulas are used to select texts for

readers and reading situations that go well beyond the limitations of the formulas, as evidenced by a number of recent studies indicating that classic readability formulas (i.e., Flesch Reading Ease, Flesch-Kincaid Grade Level and Dale-Chall) perform worse than other models for different reading tasks, reading domains and populations (Crossley et al., 2008; Crossley et al., 2017; Hiebert, 2011). For instance, recent studies by Crossley et al. (2017) and De Clercq et al. (2014) indicated that classic readability formulas were less predictive of comprehension than readability formulas developed using linguistic features based on word phrases (in the De Clercq et al. study) and on features that measure lexical and syntactic constructs, text cohesion, sentiment, topic analysis and semantics (Crossley et al. 2017).

Despite the fact that there have been several studies based on readability formulas, it seems that more research is needed to delve into different aspects of text difficulty and readability of passages. In line with the existing gap in the literature, the present study aimed at pursuing this line of research and explore the factors contributing to readability of passages in English, other than the ones included in classic readability formulas. This study has two purposes. First, it intends to investigate the relations between text difficulty and reading ability among Iranian EFL learners. The second purpose is to shed more light on the interactions between perceived text difficulty and actual text difficulty in order to give us a clearer image of their importance in reading comprehension for a pre-intermediate group of EFL learners. The following research questions can help us to guide this study is there any relationship between text difficulty level and reading comprehension among Iranian EFL learners? and what is the perception of Iranian EFL learners toward text difficulty?

## Literature Review

In this section, the possible reasons for text difficulty, as identified in the previous studies are enumerated and elaborated.

### 1. Reader Characteristics

It is specified that that among the factors internal to the subject are learners' decoding skills, their prior knowledge on the subject, their abilities to regulate comprehension, among others. Smith et al. (2021) specifies that the reading objectives are related to the student's awareness of the instructions and the intention of reading. Knowing how to read is equivalent to knowing how to do it according to the intentions of the reading activity. That is to say, the reading activity is directed by the objectives, therefore, it is not the same to read to see if it is interesting to continue reading, than to read to look for certain information, or to read to form a global idea of the content and transmit it to another person.

They allow the subject to integrate skills (or strategies) with knowledge. Their role is associated with the self-regulation of cognitive processes. For example, they allow speed and levels of attention to be adapted to objectives, goals, achievement

expectations and perceived difficulty of the task, by activating the knowledge or resources necessary to carry it out. The self-regulation of processes implies constant monitoring, or an "alert" attitude, which allows evaluating the dynamics of the processes themselves. Research is currently focused on elucidating whether monitoring constitutes a pending domain strategy, or a general skill, present in performance in different domains (Elleman & Oslund, 2019). On the other hand, metacognitive strategies lead the reader to become aware of their situation regarding comprehension and consequently adopt the appropriate corrective decisions, in case the objectives of the task are not being achieved.

Reading is not only reduced to decoding words, but also and, above all, it means understanding the written message. Some research on reading difficulties has focused on the problem of decoding, however, it happens that many children decode correctly and do not fully understand what they read. If the decoding is difficult or slow, it will require a greater effort from the student and this will cause them to lose the general meaning of what they are reading. For this reason, it is important to teach students to use the strategies in their reading and to be able to read well with security, fluency and confidence (Yildiz & Çetinkaya, 2017). On the other hand, decoding implies sequential processes that, from simple linguistic units (letters, syllables), proceed in an ascending direction towards more complex linguistic units (words, phrases, texts).

Whose role is to activate and maintain the energy level necessary for the reading to take place. Therefore, the general capacity of the individual to concentrate and sustain his attention exerts a direct action on comprehension. Attention failures can be explained by brain dysfunctions, by the influence of internal aspects of the reader, such as, for example, their attitude towards the task or by contextual aspects (noise levels, for example), Likewise, attention is important to maintain the attention of the students because a good reading and other activities that are carried out in the classroom depend on it (Yildiz & Çetinkaya, 2017).

## **2. Word Reading**

Vocabulary is the essential element both in the comprehension of the text, be it written or oral, as well as in its production. Knowing vocabulary means mastering specific words (nouns, adjectives, verbs, expressions, among others), however, it is often necessary to also know words that would be called "support" words that make possible the explanation when the one that is needed is not known. Therefore, the vocabulary of a language is limited, even if it is very extensive. In the case of Spanish, in the most important dictionaries there are around 80,000 words, some individuals work with 500 and the average that students should know when they finish the secondary period would be about 5000 (Vellutino et al., 2004).

Likewise, it is also an element that influences comprehension, and reciprocally, reading is a primary source for vocabulary. When a text contains difficult vocabulary, comprehension becomes more unattainable; Conversely, when the vocabulary of a text is more accessible to readers, comprehension and learning improve (van den Broek, 2012).

On the other hand, vocabulary, school level and the type of educational center affect reading performance, although to a different degree. The relationship found, however, taking into account that gender was not decisive in any case and that the other variables are very relevant, could be due to other causes (van den Broek & Kendeou, 2017). Since without vocabulary it would be impossible to communicate in a new language. In this way, the learning of vocabulary constitutes one of the aspects that has aroused the most concern among teachers and students of second languages and has originated a multitude of studies that have had a subsequent methodological expression in the field of specific techniques in the classroom. One of the main concerns of teachers is to determine "how much" vocabulary is necessary to learn, to achieve a certain communicative competence, either in oral or written skills.

