

Comparative Analysis of Reading Text Readability in Chinese Junior High School English Teaching Textbook Based on Corpus

Wu Mengyao

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

Helmi Norman

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
Corresponding Author: helmi.norman@ukm.edu.my

Nor Hafizah Adnan

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

Received:05/17/2023

Accepted:09/06/2023

Published: 09/24/2023

Abstract

Readability is one of the important factors for students to consider when choosing a text to read. Based on a self-built corpus of Junior High School textbooks in China, this paper mainly investigates the differences in readability between reading texts in junior high school English textbooks. From vocabulary, sentence, and comprehensive indicators, this study makes quantitative statistics on the factors that affect the readability of English textbook texts to understand the readability of different textbook texts and to explore whether there are differences in the readability of eight textbooks. Finally, the results show significant differences in the readability of the eight editions of English readers. The average word length and new words rate are the highest in Popular Science Press, which reach 4.43 words and 8.40%, respectively. Among the average sentence length and the frequency of compound sentences, the Shandong Education Press has the highest, going 11.36 words and 11.03%, and confirmed that the Shandong Education Press has the highest difficulty. The textbook arranger can take the difficulty of the reading text into account once they have determined the readability of the text. They can then arrange or select a textbook appropriate for students, enabling them to comprehend and master the text more effectively.

Keywords: comparative analysis, Junior High School English teaching, textbook corpus, reading text readability

Cite as: Mengyao, W., Norman, H., & Adnan, N. H. (2023). Comparative Analysis of Reading Text Readability in Chinese Junior High School English Teaching Textbook Based on Corpus. Arab World English Journal (AWEJ) Volume 14. Number 3 September 2023) 431-444.
DOI: <https://dx.doi.org/10.24093/awej/vol14no3.28>

Introduction

English is an international lingua franca and a critical media tool for worldwide political debate and trade, for communicating global information and technology, sharing resources, and education (Cheng, 2002). Junior High School is a crucial time to establish the groundwork for future development and a significant stage in the primary education period in China (Song, 2010). Teaching materials are the primary source of students obtaining knowledge and the main basis for teachers' teaching, and they are the key to implementing the curriculum standards (Chen, 2010). As time fast forward with the rapid increase of technology, building first-class human capital in fulfilling the needs of today's world is vital at its best (Adnan & Sayadi, 2021). Textbooks are the most essential materials in teaching tools, and the setting of the difficulty of teaching materials is also essential for teachers' and students' learning. Reading text is a necessary part of the textbook. It is vital to support and scaffold students' literacy challenges by selecting texts of difficulties appropriate for their reading abilities (Crossley et al., 2022). The readability of reading text has long been regarded as a significant aspect of English language teaching as it provides the overall picture of a text's difficulty level, especially in teaching and testing (Srisunakrua, 2019).

The six chapters that comprise the majority of this study include the introduction, literature review, research methods, results, discussion, and conclusion of the reading text readability. In this paper, we created a corpus of English textbook reading texts for junior high school students in China, briefly described the corpus-building process, organized and summarised information about textbook publishing and reading texts in-depth, and examined the readability of the textbook reading texts from the standpoint of the factors affecting readability. The goal of this paper is to answer the following questions:

- 1) How difficult are the eight editions of the Junior High School English textbook reading text?
- 2) Are there significant differences in reading text readability among the eight editions of Junior High School English textbooks?

Literature Review

The research on text readability is significant for teaching pupils according to the environment and their aptitude. Klare (1963) describes readability as "a style of writing that is easy to understand"; McLaughlin (1969) views it as "the degree to which the reading content is easy to understand for a specific group of people." In general, the text readability of English textbooks describes how simple the reading materials are to understand (Lin, 1995; Li, 2000).

Thorndike (1921) was one of the first scholars to study the readability of English textbooks, publishing 1921 *A Teacher's Word Book of 10,000 Words*, which looked closely at the vocabulary used in school textbooks. Since then, Flesch (1948), Betts (1949), Dechant & Smith (1961), and others have made significant contributions to the study of English readability. Their research broadly focuses on two topics: first, what influences English readability, and second, what metrics are utilized to quantify language readability scientifically. This essay makes a significant comparison between the readability of various editions of Junior High School English textbooks from the standpoint of the elements that influence readability.

Betts (1949) pointed out the main factors affecting readability: the average number of words, the number of simple sentences and preposition phrases, the percentage of different words,

etc. Dehant & Smith (1961) summarized the factors affecting the readability of reading materials: word length, the proportion of new words, sentence length, personal pronouns, syllables, algebras, prefixes, etc. Fry (2002) argues two factors: syntactic and Semantic difficulty. Richard & Schmidt (2003) claimed that it depends on many factors, mainly including the average sentence length of the article, the number of original words, and the complexity of the grammar. In summary, many factors influence readability, and we can roughly divide these factors into two main aspects: one aspect is the lexical level, including factors such as word length, number of new words, average number of words, etc., while the other aspect is the sentence level, including factors such as sentence length, average length of sentences, complex sentences, and so on. This study will perform statistical analysis on the readability of the text in the eight editions of widely used Junior High School English textbooks in China for these two aspects: the lexical level and the sentence level.

