



ORIGINAL ARTICLE

Vocabulary Coverage in a High School Vietnamese EFL Textbook: A Corpus-based Preliminary Investigation

Nguyet Thi Minh Le⁺,
Huong Thi Dinh

Hanoi National University of Education, Vietnam

⁺Corresponding author • Email: nguyetlm@hnue.edu.vn

Article history

Received: April 22, 2022

Accepted: May 24, 2022

Published: June 8, 2022

Keywords

English textbook, vocabulary coverage, high-frequency words, word families

ABSTRACT

In the context of English as a foreign language (EFL) teaching, textbooks play a crucial role in the classroom as the primary source of lexical input for learners. Thus, it is essential that textbooks can sufficiently facilitate learners in their overall comprehension and vocabulary learning, especially with the most frequent words in English. The study examined the vocabulary load of a locally-published EFL textbook in Vietnam for high school students focusing on the vocabulary coverage in relation to the 95% and 98% thresholds for text comprehension and the coverage of high-frequency word families in the textbook. To fulfil these objectives, a frequency-based analysis of word families using the Vocabprofilers on the <https://www.lexutor.ca/> website was conducted with the 41,137-word textbook corpus. It is revealed that students need to have a vocabulary size of 3,000 and 5,000-word families to reach the coverage of 95% and 98% of the whole textbook, which would be too challenging for Vietnamese high school students. Moreover, the textbook was found not to represent the second 1,000-word list very well with the appearance of just over half of the second most frequent 1,000-word families. These findings highlight the role of teachers in adapting the textbook for more effective vocabulary learning and general comprehension of the materials.

1. INTRODUCTION

In the setting of teaching English as a foreign language (EFL), there is a broad consensus on the central role of the textbook as a “major source of input in English class”. (Sun & Dang, 2020, Dang & Webb, 2020; Alsaif & Milton, 2012). Textbooks provide opportunities for learners to gain language exposure through reading written texts, which also holds true regarding spoken English through listening activities of the textbook. Research shows that in many EFL teaching cases, lack of vocabulary exposure in learning resources, especially textbooks - the fundamental resources for lexical input in the EFL classroom, is one of the main causes of low level of vocabulary knowledge in English (Rahmat & Coxhead, 2021; Nguyen, 2020). To maximize learning quality, the vocabulary in EFL textbooks should be thoroughly chosen to ensure learners’ understanding of the textbook content and at the same time their efficient acquisition of the most useful words for them. Investigating corpus-based information in relation to the book user’s lexical knowledge would shed light on the vocabulary load of the textbooks for their users (Sun & Dang, 2020).

In Vietnam, with the introduction and implementation of the new General Education Program 2018, a list of English textbooks is currently available for selection by English teachers and local educational leaders, which is a significant change from the premise of textbook prescription by the Ministry of Education and Training in the past (Vietnam’s MoET, 2020). Instead of sharing the same textbook which was published by the Government, now schools and teachers have a greater role and options in choosing the textbook published by both eligible public and

private publishers. However, the greater power also accompanies greater responsibilities and requires more profound knowledge from teachers, and school leaders to make a sensible decision on textbook selection. Textbook evaluation would serve as a helpful tool for teachers in this process. An aspect of textbook evaluation focuses on the vocabulary load and coverage of the book as to whether book users can comprehend it properly without struggling with novel words, and whether the books sufficiently expose learners to the most frequently used words of the English language.

Yet, there have been very few studies in Vietnam concerning this matter, especially when it comes to EFL textbooks for school students. Therefore, this study aims to examine an English textbook for high school students to (1) determine the vocabulary load students need to understand the textbook in question; (2) to examine the coverage of high-frequency words in the textbooks as it is essential for EFL learners to have a solid knowledge of high-frequency words before learning words at lower frequency levels (Nation, 2013). The research aims to provide a useful reference for teachers and educational leaders in their quest for the most appropriate textbook for their students as well as propose some recommendations for English teachers in Vietnam in particular and other EFL contexts for more effective textbook exploitation.

2. LITERATURE REVIEW

2.1. Background of EFL teaching in Vietnam

Before being treated as the number one foreign language in Vietnam, though introduced into the National education system for long, not until the overall economic reform in 1986 (aka *Đổi mới*) did the English language start to gain popularity and be pursued by more language learners. According to Hoang (2018), English was taught in the light of the ‘structural method’ which focused on language accuracy through introducing and transforming different sentence patterns, translating and addressing reading comprehension questions. Oral skills, if introduced, were limited to the development of new sentences without context, and hence little attention was paid to the aspect of language fluency.

In the context of renovation (*Đổi mới*) initiated at the Six National Congress, the increasing popularity of the language started to gain recognition, marked by a number of milestones in curriculum and textbook designs as well as in the implementation of foreign language projects.