### **3. Working Memory**

From the cognitive aspect, that is, elements and procedures required to process information, condition reading comprehension. The latter, although it constitutes one of the possible information processing, depends on more general characteristics of cognition operations, within the cognitive are intelligence, memory, perception and attention (Christopher et al., 2012). Working memory is associated with difficulties in syntactic processing when the complexity of the sentences increases but not when they are simple, so that a support option is for the student to learn strategies to understand and remember different grammatical structures that are used to express the same idea (Christopher et al., 2012).

### **4. Vocabulary and Background Knowledge**

Vocabulary (word knowledge) and background knowledge (domain-specific and/or world knowledge) are important sources of variation among readers as they are crucial components in creating a coherent mental representation of the text. In reading comprehension, word and world knowledge closely correspond with each other as depth and breadth of reader's vocabulary can be interpreted as the linguistic equivalent of one's world knowledge (McNamara & Kintsch, 1996).

They refer to the knowledge acquired prior to the reading to be carried out, to what is known about a certain subject or what the children know. This knowledge is activated to understand what is written and retain information. Regularly at school, expository texts are used more than narrative ones, the knowledge of the subjects is usually different and if there is none, they have to be developed; Likewise, the

previous knowledge of the reader is explored before reading the texts, this constitutes an activity that allows establishing relationships between what is already known and the information that has to be acquired (Kintsch, 1998). Therefore, the reader's previous knowledge is very important, since it determines the degree to which the textual meanings can be related and integrated into the schemes that exist in the reader (Magliano et al., 2007).

## 5. Developmental Age

Research suggests that working memory capacity, comprehension processes, and the ability to understand more difficult texts gradually develop throughout childhood (James et al., 2021). As such, a developmental perspective is essential in investigating individual differences in reading comprehension (Sullivan et al., 2023).

## 6. Text Features

The characteristics of the text influence the processing of the reader, but in turn are conditioned by those of the author and the context of production. However, the structure of the text and its content are two important components that influence reading comprehension. Its content is expressed according to a structure. The structure of the text refers to the organization of ideas in the text and the relationships between them (Pickren et al., 2022). All texts are characterized by a logical organization of the information they present. That is, it considers font size, text type, vocabulary complexity, complexity of syntactic structures, text structure, textual instance, and linguistic/paratextual elements.

On the other hand, Ayu (2021) indicates that there are some factors that facilitate the comprehension of texts such as factors external to the subject such as font size, type and class of text, vocabulary complexity, etc. and factors specific to the reader subject such as coding skills, prior knowledge, among others. The microstructure is obtained when a text is read and its meaning is decoded through the construction of a logical relationship, also called the text base, which consists of a set of ordered propositions, which are formed by a predicate and one or more arguments connected to each other by repeating them. The functions are: to identify the basic ideas of the text, to establish a thematic continuity between those ideas, to relate some ideas to others in causal, motivational or descriptive terms (Anne et al., 2020).

On the other hand, the microstructure of the text refers to the minor semantic units that compose it, the form of link between propositions or sentences, the use of deictics, pronouns, tenses, order and relationship between words, among others, that is, the microstructure of a text is constituted by the orderly succession of the different ideas it contains. These ideas, also called propositions, refer to the cognitive processing involved and not necessarily to the grammatical structure (Pickren et al., 2022). The macrostructure refers to the global meaning that permeates and gives meaning to the

local elements, it responds to the fact that we must appreciate those ideas that are central and lend a unitary and globalizing sense to what is read; 16 that is, the most relevant aspects of the text that represent a general vision of the subject, has the following functions: provide a global coherence to the propositions derived from the text, create the macrostructure of a text by individualizing the information, and allows reducing extensive fragments of information to a small number of ideas that can be handled smoothly and without leaving out important information. This can be done with students to find the central idea and the ideas that complement it to understand texts (Anne et al., 2020).

Therefore, the macrostructure, due to its global condition, takes place at the level of the theme or subject, that is, what provides coherence to a text, which relates paragraph to paragraph (Francis, 2018). The superstructure is what allows us to talk about discursive typologies, although not all texts present a clear superstructure. The superstructures are global structures, independent of the content, to which they impose certain limitations. It is a kind of basic scheme to which a text is adapted, for example, an argumentative text has a superstructure made up of an introduction, thesis, development and conclusion; a story has a superstructure formed by a beginning, an action, a reaction and a solution or outcome and a text (Francis, 2018).

## **7. Word Frequency, Sentence Length and Text Cohesion**

The level of inferential comprehension leads to the inference of additional details that the reader could have added, the inference of the main ideas, the inference of the secondary ideas that allows determining the order in which they should be if they do not appear in order in the text, the inference of character traits or characteristics that are not formulated in the text. This level allows the interpretation of a text. Texts contain more information than is explicitly stated. Making deductions involves making use, while reading, of information and ideas that do not appear explicitly in the text. It depends, to a greater or lesser extent, on the reader's knowledge of the world (Schmitz et al., 2017).

Likewise, it looks for relationships that go beyond what has been read, explaining the text more fully, adding information and previous experiences, relating what has been read with our previous knowledge, formulating hypotheses and new ideas. The goal of the inferential level will be the elaboration of conclusions. This level of comprehension is rarely practiced in schools, since it requires a considerable degree of abstraction on the part of the reader. It favors the relationship with other fields of knowledge and the integration of new knowledge into a whole (Follmer & Sperling, 2018).