As for the study of the readability of English textbooks, scholar Gu (2000) further discussed the improvement of intensive reading teaching and the perfection of textbooks. However, the study must pay more attention to syntax and grammatical difficulty. Lu and Guo (2014) surveyed the readability of high school English textbooks. This study mainly focused on the relationship between readability and fun and needed more in-depth research on readability. Guo (2016) evaluated the readability of the reading materials in the 1 to 5 textbooks required for senior high school English in major editions. It conducted a comprehensive survey on the factors of text readability. However, this study focuses only on one type of teaching material, and further research is needed to gather data on other teaching materials. Miftaahurrahmi et al. (2017) investigated the readability of reading texts in English textbooks used by senior high school students in West Sumatra. She only uses one method to study the difficulty of the text and few attempts at other studies on readability. Parashchuk et al. (2021) state that applying automated tools for scaling text complexity, like the TextEvaluator, is indispensable for leveling reading passages in EFL classroom teaching and assessment practices. Baker's (2023) exploratory mixed-methods study conducted in an Asian university writing center setting examines the effects of text length on the readability of model essays. He found that text length had a significant positive association with the informants' ranking of ease and difficulty. It shows that the study of readability can help students in different ways.

To sum up, most studies on the difficulty of English textbooks focus on college and high school textbooks. However, Junior High School is an essential period of English learning, and the difficulty and span of Junior High School textbook texts play a crucial role in students' learning. Secondly, the research on readability factors is scattered, and there needs to be a systematic summary and analysis of readability indicators. Finally, researchers use existing corpora to study the difficulty of the text or use the corpus of textbooks with a small amount of data. However, the author spent three months building a corpus of the eight editions of Chinese Junior High School English textbooks, with 125,800 words, to conduct quantitative research on the difficulty of providing specific references for textbook users. Therefore, this paper adopts the method of self-built corpus to quantitatively compare the readability of Chinese Junior High School English textbook texts from the lexical and syntactic levels.

Method

The language readability of the existing English resources in China is the main focus of this study. Take a look around the "Electronic Textbook Network" (www.dzkbw.com). People can

obtain eight versions of Junior High School English textbooks in China through Baidu Wenku and other websites: People's Education Press (PEP), Foreign Language Teaching and Research Press (FLT & RP), Popular Science Press (PSP), Oxford University Press (China) Limited and Yilin Press (OUPL & YP), Shandong Education Press (SEP), Hebei Education Press (HEP), Beijing Normal University Press (BNUP), and Shanghai Education Press (SEP). For complete information on the textbook information, see Appendix A.

Construction of Corpus

A corpus consists of an extensive library of written or spoken text that requires computer storage and processing for language research (Renouf, 1987). After the research content was determined, the author extracted the content and built the corpus of Junior High School English teaching materials. The specific procedure is as follows:

Establish the Corpus

Eight variations are constant despite the number of names and variants discovered on various pages utilizing the website to search. The top eight English textbooks chosen for the comparison study are PEP (59.82%), FLF & RP (10.71%), PSP (9.82%), OUPL & YP (7.59%), SEP (4.46%), HEP (3.57%), BNUP (1.79%), and SEP (1.34%), according to the ranking of local textbooks in "The Use of Junior High School English Textbooks and Textbook Analysis," published on the Zhihu website.

Corpus Gathering

We extracted the vital information from each textbook version and compiled it into an Excel table. The information includes grade, unit name, reading material text, grammar, and extracurricular knowledge. We obtained the data from the eight textbooks mentioned above, versions found on Baidu Wenku, Baidu Network Disc, Disc Duo, Sina Microdisk, and other search engines.

Corpus Combing

The text corpus of each textbook version is separately extracted into a Word document and used to view the errors, carefully check the spelling errors, remove the problems present in the corpus, and rename the text and document. The process contains spelling errors, typesetting errors, random coding, and other defects. (For instance, the PEP of the textbook text, the RJB-Article document, the Art-RJB-7-1-unit1, and so forth.

Corpus Segmentation

First, the text corpus of the textbook is converted into TXT format to facilitate segmentation and statistics. In the process of transformation, people can sort each document (ranked by the use ranking of each region), such as one in 1-PEP-Article-all, PEP the pinyin initials, all indicates all reading text in the textbook, other versions, and the like. The division of the corpus into three parts: the first includes the grade and different grades; the second is according to the unit, with each unit building a document; and the third is according to the text, calculating the number of reading materials in each textbook edition.