The international integration entailed an urgent requirement for the Vietnamese in the new era to be able to communicate effectively in an international context where English was becoming a global language. This situation fostered the launch of the national project entitled ‘Teaching and Learning Foreign Languages in the National Education System, Period 2008-2020’, which was detailed in Decision No. 1400/QĐ-TTg signed by the Prime Minister of the Government of the Socialist Republic of Vietnam in September 2008 (officially referred to as the 2020 Project).

It is highlighted in the project that by 2020, most Vietnamese young people graduating from secondary vocational schools or higher levels will be able to communicate confidently in their daily life as well as in an international environment where a foreign language is primarily used (Prime Minister of SRV, 2008). In order to achieve this long-term goal, three pilot English curricula were designed and assigned to the three successive levels of general education in the country, namely: (1) Pilot English Curriculum for Vietnamese Primary Education Schools; (2) Pilot English Curriculum for Vietnamese Lower Secondary Schools; and (3) Pilot English Curriculum for Vietnamese Upper Secondary Schools. As an integral part of the general education from Grade 3, these curricula are all aimed at learners to reach a certain CEFR (Common European Framework of Reference for Languages) level starting from A1 for elementary school leavers to B1 for high school graduates. Accordingly, all the officially approved textbooks have to cover all the four communicative language skills of listening, speaking, reading, and writing (MOET, 2010, 2012a, 2012b).

In order to achieve the general goal of the 2020 Project, together with the specific objectives of the three curricula, the development of the ten-year English textbook series under the collaboration of Vietnam Education Publishing House with MacMillan Education and Pearson Education was initiated in 2010 before the products were piloted at selected schools all over the country in 2012. Although “there occurred arguments and different, contradicting, and even conflicting ideas” (Hoang, 2015) during this process, it is undeniable that this series has proved to fulfil an urgent need for comprehensive teaching and learning resource for the three pilot curricula.

2.2. Vocabulary knowledge in EFL learning

2.2.1. Lexical coverage and thresholds

As defined by Laufer and Ravenhorst-Kalovski (2010), lexical coverage or text coverage or vocabulary coverage is the percentage of words in the text that are known by readers. In this way, if the lexical coverage of a text is 70%, it means that the readers understand only 70% of the words in the text, and an increase in the percentage would entail better text comprehension.

In order to determine how far a student is able to understand a certain language input, there should be a combination of two data sources: text coverage and lexical thresholds. Initially developed by Laufer (1989) and then further defined by other researchers, the lexical threshold is generally known as a milestone where the knowledge of lexical items can sufficiently cater for students' comprehension of the language input, "the minimal vocabulary that is necessary for adequate reading comprehension, the boundary between having and not having knowledge" (Laufer, 2010, p. 16). The minimal lexical coverage (lexical threshold) was, as first suggested by Laufer (1989), 95% for readers to cultivate a sufficient understanding of the text (Nguyen, 2020). For example, in a text consisting of 100 words, the reader needs to know at least 95 words to grasp the text's message. In later research, Hu and Nation (2000) and Nation (2001) further suggested that this figure should be extended to 98% to ensure better comprehension and incidental vocabulary learning without much external support. This seemingly modest 3% gap actually consists of a considerable number of word families, and no matter what percentage should be optimal, it is undeniable that there is a sharp link between lexical coverage and reading comprehension (Rahmat & Coxhead, 2021).

2.2.2. Word categories

The idea of word categories by Nation (2006) is essential to determine the necessary amount of vocabulary to be learned and the specific words that both ESL and EFL learners see in a corpus. Accordingly, words can be divided into three categories that are based on how often they are employed in both written and spoken contexts. The first category is high-frequency words which are the most frequent and dominant in any text and are composed of the first and second 1,000-word families.

It is confirmed that those most frequent 2,000 words may account for roughly 80% of the vocabulary load in various written texts (Alsaif & Milton, 2012; Rahmat & Coxhead, 2020), which consolidates Nation's (2013) argument about the importance of introducing the first 2,000 words to EFL learners. Teachers and learners are encouraged to do everything possible to make sure those words are learned (Alsaif & Milton, 2012). One of the most possible ways to do this is for teachers to include as many words from this category as early as possible to optimize learners' comprehension through the learning materials, specifically coursebooks. Another important feature that should also be well considered is the extent to which these words are repeated. According to Webb & Nation (2017) as cited by Sun & Dang (2020), the more frequently a word is recycled, the more likely it is learned. In other words, repetition is the key to incidental vocabulary learning and it is claimed by Nguyen (2020) that vocabulary learning can only be fostered if a word is encountered at least 6 times. In Sun & Dang's (2020) review, it is pointed out that EFL learners are unlikely to learn a new word which is encountered once only, while 7 or more occurrences may be needed for deliberate learning, 10 or more for incidental learning from reading, 15 or more for incidental learning from listening and 20 or more for significant influence on incidental vocabulary learning. (Webb & Nation, 2017; Pellicer-Sanchez & Schmitt, 2010; Webb, 2007; van Zeeland & Schmitt, 2013; Uchihara, Webb, and Yanagisawa, 2019 cited by Sun & Dang, 2020).