The sentence is the basic linguistic unit and the starting point of grammatical analysis. By such must be understood a linguistic-phonetic (or graphic, of course) entity that prototypically complies with these two requirements: to be a global unit of

representation of what is perceived or experienced and is desired to be transmitted to another; a kind of "minimum unit of communication", a statement, therefore being composed of (at least two) elements/words that have close internal relationships. On the other hand, sentence comprehension is a complex and "active" process that requires various cognitive and linguistic capacities that allow decoding and understanding the minimum constituent units until the elaboration of the sentence. mental representation of the message it transmits (Amendum, 2018). In general, interpreting what is meant through a sentence requires language processing that can be analyzed at a structural level from the phoneme itself to the speech.

Therefore, the comprehension of sentences is divided into two processes: the separation of the phrases of the sentence (subject, predicate); and the assignment of thematic roles to those phrases. The first of these processes is more superficial, linked to the grammar, and is carried out largely automatically, while the second requires conscious control (McNamara & Kintsch, 1996).

## 8. Genre

The paragraph is the thematic unit or internal coherence shared by the linguistic material housed within it, and the paragraphs have certain marks (most likely referring to full stop, line indentation, and first capital letter), which help to identify them in writing (Andreani et al., 2021). On the other hand, the paragraphs are separated from each other by full stops that show the transition from one to the other; the paragraphs can be enumeration, sequence, comparison or contrast, development of a concept, cause and effect (McNamara & Kintsch, 1996). But the types of paragraphs can be different, not only because of their location within the text (in the opening, development or closing), but also because of the function that each of them fulfills. The paragraphs can be opening (introduction and thematic), development (citation, comparative, descriptive, conceptual, argumentative, counter argumentative) and closing (Schmitz et al., 2017; Vellutino et al., 2004)

## Methodology

### Participants

Participants of this study were a group of 60 female and male EFL learners studying in a general English course. Their age ranged from 18 to 26. They were selected from original body of 112 students, based on their scores on Oxford Placement Test (OPT); that is, the learners whose scores fell within  $\pm 1$  standard deviations from the mean score ( $M = 21.08$ ;  $SD = 3.10$ ) were selected to take part in this study; so, their scores ranged from 16 to 27. The learners had studied English for six years at high school and some of them had been studying English in private institutes when they participated in data collection. Their first language was Persian. It is worth mentioning that the learners were selected via convenient sampling, that is, they were selected

based on their availability to the researcher and their volunteer participation in the study. They consented with the terms described by the researcher before their participation in this research. After the quantitative data were collected, 11 students volunteered to take part in the interview sessions. They included seven female and four male students.

## **Instruments**

Instruments which are mentioned below were used in this study.

### **1. Oxford Placement Test (OPT)**

In order to homogenize the participants, the Oxford Placement Test (OPT) was given to each EFL learner. The Oxford Placement Test is ideal for testing students' English level quickly and easily, place millions of students in the right class. The Oxford Placement Test mainly measures English use, in terms of vocabulary, grammar and reading comprehension. It contains 60 questions and the learners are allotted 40 minutes to answer the items. The reliability of the test in this study was measured to be .82, using KR-21 formula based on the data collected from 26 students.

### **2. Reading Comprehension Tests of Different Difficulty Levels in English**

The reading tests of different difficulty levels were selected. However, to calculate the readability of the test, the Flesch Reading Ease Readability Formula was employed. Two tests were used in this study, one with lower and one with higher level of difficulty (as measured by the aforementioned readability formula). The purpose of the selection of two tests with different levels of difficulty was to examine if the same sample of learners of English with a specified level of general English proficiency perform in a unified manner on two tests with different levels of text difficulty. To do so, two tests, one with lower level of difficulty and one with higher level of difficulty were selected.

#### **a. Test of Reading Comprehension at Appropriate Difficulty Level**

According to Shorkaei and Talebi (2018), the reading comprehension with a lower-level difficulty was called Appropriate Difficulty Test (ADT). This test consisted of five passages each with 8 multiple choice items. Difficulty level of the passages was measured using the Flesch and Flesch-Kincaid readability formulas, which are discussed in details in the procedures section, and was found to enjoy the observed mean of 77.64 as measured by Flesch formula and 5.64 as measured by Flesch-Kincaid formula. These statistics confirmed its appropriateness linguistic level. The researcher consulted with three experts in TEFL about difficulty of the texts. The readability indices of the passages included in the test are shown in Table 1.

**Table 1.**  
**The Readability Indices of the Passages in ADT**

Passage	Words	Paragraphs	Sentences	Sentence per paragraph	Words per sentence	Flesch	Flesch- Kincaid
1	604	10	43	4.3	14.0	74.6	6.3
2	606	7	45	6.4	13.4	74.6	6.1
3	607	8	45	5.6	13.4	76.7	5.9
4	6.4	9	47	5.2	12.8	78.6	5.4
5	603	9	50	5.5	12.0	83.7	4.5
<b>Average</b>						<b>77.64</b>	<b>5.64</b>

Day and Park (2005) proposed a taxonomy of the types of comprehension questions held to help students comprehend the text better with the aim of becoming interactive readers. Their taxonomy included literal comprehension, reorganization, inference, prediction, evaluation, personal response types of reading questions to be utilized by teachers and material developers. Nevertheless, literal comprehension, reorganization, inference, and vocabulary questions were within the scope of testing reading comprehension for determining the text difficulty and the learners' perceptions about it in this study. In addition, as Shorkaei and Talebi (2018) stated, these items were selected since they were more objective for the purpose of scoring. Literal comprehension concerns an understanding of the direct and explicit meaning of the text. Reorganization centers on the literal understanding of the text.