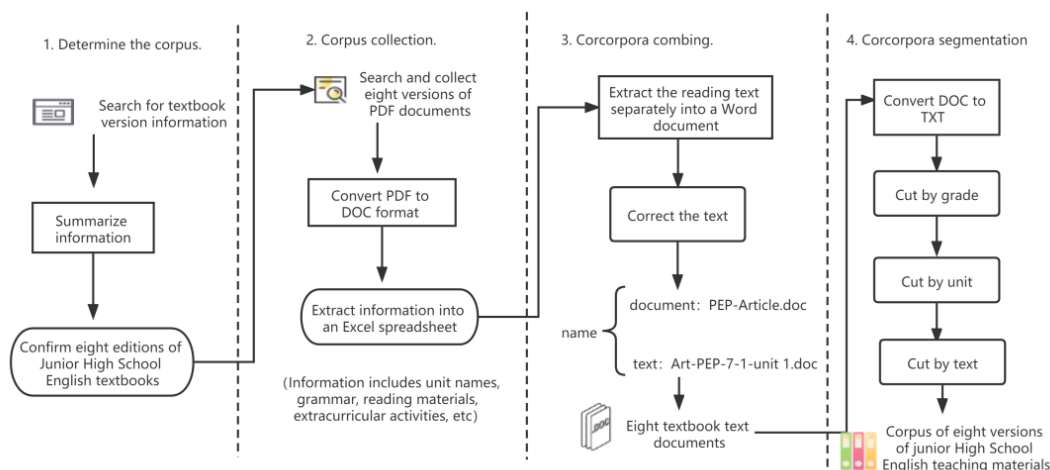


Figure 1. Corpus construction process of Junior High School English textbook

Determine the Factors Affecting Readability

Determine the readability of Junior High School English textbooks by considering the generalization of the readability factors affecting reading materials made by scholars like Dehant and Smith (1961), Fry (2002), Richards and Schmidt (2003), and referring to Li (2000), Guo(2016)and other scholars on the factors affecting the readability of the textbook text. This paper roughly divides these factors into two levels: (1) The vocabulary level, such as the word length, the rate of new words, and the average number of words, etc. (2) The sentence level, such as sentence length, average sentence length, complex sentence, and so on. Subsequently, a comprehensive analysis of the data results at the two levels was conducted, including statistics on the ease of reading the eight versions of Junior High School English textbooks widely used in China.

Results

Basic Information Statistics of Junior High School English Textbooks

Before calculating the readability of the text, we counted the primary material in each version of the textbook. The statistical content includes basic information such as grade, number of volumes, units, articles, and the total quantity of Junior High School English textbooks and other items. We ranked each textbook edition based on frequency and tallied the basic grade information, number of volumes, units, words, and textbooks. Here are the statistical findings:

Table 1. Statistics of basic information of Junior High School English Textbooks

| No. | Version | Grade | Number of books | Number of Units | Number of text | Total Number of words | Total number of sentences |
|-----|-----------|--------------------------|-----------------|-----------------|----------------|-----------------------|---------------------------|
| 1 | 《PEP》 | Grade 7th, 8th, and 9th | 5 | 55 | 79 | 18 109 | 1 618 |
| 2 | 《FLF&RP》 | Grade 7th, 8th, and 10th | 6 | 64 | 64 | 15 264 | 1 412 |
| 3 | 《PSP》 | Grade 7th, 8th, and 11th | 6 | 66 | 103 | 14 267 | 1 449 |
| 4 | 《OUPL&YP》 | Grade 7th, 8th, and 12th | 6 | 44 | 44 | 9 572 | 1 021 |

| | | | | | | | |
|---|--------|-------------------------------|---|----|-----|--------|-------|
| 5 | 《SEP》 | Grade 6th, 7th, 8th, and 13th | 7 | 61 | 85 | 19 683 | 1 732 |
| 6 | 《HEP》 | Grade 7th, 8th, and 14th | 6 | 42 | 134 | 25 552 | 3 184 |
| 7 | 《BNUP》 | Grade 7th, 8th, and 15th | 5 | 31 | 61 | 14 009 | 1 435 |
| 8 | 《SEP》 | Grade 7th, 8th, and 16th | 6 | 46 | 46 | 10 129 | 1 149 |

As pointed out in the above table, each grade is broken down into two books, making a total of six books, except for PEP and BNUP of the textbook in grade nine, which makes there only five books, and the SEP of 6th, 7th, 8th, and 9th in four grades. The other three grades are 7th, 8th, and 9th. The SEP is distinct from the others in that it has four grades in its Junior High School and seven books overall. Only the international version, OUPL and YP, SEP unit number, and articles, respectively, 64,44,46, the textbook each unit only a reading text, another has additional reading material. Each textbook edition has specific content.

In addition, the HEP of the text number, sentences, the number of articles is the most, 25,552 words, 3,184 sentences, 134 reading, much higher than different versions of textbooks, the number is twice higher than other versions, even nearly three times higher than the textbook, OUPL and YP only 9,572 words, 1,021 sentences, 44 articles, this phenomenon shows the HEP of the textbook text readability is the most significant? Moreover, the least difficult to translate is the forest textbook.