Mid-frequency words, despite less frequent occurrences, are "frequent enough to be a sensible learning goal after the high-frequency words" (Nation, 2013, p. 25) and cover the next seven thousand word families (from the third to the ninth).

Finally, low-frequency words are the least encountered in any texts by learners and include the 10th 1,000 and beyond. A summary of the three listed word categories can be seen in Table 1.

Table 1. Three-word categories (adapted from Nation, 2013)

Level	1,000-word family lists	Example
High frequency	1,000-2,000	<i>Agree, children, better...</i> Function words (<i>some, at, in...</i>)
Mid frequency	3,000-9,000	<i>earthquake, comprehend...</i>
Low frequency	10,000 on	technical vocabulary in specific subject areas (<i>ecotourism, transitive, illocutionary...</i>)

2.2.3. Previous studies on vocabulary in EFL textbooks

There have been a number of studies focusing on examining EFL textbooks. Using the Range Programme (Heatley et.al. 2002), Rahmat & Cobhead (2021) examined the three EFL textbooks officially used for 10-12 graders in Indonesian senior-high schools, concerning three major aspects: the number of words required to reach significant vocabulary coverage milestones, namely 95% and 98%; the distribution of words according to their frequency, most importantly the high-frequency words in the textbooks in question; and finally, the repetition of words in textbooks. The findings revealed the fact that there is a considerable gap between the required and the student's actual knowledge of vocabulary (5,000-6,000 and approximately 2,000, respectively) to achieve 98% lexical coverage. This big difference suggests that Indonesian students may not know a large number of words, which could hinder their comprehension of the texts. However, it is also noticeable that this series contains a large amount of high-frequency words (roughly 80%), many of which are repeated several times and therefore more likely to be learnt and remembered by students.

Similar studies on high school textbooks but in other EFL contexts were conducted by Alsaif & Milton (2012) in Saudi Arabia and Sun & Dang (2020) in China. The findings of these research papers were generally consistent with Rahmat & Cobhead (2021)'s and therefore propose critical views on the nature of EFL textbooks. Alsaif & Milton (2012) examined an outstanding account of 22 English textbooks mandatorily used from grade six to grade twelve. The research was completed employing Nation's software Range and the program *text lex compare* available at http://www.lex tutor.ca/text_lex_compare. According to the findings, only 2,800 words out of the most frequent 5,000, together with another 1,000 fewer frequency words were calculated from the textbooks examined. In terms of how many target MOET (Ministry of Education and Training) words are presented in the textbooks, the collected data showed that 55 words, which accounts for 2.75%, are included in the MOET's list but not found in the textbooks at all. These modest statistics are claimed to be "a shortcoming of the current school textbooks" because of the inadequate coverage for an optimal understanding.

Choosing the context of senior high schools in China, Sun & Dang (2020) investigated a corpus of nearly 300,000 words from both the written texts and the spoken texts transcribed from listening activities in 11 EFL textbooks. Although the examined textbooks offer a favourable condition for learners to adopt the most frequent 1,000 words, the 165 student participants are expected to encounter an annual number of 2,000-3,000 new word families which are far more than 400-word families as suggested by Webb & Chang (2012). Similarly, the oral texts were much more challenging than expected and "required a larger vocabulary size to reach the 98% coverage than the written texts."

Among the very few studies related to lexical coverage of Vietnamese high school students, Nguyen (2020) stood out with his examination of the high school English language textbooks published by the Vietnam Education Publishing House. With 422 high school Vietnamese students recruited from across the country and all three grades (Grade 10,11 and 12) of the high school level took the updated Vocabulary Levels Test (VLT) developed by Webb, Sasap and Ballance (2007), he found out that the vocabulary knowledge of Vietnamese high school students were the first two 1,000 word lists. More importantly, Nguyen (2020) mainly focused on the reading passages of the textbook series, rather than the textbook as a whole. Those textbooks were published and prescribed for high schools in Vietnam by the State for a period of about 10 years until 2020 (Hoang, 2015). This study concluded that the reading texts in the series of EFL textbooks for high school students in Vietnam were insufficient in fostering content and vocabulary gain due to the overload of novel words.

Another vocabulary study on textbooks by Dang (2017) examined the textbook "Tieng Anh 12" focusing on three specific aspects: word knowledge, vocabulary learning strategies and types of vocabulary activities. The Range program by Nation (2002) was used in addition to survey questionnaires and interviews with 180 twelfth graders and 5 English teachers. It was concluded from the findings that the lexical input in the textbook is not well selected and inadequate in terms of number and desirable aspects of knowing a word. Also, vocabulary activities are not diverse and explicitly designed, which may demotivate students and pose a challenge for teachers in exploiting the available activities.