Nonetheless, it is more complex than literal comprehension questions, as students should move to a more holistic, global view rather than a sentence-by-sentence understanding of the text. Readers must piece together information collected from various parts of the text for more comprehension. To reply to an inferential question, students need to use a combination of the literal comprehension of the text with their knowledge and intuitions, as the answer to this type of question is not explicitly stated in the text. Making inferences involves more than a literal understanding. Students may initially have a difficult time answering inference questions because the answers are based on material that is in the text but not explicitly stated. An inference involves students combining their literal understanding of the text with their own knowledge and intuitions. Finally, the researchers decided to put vocabulary knowledge into the respective category as without knowing a word and its structure, finding its meaning is to a great extent difficult. The reliability of the test was taken care of at the piloting stage through the KR-21 formula which turned out to be 0.79. The time allowed was 40 minutes as determined at the piloting stage.

b. Test of Reading Comprehension at High Difficulty Level

According to Shorkaei and Talebi (2018) this test was called High Difficulty Test (HDT). Similar procedure, as described above for ADT, was followed for developing this test; however, the only difference was that the texts selected for making this test were of higher levels of difficulty, as measured by Flesch and Flesch-Kincaid readability formulas. In developing this test five passages were included and each of them was followed by 8 multiple choice questions. Difficulty level of the passages was measured using the Flesch readability formula and the observed mean of the difficulty level of the passages was found to be 60.76 as measured by Flesch formula and 8.62 as measured by Flesch-Kincaid formula. In addition, the test was shown to three experts who evaluated ADT in terms of its appropriacy of difficulty level and they confirmed that HDT is beyond the participants current level of reading ability in terms of its difficulty. The types of questions were similar to those of the reading test at low difficulty level and included literal comprehension, reorganization, inference, and vocabulary questions. The readability indices of the passages included in the test are shown in Table 2.

**Table 2**  
**The Readability Indices of the Passages in HDT**

Passage	Words	Paragraphs	Sentences	Sentence per paragraph	Words per sentence	Flesch	Flesch- Kincaid
1	600	10	38	3.8	15.7	56.5	9.2
2	601	6	38	6.3	15.8	66.3	7.9
3	591	7	40	5.7	14.7	58.5	8.7
4	601	9	38	4.6	15.8	57.3	9.1
5	619	8	37	4.6	16.7	65.2	8.2
<b>Average</b>						<b>60.76</b>	<b>8.62</b>

Similar to ADT, Day and Park (2005) proposed taxonomy for the types of comprehension questions were the basis for developing comprehension question, including literal comprehension, reorganization, inference, prediction, evaluation, personal response types of reading questions. In this study, as described in the previous section, literal comprehension, re-organization, inference, and vocabulary types of questions were included due to the fact that these types of questions were in line with the purpose of testing reading comprehension for determining the text difficulty and the learners' perceptions about it in this study; that is, as Zabihi Shorkaei and Talebi (2018) stated, these reading comprehension question types were more objective for the purpose of scoring. Literal

comprehension questions measured learners' understanding of the direct and explicit meaning of the sentences. Reorganization question types focused on the literal comprehension of the sentences, at a more complex level than that of literal comprehension questions, since the test takers need to consider a more holistic, global view rather than a sentence-by-sentence understanding of the text. In other words, the learners needed to scan for information from different parts of the text to attain a deeper comprehension of the passage. Inferential question types required the learners to combine their literal comprehension of the text with their knowledge and intuitions, since these types of questions did not include explicit statement of message. Inferential question types involved a deeper understanding and measured a level of understanding which was beyond literal understanding. Finally, vocabulary question types were also included in HDT which required the learners to find meaning of the target word with reference to the co-text. The reliability of the test was taken care of at the piloting stage through the K-R21 formula, which turned out to be 0.77. The learners had 40 minutes, similar to the parallel test, ADT, to take the test.

### **3. A Semi-structured Interview**

A semi-structured was conducted with volunteering participant. In-depth semi-structured interviews are used to build up a complete picture of the participants' thoughts (Hashemi, 2014, p.175). After receiving permission from each interviewee, the interview was conducted and recorded after their reading comprehension tests. In the first part of the interview, the focus of the questions perceived attitude and perceptions toward the two tests of reading comprehension they took. In the second part of the interview the focus of the questions was on students' perception toward the difficulty of the given tests. the interview was conducted in Persian and on one-on-one basis in the institute, after the researcher described the purpose of the interview to the participants and got their consent to participate in the research with regard to the research terms. The interviewees were promised that their interviews will be kept confidential.

### **Data collection procedure**

The following procedures were taken in order to collect the data in this study. After obtaining consent of students to participate in the study, an Oxford Placement Test was given to homogenize the learners in terms of their general English proficiency. To this end, the researcher entered 11 classes in the institute and gave the test to the intact classes. Their scores were used to select the learners. Those who scored within two standard deviations ( $M = 21.08 \pm 3.10$ ) above and below the mean were selected; so, the scores ranged from 16 to 27.