Comparison and Analysis of Text Readability at the Vocabulary Level

There are many text readability indicators at the vocabulary level, such as the average number of words per sentence, the number of uncommon words, the proportion of different words, personal pronouns, the number of syllables, the number of affixes, the number of new words, etc. According to the nature of the corpus, the total number, characters, average word length, new word number, and the rate of new words in the textbook were selected. The results are as follows:

Table 2. *Statistical table of text readability indicators at the vocabulary level*

| N o. | Version | Total Number of words | Number of characters | Average word length | Number of new words | Rate of new words |
|------|-----------|-----------------------|----------------------|---------------------|---------------------|-------------------|
| 1 | 《PEP》 | 18109 | 78332 | 4.17 | 1014 | 5.60% |
| 2 | 《FLF&RP》 | 15264 | 67098 | 4.35 | 946 | 6.20% |
| 3 | 《PSP》 | 14267 | 64780 | 4.43 | 1198 | 8.40% |
| 4 | 《OUPL&YP》 | 9572 | 42565 | 4.36 | 641 | 6.70% |
| 5 | 《SEP》 | 19683 | 86142 | 4.24 | 1161 | 5.90% |
| 6 | 《HEP》 | 25552 | 109891 | 4.14 | 1431 | 5.60% |
| 7 | 《BNUP》 | 14009 | 61292 | 4.22 | 841 | 6.00% |
| 8 | 《SEP》 | 10129 | 45145 | 4.31 | 729 | 7.20% |

The eight versions of the textbook can be categorized into three levels using the table above: Between 18,000 and 15,000, HEP, SEP, and PEP fall into various groups; PSP and BNUP are medium; between 14,000 and 15,000; and SEP, with between 9,000 and 10,000. The enormous number of articles, up to 134, compared to the OUPL and YP of 44, account for the majority of the difference in word count, together with the length of each edition.

The average word count is 4.14-4.43 in the eight textbook versions, meaning each has four to five words. Because some statements only contain two or three words in a sentence, the average readability of the eight textbooks is low. Although the HEP has many words, the average word

length is just about 4.1 words, suggesting that reading the HEP might not be too harsh even if there are many words. The average word length in the OUP and YP with the lowest word count is high, at 4.36, suggesting that it would be more challenging to read.

The rate of new words in the eight versions is 5-9%, with five to nine new words in 100. Laufer (1992) proposed that the rate of new words should not exceed 5% in the reading materials, but the overall rate of the eight versions exceeds 5%, and the rate of the PSP reaches 8.40%, followed by the highest version of the SEP, 7.20%. The high rate of the two versions shows that the reading readability is relatively higher than other versions of the textbook. Compared with the above two versions, the rate of new words of the PEP, HEP, and SEP is relatively low, with 5.60%, 5.60%, and 5.90%, indicating their reading readability is relatively low.

Upon comparing the above indicators, we found that obtaining the readability of the textbook from a single index is not possible, and the readability of the textbook also differs from a single index.

Comparison and Analysis of Text Readability at the Sentence Level

In addition to the major factors affecting the text's readability, it is important not to disregard sentences as they constitute significant aspects. Long sentences are more complicated structurally and feature more modifiers, clauses, and other words than short sentences. This essay provides readability statistics for the eight English textbook iterations. The following table provides an overview of the key indicators, which are the typical sentence length and the frequency of compound sentences:

Table 3. *Statistical table of text readability indicators at the sentence level*

| N o. | Version | Total Number of words | Total Number of sentences | Average sentence length | Number of complex sentences | Frequency of complex sentences |
|------|----------|-----------------------|---------------------------|-------------------------|-----------------------------|--------------------------------|
| 1 | 《PEP》 | 18 109 | 1 618 | 11.19 | 169 | 10.44% |
| 2 | 《FLF&RP》 | 15 264 | 1 412 | 10.81 | 127 | 8.99% |
| 3 | 《PSP》 | 14 267 | 1 449 | 9.85 | 105 | 7.25% |
| 4 | 《OUP&YP》 | 9 572 | 1 021 | 9.38 | 68 | 6.66% |
| 5 | 《SEP》 | 19 683 | 1 732 | 11.36 | 191 | 11.03% |
| 6 | 《HEP》 | 25 552 | 3 184 | 8.03 | 182 | 5.72% |
| 7 | 《BNUP》 | 14 009 | 1 435 | 9.76 | 109 | 7.60% |
| 8 | 《SEP》 | 10 129 | 1 149 | 8.82 | 53 | 4.61% |

The SEP of the textbook has the longest sentence, as seen from the table above, which shows that the average sentence length for the eight versions of the textbooks is between 8 and 12 words. The PEP comes in second with an average sentence length of 11.19 words, while the external research version has an average of 10.81 words. The average sentence length reaches 11.36 words. The writer discovered that some sentences' Rujiao version, PEP, FLF, and RP should be shorter when counting the corpus, such as the line that describes the festival: Chinese people have celebrated Mid-Autumn Festival and enjoyed mooncakes for millennia. (PEP 9th, unit 2); Many would concur that we most likely associate Christmas with giving, Christmas trees, and Santa Claus. Since the first English settlers landed in America via ship in the seventeenth century, we have celebrated the holiday (SEP 8th, unit 8; FLF&RP 9th, module 1). The three festival descriptions are lengthy, with an average length of 15 words, as seen by the samples above. Some people wear Santa Claus hats during Christmas (PSP 8th, unit 8); we will celebrate Mid-Autumn Festival this week (HEP 8th,

unit 3); compared, some textbooks versions feature shorter sentences that, on average, contain fewer words.