Although the focus is on tertiary rather than high school level, the study by Cao (2018) is well worth mentioning as she also discussed the lexical aspect of an EFL textbook entitled "Life A2-B1". Adopting the lexical program VocabProfile (Cobb, 2009), the author indicates that learners have little "exposure to the productive use of vocabulary" to reach the target CEFR level of B1. Therefore, additional activities to provide more chances for learners to adopt new vocabulary as well as to promote their learning autonomy are an urgent and essential part of teachers' lesson planning process.

Apart from these studies, no published research papers have focused on the evaluation of the new ten-year series of English textbooks for high school level, specifically on the dimensions of lexical input in both written and spoken texts in the “Tieng Anh 10”. Therefore, this study was conducted to aim at filling this paucity.

3. MATERIALS AND METHODS

3.1. Research questions

The research aims to evaluate the vocabulary load of the English textbook Tieng Anh 10 for grade 10 students by determining the number of word families that students need to understand 95% and 98% of the vocabulary load of the textbook and the coverage of high-frequency words in the whole textbook, including both spoken and written texts. The researchers aim to point out if the textbook could facilitate learners’ comprehension and vocabulary acquisition of the most useful English words.

Specifically, there are two research questions to be addressed in this study:

1. How many word families do students need to reach 95% and 98% coverage of the Tieng Anh 10 textbook?
2. To what extent are high-frequency words covered in the textbook?

3.2. The textbook

As mentioned earlier, the new Tieng Anh 10 is part of the 10-year series which was first piloted at selected high schools in Hanoi and Ho Chi Minh City in 2010. Table 2 illustrates the general outline of the book being investigated.

Table 2. Book structure, unit structure, component headings and time allocated for each heading
(adapted from Hoang, 2015)

The new “Tieng Anh 10”	
Number of units	10
Number of components per unit	8
Component heading	Getting Started Vocabulary, Pronunciation, Grammar Reading Speaking Listening Writing Communication & Culture Looking Back & Project
Time allocated for each heading	1 period

As can be seen from Table 2, Tieng Anh 10 textbook consists of ten units with five units per volume. Each unit consists of eight component headings and is allocated with 8 equivalent periods. *Getting Started* aims for learners to get an overview of the whole unit while *Language* focuses on the three target language components: Vocabulary, Pronunciation, and Grammar. This heading is followed by the four periods addressing the four essential language skills (Reading, Speaking, Listening, and Writing). *Communication and Culture* help students enrich specific socio-cultural knowledge and then further apply this knowledge to communicate in daily conversations. Finally, *Looking Back and Project* is introduced within one class hour, with *Looking Back* providing an opportunity for revision of the whole unit and *Project* requiring students’ hard work and collaboration to participate in a real-life project closely related to the topic of the unit. Apart from the ten units, there are four Review lessons during which, as the title suggests, students revise and use the knowledge and skills they have acquired through two or three successive units in different tasks.

3.3. Corpus

A textbook corpus of 41,137 words was developed from the soft copy of the selected textbooks Tieng Anh 10 Volume 1 and Volume 2, published by the Vietnam Education Publishing House. In the scope of this study, both the written texts in the students’ books and transcripts retrieved from the audio scripts of the listening activities in the teacher’s books were involved in the corpus development. It is noteworthy that only the listening audio scripts from

the teachers' books were collected for the purpose of the study, not the whole corpus of the teachers' books. Thus, the textbook corpus comprises two sub-corpora of the written text and listening transcripts, which consist of 38,113 and 3,024 words respectively.

This corpus-based study employed the method of document analysis, which is a 'systematic procedure for reviewing or evaluating documents – both printed and electronics (computer-based and Internet-transmitted) material', together with the online lexical analysis tools to obtain the desired answers to the two abovementioned research questions. According to Bowen (2009), this analytical method involves the examination and interpretation of data with the aim of eliciting meaning, archiving new insights, and developing empirical knowledge.

3.4. Research instrument

The textbook materials (students' book and teachers' guide) were processed in a frequency-based analysis of word families using the BNC-COCA 1-25K program in the Vocabprofilers on the <https://www.lexutor.ca/> website. Vocabprofilers is "a simpler, web-based alternative to the RANGE program (Nation & Heatley, 2002). RANGE program can be found on Nation's website <https://www.wgtn.ac.nz/lals/about/staff/paul-nation#vocab-programs>.

3.5. Procedure for data collection

The electronic copies of Tieng Anh 10 students' book and teachers' guide were processed and examined to produce a 'ready-to-analyse' corpora. The text processing procedure was adapted from Rahmat & Coxhead (2021) with the following steps:

- The materials were downloaded in PDF format and converted into text files (.txt). Any parts which were not convertible were typed manually.