Then, the researcher prepared two sets of reading tests with two different levels of difficulty. The next stage was to pilot the reading comprehension tests. In order to do so, ten learners from the sample were asked to take the reading tests first. Viewing the fact that the tests were practical and reliable enough for the purpose of the research, they were distributed among the rest of 50 learners on other classes. Care was taken to keep the administrative conditions the same when the tests were conducted in two administration sessions. For the sake of practicality, the test with low level of difficulty was administered first, and a week later the second test, the one with high level of difficulty was administered.

After the researcher scored the papers, the students were asked to take part in the interview. The purpose of the interview was described and the learners were asked to take part in the interview voluntarily. They were told that the interview was just for the sake of research and their comments and participation would not affect the scores they obtain in that course. 11 learners participated in interviews. Meanwhile, the interview prompts were prepared via consulting the available literature on reading comprehension studies which have been conducted via qualitative methods. Each interview was held in the institute in a room which was quite enough to voice record the session and the researcher and interviewee had enough privacy to do the interview without any disturbance. The interviews were held in Persian.

Finally, after collecting the data, the qualitative data were analyzed based on undirected content analysis method in which open coding was the main technique used for identified relevant codes and grouping them into themes. The audio-recorded content was transcribed, and categorized; the researcher then read through transcripts several times and made notes of important features with the intention of finding patterns in the data set; finally, the relevant codes were classified and grouped into the themes.

The quantitative data were analyzed with SPSS 26.0 software, a software package for analyzing quantitative data. To analyze the scores from the reading comprehension tests, descriptive indices including mean, standard deviation, skewness and kurtosis of the scores were calculated. The scores were then checked to see if the distributions were normal. Then, Pearson correlation test was conducted to see if the two sets of scores were correlated. This was part of paired samples t-test analysis which aimed at determining if text difficulty is a significant factor in comprehensibility of passages.

## Result and Discussion

### Results of the First Research Question

The first research question of the study was:

*Is there any relationship between text difficulty level and reading comprehension among Iranian EFL learners?*

The following null hypothesis was formulated based on this research question:

*There is not any relationship between text difficulty level and reading comprehension among Iranian EFL learners.*

To test the hypothesis, first descriptive analysis of the scores from the two reading tests was conducted. The results are shown below.

**Table 3.**  
**Descriptive Statistics for Reading Tests**

	N	Mean	SD	Skewness	Kurtosis
Reading (Low Difficulty)	60	22.10	2.04	.44	-.36
Reading (High Difficulty)	60	16.83	2.26	.40	-.18

As shown in Table 3, there is a huge difference between the observed maximum and minimum scored in each set of data. In addition, the learners scored much higher on the test with less difficult passage. However, the standard deviation of the scores were nearly similar. In terms of the distribution of the scores, it has to be mentioned that since the observed skewness and kurtosis were between  $\pm 1$ , it was concluded that the distribution was normal.

To further test the distribution of the reading test scores in terms of their normality of distribution, Shapiro-Wilk test was used. This test, according to Dornyei (2007) is most suitable with small sample sizes. The results are shown below.

**Table 4.**  
**Tests of Normality for Reading Test Scores**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Reading (Low Difficulty)	.18	60	.20	.95	60	.06
Reading (High Difficulty)	.19	60	.20	.95	60	.07

a. Lilliefors Significance Correction

Based on the observed statistics in Table 4, it was concluded that the distributions of the reading tests, both the one with less difficult passages and the one with more difficult passages were normal ( $\text{sig.} = .06$  and  $.07 > .05$ ). Accordingly, it was concluded that parametric test, paired sample t-test and Pearson correlation test could be used. The results are shown below.

**Table 5.**  
**Paired Samples Correlations**

	N	Correlation	Sig.
Reading (Low Difficulty) & Reading (High Difficulty)	60	.841	.000

Based on the statistics in table 5, it was argued that the first hypothesis of the study, stating that there is not any relationship between text difficulty level and reading comprehension among Iranian EFL learners, was rejected since there was a significant direct strong correlation between the test scores of the two reading tests ( $r = .84$ ,  $p = .00 < .05$ ). This implies that the learners approach the two types of texts with different levels of difficulty in the same manner.

### Results for the Second Research Question

The second research question of the study was:

*What is the perception of Iranian EFL learners toward text difficulty?*

As stated earlier, a semi structured interview was conducted and the data from the interview were analyzed. The results are presented in the following sections which are indeed the themes emerged from the data. The analysis was conducted based on the guide for content analysis and thematic analysis provided by Neuendorf (2018). In accordance with this framework, what participants share in the interviews is analyzed in three steps for the sake of qualifying the results. The first step is analyzing narrative materials of what the participants have experienced. The researcher adopts a realist/essentialist and constructionist, factist perspective to the narrations. Description and interpretation of the narrative is the first step within a realist frame of reference and examples provided; both inductive and deductive analyses are conducted, emphasizing on the context of the participants' experience within constructionist frame of reference; then, integration of manifest and latent contents is established, drawing thematic map. In this approach, no peer checking is necessary and for the sake of reliability of analysis, a second confirmatory analysis is done.

It was found that, with reference to the related codes arisen in the comments and answers of the 11 participants, the reader traits seem to be the most dominant in explaining the test scores, as their cognitive abilities can pave the way for comprehension of the passages regardless of their type and features. Meanwhile, it

was seen that textual features, and processes involved in comprehension are less vital in justifying the variations occur while reading. Consequently, it was concluded that variations observed in responses to reading test items may be a consequence of inconsistency in test-takers' abilities and features of items or texts are that strong to determine what was observed.