The textbook contains compound sentences between 4.61% and 11.3% of the eight versions of the complex sentences, with numbers ranging widely. The SEP had the most compound sentences and had the highest frequency, reaching 11.03%, compared to the SEP and HEP, whose respective frequencies were just 4.16% and 5.72%. The conclusions drawn at the vocabulary and sentence levels must be more consistent. Then the author will use the comprehensive analysis method to judge the differences in the difficulty of the textbook texts.

Comprehensive Analysis of the Influencing Factors of Text Readability

When comparing the readability factors of vocabulary and sentence-level texts in the Junior High English textbook, it is crucial to consider factors such as word count, average word length, rate of new words, average sentence length, sentence readability, and frequency of compound sentences. It is essential to approach this comparison from an overall perspective and conduct a thorough analysis to ensure everything is consistent. However, the properties of the influencing factors are different, not only in number and frequency, but the level between the indicators is very different. To enable a unified comparison, we standardized the raw data using Min-max normalization, which scaled the values to a specific interval [0, 1]. This normalization process facilitated the comparison and weighting of different units and properties. Readers can find the calculation results in Table Four. Next, we used the expert scoring method (Zhang, 2004) to determine the weight of each index. (0.05, word rate of 0.15, average word length of 0.25, total number of sentences of 0.05, average sentence length of 0.25, compound sentence frequency of 0.25), standardized data * weight = result, and finally, the data is comprehensively sorted (see Table five for structure).

Table 4. *Results of standardizing text readability indicators of each version*

| N o. | Version | Total Number of words | Average word length | Rate of new words | Total Number of sentences | Average sentence length | Frequency of complex sentences |
|---------|--------------------|-----------------------------|---------------------------|-------------------------|------------------------------------|-------------------------------|--------------------------------------|
| 1 | 《PEP》 《FLF&RP》 | 0.71 | 0.95 | 0.96 | 0.51 | 0.99 | 0.95 |
| 2 | 》 | 0.6 | 0.98 | 0.61 | 0.44 | 0.95 | 0.82 |
| 3 | 《PSP》 《OUPL&YP》 | 0.56 | 1 | 0.68 | 0.46 | 0.87 | 0.66 |
| 4 | 》 | 0.37 | 1 | 0.83 | 0.32 | 0.83 | 0.6 |
| 5 | 《SEP》 | 0.77 | 0.95 | 0.88 | 0.54 | 1 | 1 |
| 6 | 《HEP》 | 1 | 0.93 | 1 | 1 | 0.71 | 0.52 |
| 7 | 《BNUP》 | 0.55 | 0.95 | 0.91 | 0.45 | 0.86 | 0.69 |
| 8 | 《SEP》 | 0.4 | 0.98 | 0.97 | 0.36 | 0.78 | 0.42 |

Table 5. Ranking results of text readability indicators of each version (Sorted by comprehensive results)

| No. | Version | Total Number of words | Average word length | Rate of new words | Total Number of sentences | Average sentence length | Frequency of complex sentences | Comprehensive results | Rank |
|-----|------------|-----------------------|---------------------|-------------------|---------------------------|-------------------------|--------------------------------|-----------------------|------|
| 5 | 《SEP》 | 0.039 | 0.239 | 0.133 | 0.027 | 0.25 | 0.25 | 0.156 | 1 |
| 1 | 《PEP》 | 0.035 | 0.239 | 0.144 | 0.025 | 0.246 | 0.237 | 0.154 | 2 |
| 2 | 《FLF&RP》 | 0.03 | 0.244 | 0.092 | 0.022 | 0.238 | 0.204 | 0.138 | 3 |
| 7 | 《BNUP》 | 0.027 | 0.239 | 0.136 | 0.023 | 0.215 | 0.172 | 0.135 | 4 |
| 6 | 《HEP》 | 0.05 | 0.233 | 0.15 | 0.05 | 0.177 | 0.13 | 0.132 | 5 |
| 3 | 《PSP》 | 0.028 | 0.25 | 0.102 | 0.023 | 0.217 | 0.164 | 0.131 | 6 |
| 4 | 《OUPL&Y P》 | 0.019 | 0.25 | 0.125 | 0.016 | 0.206 | 0.151 | 0.128 | 7 |
| 8 | 《SEP》 | 0.02 | 0.244 | 0.145 | 0.018 | 0.194 | 0.104 | 0.121 | 8 |

According to Table five, after the index of the text readability, get the ranking results from difficult to easy: SEP, PEP, FLF, BNUP, HEP, BNUP, OUPL, YP, and SEP. The results and the individual text readability affect the arrangement of the writer of the text readability, namely, cannot simply determine a particular index as the basis of text readability, but a comprehensive analysis of the indicators, and multiple comparisons, to ensure the effectiveness of the conclusion.