- Every page was thoroughly inspected for spelling mistakes or technical errors while being converted. Any mistake identified was corrected in close reference to the original texts, for instance, *columns, kinds, environment, Itnass, environment, unplugging, and theatres*. The unrecognized text and symbols were eliminated, including phonetics symbols, special icons or abbreviations which were explained explicitly in the textbook.

- Regarding hyphenated multi-word items, 'depending on the overlap in meaning between the individual words and the multi-word item' (Webb, 2008), the researchers either replaced the hyphen with space to break them into single lexical items or eliminated the hyphen, which would categorized the multi-item words into the Off-List group for later analysis. Some examples of hyphenated words to be separated included *child-minder, large-sized, rain-making, environmentally-friendly, etc.*

- As for proper nouns, in this study, proper nouns were identified and classified into the Off-lists group together with the compounds.

- The input texts were fed into the Vocabprofilers on the website [lexutor.ca](https://www.lexutor.ca/) for piloting analysis multiple times to identify the spelling or technical mistakes. The corpora then were confirmed and ready to analyse.

3.6. Procedure for data analysis

The corpora were analysed using the BNC-COCA 1-25K program in the Vocabprofilers on the <https://www.lexutor.ca/> website. In this program, each word in a text counts as one token and each different word in a text counts as one type (Webb, 2008); a word family (*employ*) includes a base form (*employ*) and its inflexions (*employed, employs, employing*) and closely related derivations (*employee, employees, employer, employers, employment, employable, unemployable, employability, unemployed, unemployment*) (Sun & Dang, 2020).

The procedure to analyse the vocabulary in the corpora was adapted from Rahmat & Coxhead (2021) and Sun & Dang (2020) as follows:

To answer the first research question on the number of word families learners need to cover 95% and 98% of the textbook, the researchers drew on the result table from Vocabprofilers including the number of word tokens, types and families of all the words in each of the different 1,000-word lists. The two sub-corpora were also analysed independently for better comparison between the written and spoken language used in the textbook.

To address the questions of how much the textbook covers the high-frequency words (the 1st and 2nd 1,000-word lists), the researchers also looked at the frequency-based word list which would show how many times each word family is mentioned in the texts by percentage. The number of tokens of each word family was counted and categorised into 5 levels: 0-1 occurrences, 2-6 occurrences, while 7 or more occurrences, 10 or more occurrences, 15

or more occurrences, and 20 or more occurrences. There has been no strong consensus on the minimum number of occurrences for vocabulary learning in the recent literature; however, according to Sun & Dang's review (2020), the findings of previous studies revealed that EFL learners are unlikely to learn new words which are encountered once only, 7 or more occurrences may be needed for deliberate learning, 10 or more for incidental learning from reading, 15 or more for incidental learning from listening and 20 or more for significant influence on incidental vocabulary learning. (Webb & Nation, 2017; Pellicer-Sanchez & Schmitt, 2010; Webb, 2007; van Zeeland & Schmitt, 2013; Uchihara, Webb, and Yanagisawa, 2019 cited by Sun & Dang, 2020).

4. RESULTS AND DISCUSSION

4.1. Vocabulary coverage in *Tieng Anh 10*

The first research question of the study focuses on the number of word-families textbook users need to reach the two lexical thresholds of 95% and 98% of the vocabulary load in the textbook *Tieng Anh 10*. The answer to this question can be found in Table 3 which illustrates the cumulative coverages of the whole textbook corpus, together with the analysis results of the two sub-corpora of the spoken text (developed from listening scripts) and the written text, with and without the Off-list words.

It can be seen that without the Off-list words, to understand 95% of the whole corpus, Vietnamese 10th graders needed to have a receptive knowledge of 6,000-word families in the English language, which was also true for spoken text (which was retrieved from listening and some pronunciation activities in the textbook) and written text.

However, when combined with words from the Off-list groups, 95% coverage of the whole textbook corpus, as well as the spoken and written text, could be reached with 3,000-word families. The proportion of the Off-lists for all three categories (whole corpus, spoken text and written text) was roughly equal at around 4%.

Table 3. Lexical coverage of the whole textbook corpus, spoken text and written text

Freq. Level	Whole corpus		Spoken text		Written text	
	Without off-list words	With off-list words	Without off-list words	With off-list words	Without off-list words	With off-list words
K-1 :	79	82.95	80	84.3	78.9	82.8
K-2 :	88	91.95	89	93.3	87.9	91.8
K-3 :	92.5	96.45	92.1	96.4	92.5	96.4
K-4 :	93.6	97.55	93.7	98	93.5	97.4
K-5 :	94.6	98.55	94.6	98.9	94.5	98.4
K-6 :	95.1	99.05	94.9	99.2	95	98.9
K-7 :	95.3	99.25	95.1	99.4	95.2	99.1
K-8 :	95.5	99.45	95.5	99.8	95.4	99.3
K-9 :	95.6	99.55			95.5	99.4
K-10 :	95.7	99.65	95.6	99.9	95.6	99.5
K-11 :	95.8	99.75			95.7	99.6
K-16 :	95.9	99.85	95.7	100	95.8	99.7