### 1. Reader characteristics

The findings of qualitative analysis also unveiled that background knowledge and vocabulary were among the most influential factors explaining the learners' perceptions of text difficulty. Those learners who had higher levels of background knowledge and vocabulary, in comparison with readers with lower levels, enjoyed higher level of confidence and scored higher."

For example, a learner talking about the reading text with more difficult texts mentioned that:

Excerpt 1:

*"When I was reading the passage sometimes there were a chain of new words so that I could not grasp head or tail of the passage or at least those few sentences. I could not relate it to any question". (Shiva)*

Similarly, with regard to the background knowledge, a learner mentioned that it helped her when reading the passage with higher level of difficulty.

Excerpt2:

*" When coming up with a familiar issue in the passage I could easily read it... I mean it was easier to read because I could partly guess what was said. I mean I could better relate the choices in the question with the text". (Mina)*

Similarly, another learner stated that:

Excerpt 3:

*"Sometimes, you find a passage similar to what you had read before, I mean in terms of the content. In this case, sometimes I felt that I had read this passage in Persian for example and I could better understand what was going on". (Mehdi)*

### 2. Features of the Passage

The results revealed that certain passage features explained some of the variation in item difficulties. For example, a learner talking about the first test with less difficult passages stated that:

Excerpt 4:

*“The passage had short sentences... I mean the sentences were manageable, ... when you read a sentence sometimes you see there is a sentence in the heart of the other sentence .. it is like a puzzle and it is hard to find the relation of ideas un a sentence... let alone the relation between two or more sentences”. (Telma)*

Similarly, another learner mentioned that commonly used conjunctions could be a great help.

Excerpt 5:

*“There are a number of clues like and, but, however, because, if ... which always help you understand sentences or ideas are related. In the second test [the one with more difficult texts], there were fewer words like these and you had to guess the relations, you know what I mean”. (Pouria)*

### 3. Text-reader interactions

The results indicated that interactions between referential cohesion and comprehension processes “tapped by a test item (i.e., text memory or text inference), referential cohesion and vocabulary, word frequency and vocabulary as well as working memory and deep cohesion were the most important in explaining variation in reading comprehension test scores.”

With regard to the reading test which included more difficult passages, one of the learners stated that:

Excerpt 6:

*“There was a kind of relation between the options we had for a question and the text... for example, I could see a word in the choices and this word was kind of related to what I remembered from the passage ... I mean a word or a phrase or a name or something. This is important to have your key words in your memory when answering reading questions”. (Hana)*

### Discussion

The purpose of this study was to investigate the sufficiency of readability formulas in terms of accounting the factors influencing text difficulty for EFL learners. The results of the study revealed that the factors included in readability formula such as the number of morphemes, the length of the sentences, among others, can predict text difficulty; however, they are not enough in terms of judging text difficulty for EFL learners. The results of qualitative analysis depicted that there a number of learners variables such as learners’ background knowledge and vocabulary knowledge have to

be included in text difficulty assessment. The quantitative analysis of the scores showed that HDT mean score was lower than that of the ADT. The reactions of the learners to the texts of the two tests when answering the first interview questions also showed that they perceived HDT to be considerably more difficult than ADT. Based on the learners' comments on their perceived factors leading to their perceptions of text difficulty, there were some comments which were common to both test types, the with low difficulty level and the one with high difficulty level. The findings are consistent with the previous finding that making a clear mental picture of a text is dependent on learners' breadth and depth of vocabulary and background knowledge (e.g., Kintsch, 1998; Magliano et al., 2007; McNamara & Kintsch, 1996). Broader and deeper vocabulary knowledge impacts fluency and accuracy of word reading, and capacity for comprehension; additionally, it helps them with extracting new word meaning and fitting words within their mental text depiction (Van den Broek et al., 2014). The results obtained from the interview with regard to breadth and depth of background knowledge and vocabulary knowledge are in agreement with what was observed in the quantitative section of the study. What we measure via the readability formulas, the vocabulary difficulty, sentence length and structural complexity are perceived to be the sources of difficulty; however, we need to move beyond this frame and take account of what is not measured by the number of words, sentence length, and other structural features. The qualitative findings of this research seem to be beyond what is measured via readability formula, that is, background knowledge or cognitive familiarity with the text.

Based on the "comments collected about the test with more difficult texts, higher level of background knowledge helps learners with digesting information embedded in the passage, connecting textual elements, filling in informational gaps, and linking in formation from texts with background" knowledge (Golke et al., 2022; Strohmaier et al., 2023). Those who do not have sufficient vocabulary or background knowledge suffer from limited ability in understanding different texts since knowledge of words and their meanings (breadth of vocabulary) is inadequate for making successful inference and integration of text information with their background knowledge (Zano, 2022).

The results clearly supported that understanding of texts be enhanced through acquiring new words and expanding general knowledge. Familiarity with more words and topics augments reading texts of different levels of difficulty. The results are in agreement with the previous findings in that expository may be difficult since familiarity with different topics are needed for grasping the meaning of the text (Kintsch, 1998). Additionally, the difficulty may be due to the lack of frequent words and existence of informational blanks which may be hard to fill if the test taker lacks enough background knowledge (Magliano et al., 2007). Difficult texts usually require deeper background (domain-specific) knowledge (McNamara & Kintsch, 1996).