Overall, among the eight versions of Chinese Junior High School English textbooks, the SEP is the most difficult, followed by the PEP, FLF, and RP, and the least difficult are the OUPL and YP, and SEP. Accordingly, teachers can choose different versions of the text according to the readability indicators. At the same time, the textbook writers can also adjust the readability of the textbook version according to the data to better practice the principle of taking measures by local conditions and their aptitude.

Discussion

The main findings from the chapter, as mentioned earlier, include the following: There are some differences in the lexical and syntactic levels of readability of reading texts in the eight editions textbooks. The range of factors affecting readability are as follows: total number of words (9572-25552), average word length (4.14-4.13), rate of new words (5.6%-8.4%), average sentence length (8.03-11.36), frequency of complex sentences (4.61%-11.63%). Among them, the average word length and new word rate are the highest in PSP, which reach 4.43 words and 8.40%, respectively. The SEP has the enormous average sentence length and frequency of compound sentences, at 11.36 words and 11.03%, respectively. After comprehensively comparing and standardizing various indicators that affect the readability of the text and using the expert scoring method to set the weight, we confirmed that the SEP has the most extensive difficulty.

Here, we contrast the findings of the earlier research. As reported by Guo (2016), the new words rate of high school English textbooks is 5.9% to 12%, and that of reading texts from Junior High School English textbooks in this study is 5.6% to 8.4%, indicating that the difficulty of some textbooks in this study has exceeded the difficulty of high school textbooks. The challenge of new

words is far greater than their capacity for pupils with a Junior High School level of education. Students may eventually lose interest in the English language learning process if they find it difficult to read the material. The average sentence length of this study's Junior High School textbooks was 8 to 11 words. Miftaahurrahmi's (2017) study of senior high school reading materials in West Sumatera discovered that the average sentence length was 19.5 words. It demonstrates that the average sentence length of these eight textbook editions is fair and that students may comprehend the reading content more easily.

While there is no consensus regarding major differences in the readability of textbook texts, the statistics in this research indicate significant differences in their difficulty. Some textbooks have an excessively high difficulty level, while others are deficient. According to Nuttal (1996) and Miftaahurrahmi et al. (2017), a good-written book has a lexical and structural difficulty that tests pupils without overwhelming them. So that students may comprehend the text and better master it, the difficulty of reading should be taken into account while organizing and choosing the text for textbook reading.

Due to the large number of textbook versions and the difficulty of the experiment, we decided not to investigate the opinions of the same group of students on different versions of textbooks and their experiences after using them. It will be easier to explain the differences in the readability of reading texts between textbook versions and to comprehend students' learning experiences of reading texts in different versions of textbooks. If the same group of students can study different versions of texts on the same topic and then comment on various versions, it will be more effective to select the most suitable reading texts of textbooks for students, which may be the best option.

The author aims to provide data resources for future research, focusing on studying text readability and facilitating teaching and the advancement of education. Norman (2011) introduced the concept of ubiquitous learning, which involves integrating mobile technology, web services, and context-awareness technology to provide virtual information based on the learner's situation or environment. Currently, researchers are exploring online and digital learning using network media (Norman et al., 2022) and MOOC learning (Nordin et al., 2015; Nordin et al., 2016). Paper textbooks will likely adapt to mobile technology, web services, and context-awareness technology, enabling students to access English learning materials anytime and anywhere.

Conclusion

Using the self-built corpus of Junior High School English textbooks and the corpus research method, this study compared the difficulty of reading texts in the eight editions of Junior High School English textbooks widely used in China. Based on the results of this study, there are evident differences in the readability of reading texts in the eight editions of Junior High School English textbooks in China. This finding highlights the importance of considering reading difficulty when arranging and selecting textbooks to enhance students' comprehension and mastery of the texts. Meanwhile, it can help teachers use textbooks flexibly and provide references for textbook writers in developing quantitative standards for the difficulty of textbook texts.

Upon completion of this study, the researcher will provide recommendations. To enhance students' reading experience with more comprehensible and engaging materials, textbook authors and test developers are advised to prioritize factors such as ensuring text readability, accurately determining the difficulty level, and aligning with students' proficiency level. To be able to choose textbooks that are appropriate for the level of difficulty of the school, the region, and the nation, it

is encouraged that the selectors and users of the textbooks consider the differences in the readability of the textbooks when making their selections. Finally, professionals carefully choose and create each textbook edition for students and teachers. They can also consider selecting additional textbooks of varying difficulty as ideal supplementary teaching resources and reading materials to complement their already familiar ones.