Note: The frequency-based 1,000 word lists with no word families mentioned in the text are excluded from the tables

Given the significant gap between the lexical coverage with and without the Off-lists, it seems necessary to investigate more closely into the words in the Off-lists. The words in the Off-lists of the whole corpus mainly belong to three sub-groups: (1) proper nouns; (2) compounds; and (3) abbreviations. The mentioned proper nouns includes both Vietnamese (e.g. *Anh, Bac, Viet Nam, Binh, Lao Cai, Sa Pa, Can Gio, Cuc Phuong, Giang, Nam, Mai, Vinh, Trinh, Truong, Thanh...*) and foreign-originated words (*Australians, Indonesians, African, British, Beijing, China,*

Chinese, Chopin, Collins, Kevin, Francois, Korea, Korean, Pennsylvania, Queensland, Tchaikovsky, etc.). As stated by Webb (2008), many researchers have taken the approach that proper nouns have a minimal learning burden and may be easily understood by readers. This statement should also hold true in this study as almost all the proper nouns mentioned were familiar and easily recognisable for Vietnamese students. Also, when compounds were considered, most of the compounds used in the textbook were composed of simple and frequent single lexical items, for instance: *backpacks, bathroom, bedrooms, birdwatchers, birthplace, blackboard, campfire, caregivers, caretakers, coworkers, desktop, e-book, headphones, newcomer, seawater, seashore*. As stated on the website lextutor.ca. “*The vast majority of compounds are made up of small, frequent words that learners can easily parse or be taught to parse*”. Also, there was only a limited number of abbreviations mentioned, including VND, USD, USB, UNESCO, WWF, etc.

This finding coincided with those from the corpus-based analysis of the Indonesian EFL textbook series by Rahmat and Coshead (2020), in the common use of mother tongue-originated words, especially proper nouns in locally-published English textbooks. The integration of local language items is generally acclaimed by educators and researchers thanks to its effect on cultural values appreciation and exploitation (Nguyen et al., 2021). Moreover, the local students would find the content potentially interesting because the concepts and places are culturally and socially relevant (Rahmat & Coshead, 2020).

Drawing on the above breakdown of Off-lists, it is likely that the words from Offlists would not significantly impede the textbook readers’ overall comprehension, thus, would be rationally combined with the frequency-based lexical coverage to access the number of word families needed to understand the textbook.

Therefore, with the Off-lists, Vietnamese junior high school students needed 3,000-word families to reach 95% coverage of the textbook, which indicates acceptable or reasonable comprehension (Sun & Dang, 2020), regardless of spoken or written text. To be more precise, with a vocabulary size of 3,000-word families, learners could reach a roughly 96.4% coverage of the whole textbook corpus, spoken text and written text.

However, to ensure a better comprehension and incidental vocabulary learning, with 98% lexical coverage of the whole textbook (Nation, 2001), the students needed 5,000-word families, which was the same as the figure needed for the written text, whilst 98% coverage of the spoken text could be achieved with a vocabulary size of 4000 word-families. These results strongly resonated with Nguyen’s (2020) conclusion on the lexical coverage of reading passages in the same EFL textbook series in Viet Nam. He found out that in order to reach the 95% and the 98% coverage, the students needed to have a receptive knowledge of the first three and the first five 1,000-word lists, respectively. It can be seen that the vocabulary load compiled from other parts of the textbook such as Getting Started, Language, Skills (Speaking, Listening, Writing), Communication and Culture, and Looking Back and Project could not make any difference to the level of difficulty of the reading passages’ lexical resources.

More importantly, the required vocabulary size of 3,000-word families for 95% coverage and 5,000 word families for 98% coverage was seemingly a huge challenge for Vietnamese junior high school students, whose working receptive knowledge was identified at 2000 word families (Nguyen, 2020). It is noteworthy that this level of lexical knowledge was applied to all three grades of the high school level as a whole group, but it’s highly likely that the junior high schoolers in grade 10 have a smaller vocabulary load than the seniors students of grade 11 and grade 12, which may be much less than 2,000-word families.

Even if 2,000-word families were considered as the gauge for measuring lexical coverage of the 10th-grade high school students, they could only reach around 92% coverage of the analysed textbook corpus, regardless of spoken or written text. In other words, the students would need to learn 1000 new word families to reach the threshold of 95% coverage to properly comprehend the textbook. This target seems far from achievable when research shows EFL learners are likely to take in 430 words as the maximum number of new words per year (Webb & Chang, 2012). A large number of novel words in the textbooks would negatively impact learners’ comprehension and skills development. As students struggle with too large a bulk of unknown vocabulary, especially in listening activities, they would consequently fail to cultivate sufficient understanding of the text and be unlikely to be good readers of English (Nation, 2013). Moreover, as they focused on working out the meaning of the novel words, other aspects of learning such as sub-skill development would be neglected. (Nguyen, 2020; Rahmat & Coxhead, 2020, Sun & Dang, 2020).