Based on the participants' beliefs expressed about difficult texts, it was seen that cohesion has a central role in helping test-takers subordinate the concepts in the text (Schmitz et al., 2017). That is, deep textual cohesion is chiefly valuable if the test-takers try to understand harder texts. Consequently, comprehending more coherent texts is less difficult due to use of connecting words clarifies the relations among elements in the passage. That is, these passages ease comprehension as cohesion among ideas are more explicit."

The results revealed traditional indices of readability (counting on word frequency and sentence length) "and cohesion might not be useful in determining reading test scores in all situations. It can be argued that passages with less familiar words, longer sentences and low cohesive features are harder to understand" (Amendum, 2018; Follmer & Sperling, 2018). Overlap of familiar concepts and words in highly cohesive passage makes recollecting information from the passage easier since it stimulates comprehension via stating the ideas presented in the passage. Thus, reactivating concepts provokes better comprehension as they intermittently appear in a text. In the same vein, texts which are more cohesive are easier because they involve smaller number of informational gaps, so making inference would not be tough since there is overlapping information."

Also, a less skilled reader tends to struggle with texts which involve many unacquainted lexical items and use more explicit word to reduce informational gaps which he or she has to infer (James et al., 2021). Moreover, the results showed that a reader who has better working memory enjoyed a more positive attitude toward the test items from passages that have smaller cohesive links. It was found that if the learners had better working memory, they would have more resources to manipulate and store processed information, so they can infer and integrate information from the passage with their background knowledge to form a coherent picture (Sullivan et al., 2023).

It can be concluded, based on what was found in this study, the test-takers process more or less difficult texts almost in the similar manner. That is, good readers can process texts with difficult levels of difficulty, well. They have compensatory skills that help them with understanding difficult passages and read a variety of passages with different levels of difficulty."

McNamara and Kintsch (1996) showed that texts with fewer cohesive devices are more helpful if readers have extensive background knowledge, but texts with more cohesive devices benefit readers with lower level of background knowledge. In the same manner, informational gaps make readers to infer better and fill the blanks easier. That is, they require to connect and construct a clear picture of the passage.

An important implication of this study is that teaching and reading assessment need to consider factors such as students' vocabulary knowledge, knowledge background, coherence of the text, as well as the interaction between reader

characteristics and text features. To improve their reading skills, students need to expand their vocabulary and general knowledge, while educators should provide texts with appropriate levels of difficulty – not so easy that they don't encourage active thinking, but also not so difficult that they hinder comprehension. Material developers, syllabus designers, and EFL reading test makers need to adopt a more comprehensive approach to assessing text difficulty that goes beyond traditional readability formulas, taking into account the varied characteristics of the reader and the complex interactions between the reader and the text.

## Conclusion

The results of this research can be informative since they showed how there is an interplay of text features, reader characteristics and comprehension processes. We saw that the links between text and reader leaves an impact on performance variability of learners on test items. More precisely, it can be argued that texts of certain types might be more demanding for test-takers of different abilities and skills, if linguistic characteristics of a passage, learners' vocabulary and background knowledge is taken into account.

The results also demonstrated less-skilled test-takers are challenged more seriously if texts have specific characteristics. As a result, these features help readers identify reader's ability, and make a foundation for foreseeing how students will perform on particular types of items addressing particular kinds of texts on future examinations of reading comprehension. This study also paves the way for planning interventions for changing learners' performance."

The results of this research have a number of implications in terms of pedagogy and assessment. That is, the results can be pedagogically valuable for EFL teachers, material developers and syllabus designers. It can also be valuable for those educators who are developing reading tests.

"There are several important implications which emerged from the present study. First of all, understanding and learning from the text may be improved by learning new words and familiarizing oneself with novel concepts. As we learn new words, we become more experienced and gain better understanding of what authors are trying to convey. As such, improving vocabulary affords readers the opportunity to extend their reading to ever more complex text. It is clear that word frequency affects text difficulty, and knowledge of word meanings mitigates that effect on comprehension."

"Second, good readers read well across a variety of texts as they are more equipped to deal with challenges introduced by the texts. However, it is important to also keep in mind that readers' learning experiences are maximized when they are forced to actively think about ideas presented in the text. In other words, texts should

not be too easy as simple texts do not encourage the reader to think about conveyed information.”

“Comparing the results of the quantitative results and qualitative findings of the study, we can argue what we discovered through the analysis of the interview data in terms of the role of linguistic features of the texts, the role of vocabulary knowledge, and the contribution of the learners’ background knowledge, is far beyond what we can measure quantitatively. The results, in addition to the identification of the elements which may be influential in determining text difficulty, prove that the quantitative approach to the estimation of text difficulty, most importantly, the readability formulas, are enough but not comprehensive. It seems that the formula has to be modified to account for other factors such as the degree of familiarity of the readers; however, in this case the results of readability formula may not be as objective as they are known.”

Although it was not examined in the present study, a related concept to knowledge is that of student interest. Accordingly, future studies may include the reader’s motivation and attitude into consideration. One of the issues raised in this study, especially when the qualitative data were interpreted, was the significance of working memory and its contribution in making differences between the learners in terms of their performance on reading comprehension tests. Future research can encompass its role in the learner’s performance and their achievement on reading tests.