Acknowledgments:

This project was supported by [the Malaysian Research Universities Network \(MRUN\)](#) grant [number: MRUN-RAKAN RU-2019-003/2](#).

About the authors:

Wu Mengyao is a Ph.D. student in the Faculty of Education, Universiti Kebangsaan Malaysia (UKM), majoring in Learning Technology and Innovation. Her research focus on English online learning and corpus linguistic, including study of Automatic Question Generation, textbook analysis, and curriculum research and so on. ORCID: <https://orcid.org/0009-0005-2916-2008> Email: p119844@siswa.ukm.edu.my

Associate Professor. Ts. Dr. Helmi Norman is currently the Deputy Director of Instructional Technologies for the Center of Teaching and Curriculum Development, Universiti Kebangsaan Malaysia (UKM). He is also an Associate Professor at the Department of Teaching and Learning Innovations, Faculty of Education, [UKM](#). His [expertise](#) is [in the field of](#) digital and futuristic education. ORCID: <https://orcid.org/0000-0002-4717-3166>. Email: helmi.norman@ukm.edu.my

Dr. Nor Hafizah Adnan is a senior lecturer of Educational Technology at the Faculty of Education, Universiti Kebangsaan Malaysia (UKM). Her areas of concentration are design and development of technology-enhanced learning environments, augmented reality in education, and teaching practices in technology-oriented curriculum. ORCID:<https://orcid.org/00000001-9368-7646>. Email: norhafizah@ukm.edu.my

References

- Adnan, N. H., & Sayadi, S. S. (2021). ESL Students' Readiness for Self-Directed Learning in Improving English Writing Skills. *Arab World English Journal*, 12(4), 503-520. <https://doi.org/10.24093/awej/vol12no4.33>
- Baker, J. R. (2023). The Effects of Text Length on the Readability of Model Essays. *GEMA Online Journal of Language Studies*, 23(1), 60-73. <http://doi.org/10.17576/gema-2023-2301-04>.
- Betts, E. A. (1949). Readability: Its application to the elementary school. *Journal of Educational Research*, 42(3), 202-211.
- Chen, A. (2010). *Comparison of Oxford and Human teaching high school English textbooks*, (Unpublished Doctoral dissertation). Shanghai Normal University, China.
- Cheng, X. (2002). *English textbook analysis and design*. Foreign Language Teaching and Research Press.
- Crossley, S. et al. (2022). A large-scaled corpus for assessing text readability. *Behavior Research Methods*, 55(2), 491-507. <https://doi.org/10.3758/s13428-022-01802-x>

- Dechant, E. V., & Smith, H. P. (1961). Psychology in teaching reading. *Prentice-Hall, Inc.*
- Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, 32(3), 221-233.
- Fry, E. (2002). Readability versus leveling. *Reading Teacher*, 56(3), 286-291.
- Gu, Q. (2000). On the accessibility of college English intensive textbooks. *Journal of Shaoxing University of Arts and Sciences (Philosophy and Social Sciences Edition)*, (2), 89-91.
- Guo, L. (2016). *Analysis of major school English*, (Unpublished Doctoral dissertation). Chongqing Normal University, China.
- Klare, G. R. (1963). *The measurement of readability*. Iowa: Iowa State University Press, 17.
- Laufer, B. (1992). Vocabulary and applied linguistics. *London Journal*, 20(2), 134-136.
- Lin Z. (1995). The Measurement of English Readability. *Foreign Language Teaching and Research*, 12(04), 38-42.
- Li Shao-shan. (2000). Overview of readability research. *Journal of PLA Foreign Languages Institute*, 10(04), 2-5.
- Lu, S., & Guo, S. (2014). Evaluation of English textbooks. *Journal of Shenyang Normal University Social Sciences Edition*, 38(2), 147-149.
- McLaughlin, G. H. (1969). SMOG grading - a new readability formula. *Journal of reading*, 23(22), 639-646.
- Miftaahurrahmi, F., & Syarif, H. (2017). The Readability of Reading Texts in English Textbook Used by Senior High School Students in West Sumatera. In Y. Rozimela, et al. (Eds.), *Proceedings of the Fifth International Seminar on English Language and Teaching* (Vol. 110, pp.199-203). Atlantis Press. Available at <https://www.webofscience.com/wos/woscc/summary/97f49a3a-acdb-4a1d-a801-f804b56ffebe-8a1f6d63/relevance/1>
- Nordin, N., Norman, H., & Embi, M. A. (2015). Technology acceptance of massive open online courses in Malaysia. *Malaysian Journal of Distance Education*, 17(2). <https://doi.org/10.21315/mjde2015.17.2.1>
- Nordin, N., Norman, H., Embi, M. A., Mansor, A. Z., & Idris, F. (2016). Factors for Development of Learning Content and Task for MOOCs in an Asian Context. *International Education Studies*, 9(5), 48. <https://doi.org/10.5539/ies.v9n5p48>
- Norman, H., Adnan, N. H., Nordin, N., Ally, M., & Tsinakos, A. (2022). The Educational Digital Divide for Vulnerable Students in the Pandemic: Towards the New Agenda 2030. *Sustainability*, 14(16), 10332. <https://doi.org/10.3390/su141610332>
- Norman, H., Din, R., & Nordin, N. (2011). A preliminary study of an authentic ubiquitous learning environment for higher education. *learning*, 3(4), 89-94.
- Nuttal. (1996). *The use of authentic materials in the teaching of reading*. Available at https://scholar.google.co.id/scholar?q=Nuttall,+1996+onReadability&hl=id&as_sdt=0&as_vis=1&oi=scholart&sa=X&ve=0ahUKEwiwgqmNnaHOAhUBMo8KHQHhB8oQgQMIIDAA
- Parashchuk, V., Yarova, L., & Parashchuk, S. (2021). Automated Complexity Assessment of English Informational Texts for EFL Pre-service Teachers and Translators. *Arab World English Journal (AWEJ) Special Issue on CALL*, 7, 155-164.
- Renouf, A. (1987). Looking Up: An Account of the COBUILD Project. *Harper Collins*, 117.
- Richards, J. C., & Schmidt, R. (2003). *Longman Dictionary of language teaching & applied linguistics*. Pearson Education.
- Song, L. (2010). *Comparative study of Human Junior High School English textbook JEFc and Go for it!* (Unpublished Doctoral dissertation). Chongqing University, China.