When comparing the desired vocabulary size to reach the 98% thresholds of lexical coverage in the spoken and written text in question, Vietnamese 10th graders were expected to know 4,000 word-families to independently listen and understand the spoken text, which was smaller than the required vocabulary size of 5,000 word families for the

written text. This implies that the language presented in the spoken discourse of the textbook seemed to be less demanding than that of the written ones, which confirmed the results of previous studies on the vocabulary coverage in different kinds of spoken and written discourse (Nation, 2006) where the vocabulary size to comprehend spoken discourses are generally smaller than that for written ones. This finding seemed irrelevant in relation to the objectives of the study, however, it would serve as evidence that the textbook writers must have taken this issue into consideration while selecting the lexical resources for the listening activities of the textbook.

4.2. Coverage of high-frequency word families in the textbook

The second research question focuses on the coverage of high-frequency word families in the textbook, which includes the 1st and 2nd 1,000-word lists.

Table 4. Coverage of the high-frequency word families in the textbook

Frequency Level	Number of word-families	Percentage of word-families
1st 1,000 word list	868	86.8%
2nd 1,000 word list	548	54.8%

Table 5. Occurrences of the high-frequency word families in the textbook (by percentage)

Number of occurrences	1st 1,000 word-families	2nd 1,000 word-families
0-1	10.6%	27.4%
2 to 6	28.4%	44.9%
7 or more	61.0%	27.7%
10 and more	49.1%	17.3%
15 and more	37.9%	10.0%
20 and more	31.3%	7.1%

As can be seen from Table 4, the textbook covered 86.8% of all the word families in the first 1,000-word list. This figure revealed that the textbook covered the most frequent word families in English fairly well. Meanwhile, Table 5 shows that about 61% of those words from the 1st 1,000-word list occurred more than 7 times, which might be the threshold for deliberate vocabulary learning to happen. However, only 49.1% of the 1st 1,000-word families mentioned in the text are encountered 10 times and more; 37.9% encountered 15 times and more; only 31.3% were encountered 20 times and more. This means the possibilities for learners to acquire new vocabulary through incidental learning with reading activities and listening activities were not really significant.

Meanwhile, the coverage of the 2nd 1,000-word list in the textbook was not really sufficient with only 54.8% of the 2,000-word families mentioned in the corpus. As for the occurrences of word families in the 2nd 1,000-word list, only 27.7% of those word families were encountered 7 times and more, 17.3% encountered 10 times and more, 10% encountered 15 times and more, and only 7.1% encountered 20 times and more. It can be seen that the textbook did not represent the second 1,000-word list very well.

For both of the two first 1,000 word lists, there were altogether 584-word families of the English language that never appeared in the textbook. In the context of EFL teaching and learning in Vietnam, access to the English language outside the classroom would be fairly limited, which highlights the crucial role of textbooks as the primary and even only source of linguistic input for EFL learners. Such representation of the second most frequent and dominant word families would result in insufficient and inauthentic lexical knowledge of the learners.

In comparison with the findings of similar research, the coverage of the high-frequency word families in the textbook in question was even lower than the coverage of 98.5% and 86.7% for the 1st and 2nd 1,000-word list respectively of the Indonesian EFL textbook series examined by Sun and Dang (2020). However, the findings were similar to the previous ones in pointing out the generally poor representation of the locally-published EFL textbook over the high-frequency word families (Rahmat & Coxhead, 2020; Alsaif & Milton, 2012).

4.3. Implications for English teaching

The findings indicated that the textbook could be challenging for Vietnamese 10th grade students to read without assistance (Webb, 2008), which highlights the role of teachers in the classroom exploitation of the textbook. In the case of this very study, it seems to be true that Vietnamese teachers are not granted great freedom over the selection of textbook. Plus, this textbook has been widely used across the country, which means the level of English proficiency among learners would dramatically vary. As a consequence, textbook adaptation would be highly desirable and advisable.

First, teachers are advised to identify the vocabulary size of the targeted students prior to the beginning of the course/term with the Updated Vocabulary Levels Test (UVLT) developed by Webb, Sasao and Ballance (2017). The results of the UVLT would serve as a reference for teachers in adapting the textbooks to help learners achieve the best language acquisition from the materials.

Secondly, teachers should familiarize themselves with the English 1,000 word lists, especially ones of the high-frequency word families so that they can make a sensible selection of words for pre-teaching and emphasizing higher possibilities of vocabulary retention. Those word lists can be found and explored thoroughly on the website <https://www.lex tutor.ca/>.