The recommendations of this study are for English as a foreign language (EFL) educators to develop a more holistic approach to selecting and using reading texts by considering not only traditional readability formulas, but also other important factors such as students' vocabulary knowledge, knowledge background, and level of text coherence. Teachers are advised to actively expand students' vocabulary and build background knowledge before reading complex texts, as well as to use texts of appropriate difficulty –challenging enough to encourage active thinking but not so difficult as to hinder comprehension. Curriculum developers and reading test makers should revise the text selection criteria to include aspects of reader-text interaction, consider the use of appropriate cohesive tools for readers with different skill levels, and design learning interventions that improve working memory and compendative reading strategies. Finally, it is recommended to develop a more comprehensive reading assessment system that measures not only the understanding of the text but also the ability of students to make inferences, integrate information, and relate new knowledge to existing knowledge.

## References

- Amendum, S. J., Conradi, K., & Hiebert, E. (2018). Does text complexity matter in the elementary grades? A research synthesis of text difficulty and elementary students' reading fluency and comprehension. *Educational Psychology Review, 30*, 121-151.
- Andreani, S., Muniroh, S., Suharyadi, S., Astuti, U. P., & Yulizar, Y. (2021). The contribution of genre awareness and reading habits towards students' reading comprehension. *Indonesian Journal of Applied Linguistics, 11*(2), 463-476.
- Anne E. Mesmer, H., Hiebert, E. H., Cunningham, J. W., & Kapania, M. (2020). Does one size fit all? Exploring the contribution of text features, text content, and grade of use on comprehension. *Reading Psychology, 42*(1), 42-72.
- Arthur, J. (2016). Convergence on policy goals: Character education in East Asia and England. *Journal of International and Comparative Education (JICE), 59-71*.
- Ayu, M. (2021). Applying Text Feature Walks Strategy in Higher Education to Improve Students' Reading Comprehension. *Journal of English Teaching, 7*(3), 354-365.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *Tesol Quarterly, 42*(3), 475-493.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior research methods, 49*(3), 803-821.
- Day, R. R., & Park, J. S. (2005). Developing Reading Comprehension Questions. *Reading in a Foreign Language, 17*(1), 60-73.
- De Clercq, O., Hoste, V., Desmet, B., Van Oosten, P., De Cock, M., & Macken, L. (2014). Using the crowd for readability prediction. *Nat. Lang. Eng., 20*(3), 293-325.
- Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences, 6*(1), 3-11.
- Follmer, D. J., & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences, 67*, 177-187.
- Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the simple view of reading to account for variation within readers and across texts: The complete view of reading (CVR i). *Remedial and Special Education, 39*(5), 274-288.
- Golke, S., Steininger, T., & Wittwer, J. (2022). What makes learners overestimate their text comprehension? The impact of learner characteristics on judgment bias. *Educational Psychology Review, 34*(4), 2405-2450.
- Hashemi, M. R. (2014). *Understanding Research in Applied Linguistics: SAMAT*.

- Hiebert, E. H. (2011). Beyond single readability measures: Using multiple sources of information in establishing text complexity. *Journal of Education, 191*(2), 33-42.
- James, E., Currie, N. K., Tong, S. X., & Cain, K. (2021). The relations between morphological awareness and reading comprehension in beginner readers to young adolescents. *Journal of Research in Reading, 44*(1), 110-130.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review, 95*(2), 163.
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. *Reading comprehension strategies: Theories, Interventions, and Technologies, 107-136*.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse processes, 22*(3), 247-288.
- Pickren, S. E., Stacy, M., Del Tufo, S. N., Spencer, M., & Cutting, L. E. (2022). The contribution of text characteristics to reading comprehension: Investigating the influence of text emotionality. *Reading Research Quarterly, 57*(2), 649-667.
- Schmitz, A., Gräsel, C., & Rothstein, B. (2017). Students' genre expectations and the effects of text cohesion on reading comprehension. *Reading and Writing, 30*, 1115-1135.
- Schmitz, A., Gräsel, C., & Rothstein, B. (2017). Students' genre expectations and the effects of text cohesion on reading comprehension. *Reading and Writing, 30*, 1115-1135.
- Shorkaee, H. Z., & Talebi, S. H. (2018). Effects of reading strategy instruction on attitude toward strategies and performance in reading texts of different difficulty levels. *The Reading Matrix: An International Online Journal, 18*(1), 64-81.
- Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology, 42*(3), 214-240.
- Strohmaier, A. R., Ehmke, T., Härtig, H., & Leiss, D. (2023). On the role of linguistic features for comprehension and learning from STEM texts. A meta-analysis. *Educational Research Review, 39*, 100533.
- Sullivan, S. D., Grant, C. E., & Patel, S. (2023). Early Intervention for Kindergarten-Aged Students: Providing an Academic Advantage to Promote Reading Growth. *Educational Research Quarterly, 46*(3).
- Van den Broek, P., & Kendeou, P. (2017). Development of reading comprehension. *Theories of Reading Development, 15*, 283-306.
- Van den Broek, P., Virtue, S., Everson, M. G., Tzeng, Y., & Sung, Y. C. (2014). Comprehension and memory of science texts: Inferential processes and the

construction of a mental representation. In *The psychology of science text comprehension* (pp. 131-154). Routledge.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2-40.

Yildiz, M., & Çetinkaya, E. (2017). The Relationship between Good Readers' Attention, Reading Fluency and Reading Comprehension. *Universal Journal of Educational Research*, 5(3), 366-371.

Zano, K. (2022). Breadth and Depth-vocabulary knowledge and reading comprehension in an English first additional language context. *Journal of Languages and Language Teaching*, 10(2), 223-233.