Srisunakrua, T., & Chumworatayee, T. (2019). Readability of Reading Passages in English Textbooks and the Thai National Education English Test: A Comparative Study. *Arab World English Journal*, 10(2), 257-269. <https://doi.org/10.24093/awej/vol10no2.20>

Thorndike, E.(1921). *A Teacher’s Word Book of 10,000 Words*. New York: Cambridge University Press.

Zhang, P. (2004). *Comprehensive evaluation study based on principal component analysis*, (Unpublished Doctoral dissertation). Nanjing University of Technology, China.

Appendices

Appendix A

The publication information of eight versions of Junior High School English textbooks

| | | | | | | | | |
|----------------|--|--|-----------------------------------|--|--------------------------------|--|---|---|
| Textbook Title | Go For It | New Standard English | Project English | Fun with English | English | Compulsory Education Coursebook English | English | English |
| Approval Time | 2012 | 2013 | 2012 | 2012 | 2013 | 2012 | 2013 | 2013 |
| Publisher | People's Education Press (PEP) | Foreign Language Teaching and Research Press (FLTRP) | Popular Science Press (PSP) | Oxford University Press (China) Limited and Yilin Press (OUP&YP) | Shandong Education Press (SEP) | Hebei Education Press (HEP) | Beijing Normal University Press (BNU P) | Shanghai Education Press (SEP) |
| Write Unit | Research and Development Center of English Curriculum Materials, Institute of Curriculum Materials, People's Education Press (Ch | Curriculum and Teaching Materials Research Institute | Beijing Renai Education Institute | English Textbook Compilation Committee of Oxford University Press (China) Co., LTD | | Hebei Education Press in cooperation with DC Canada International Communicative Center | Beijing Education Research Institute | English Textbook Compilation Committee of Oxford University Press (China) Co., LTD. Shanghai Century Publishing |

| | | | | | | | | |
|--------------------------------------|---|--|--|---|-----------------------------------|---|--|---|
| | ina) Cen gage Lear ning Gro up(USA) | | | | | | | g Co., L TD |
| Editor | Liu Daoy i Zheng Wangqua n David Nunan(U SA) | Cheng Li Sim on Gre enall (UK) | Jim Gr eenIaw (Canad a) Wa ng Dec hong | Catheri ne Daw son(U K)Wan g Shoure n He F eng | | Tian Gui sen Jim Pars ons (Canad a) | Beijing Acade my of Educati on Scie nces | Zhang C hunbai S hu Yunx iang |
| Edition | 1st editio n, October 2 012 | 1st edi tion, J uly 2013 | 1st edit ion, Oc tober 2 012 | 1st editi on, No vember 2012 | 1st edi tion, J uly 2013 | 1st editi on, Octo ber 2012 | 1st edit ion, Jul y 2013 | 1st editi on, July 2013 |
| Main a reas of use (C hina) | National | Liaoni ng, In ner Mong olia, S handon g, Ti anjin, Zhejia ng, Gu angdo ng, an d Gua ngxi. | Anhui, Fujian, Haina n, Guiz hou, Y unnan, Chong qing, the No Heilon gjiang, Guang dong. | Jiangs u, Sich uan, Gu angdon g, Anhui, Hunan | Shand ong | Hebei, G ansu, Sh aanxi, an d Fujian | Sichua n, Chen gdu, Sh aanxi, Guangd ong, Be ijing | Guangzh ou, Shen zhen, an d Sheny ang |