Once being aware of learners' existing lexical knowledge, teachers can make decisions about how to assist them in comprehending the textbook and reaching the desired vocabulary. As Webb & Nation (2008) proposed, teachers can consider taking the following actions:

- (i) pre-teaching vocabulary
- (ii) allowing learners to use a dictionary and look up new vocabulary during the class
- (iii) using the glossing technique for novel vocabulary
- (iv) simplifying text for lower-level students
- (v) replacing the too difficult text with more appropriate ones

In the context of teaching English in public schools in Viet Nam, the last suggestion seems not very feasible for teachers as they are commonly required to closely follow the prescribed subject programs and textbooks. However, other solutions are worth considering and applying in their English classes.

Pre-teaching vocabulary

In skill lessons, integrating vocabulary pre-teaching as a step prior to the while-skill stage would be a helpful solution to facilitate learners' comprehension as well as skill development in both reading and listening lessons (Webb, 2009). Otherwise, learners are highly likely to be puzzled by too many unknown words and struggle to focus on developing the targeted skills. Effective vocabulary pre-teaching requires teachers' ability to anticipate learners' novel words based on their understanding of learners' proficiency and background knowledge.

Moreover, it's important to avoid over-teach the vocabulary, i.e. teaching too many new words from the targeted text, which might undermine the following comprehension and skill-focusing activities. Nation (2001) also stated that pre-teaching vocabulary should work more effectively when dealing with a small number of unknown word families.

As revealed from the listening scripts analysis, with the working lexical knowledge of 2,000-word families, Vietnamese 10th graders could probably understand 93.3% of the spoken corpora, which was only 1.7% short of the first threshold of 95% for "general understanding with support". Thus, pre-teaching vocabulary should be effective in listening lessons in the textbook.

Using a dictionary and glossing

When there is a big gap between the learners' vocabulary coverage with the desired one, learners may find it more helpful to refer to a dictionary or a customized glossary for the lesson. Actually, it would be nearly impossible for teachers to pre-teach a large number of novel lexical items within the time limit of in-class sessions, which is 45 minutes in Viet Nam. Instead, glossing allows learners to read a text relatively quickly by checking the relevant definitions or L1 translations when necessary (Webb & Nation, 2008). Regarding the glossing technique, Nguyen (2020) made a further note that inter-liner glossing faring is far better than a footnote, marginal and appendix glossing. In other words, learners seem to take in the new words faster when they are explicitly displayed in the text.

In developing the glossary for learners of this textbook, which were found not to represent the high-frequency words very effectively, teachers should include the most frequent words in the glossary to increase the chances of vocabulary encountering and learning.

Simplifying the text

When dealing with difficult texts, teachers may try replacing the lower-frequency words with the more simple and high-frequency ones to enhance learners' potential vocabulary coverage of the text, meanwhile increasing the possibilities of reencountering the high-frequency word families, which, in turn, would promote incidental vocabulary learning with these fundamental word-families. In this regard, teachers may also refer to graded readers as a source of simplified text for 'students with vocabulary knowledge lower than the most frequent 3,000 words. Many words in graded readers are repeated often, many more than ten times, which can support students' vocabulary acquisition (Webb & Macalister, 2013).

5. CONCLUSION

The study examined the vocabulary load of a locally-published EFL textbook in Viet Nam for high school students focusing on the vocabulary coverage in relation to the 95% and 98% threshold for text comprehension and the coverage of high-frequency word families in the textbook. It is revealed that students need to have a vocabulary size of 3,000 and 5,000-word families to reach the coverage of 95% and 98% of the whole textbook. However, these requirements would be unrealistic to Vietnamese high school students due to their much smaller vocabulary knowledge. Moreover, the textbook was found to insufficiently cover the second 1,000-word list with the appearance of only over a half of the second most frequent word families. These findings highlight the role of teachers in adapting the textbook for more effective vocabulary learning and general comprehension of the materials by pre-teaching vocabulary, using dictionaries, glossing and simplifying the text. Teachers are also advised to improve their knowledge of vocabulary evaluation tools such as the Updated Vocabulary Levels Test and the BNC-COCA 1-25K wordlists.

This study only covers a small corpus of one textbook in a local EFL teaching context. Therefore, the findings from this study would not be generalizable for other EFL contexts or common EFL learners. Future research would extend the scope of this study to the whole series of the English textbook for all three levels of education in Viet Nam or the newly-published textbooks by various private publishers in the country for a broader and different perspective.

Conflict of Interest: No potential conflict of interest relevant to this article was reported.

REFERENCES

- Alfaki, I. M. (2015). Vocabulary Input in English Language Teaching: Assessing the Vocabulary Load in Spine Five. *International Journal of English Language and Linguistics Research*, 3(1), 1-14. <https://doi.org/10.1017/CBO9781107415324.004>
- Alsaiif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *Language Learning Journal*, 40(1), 21-33. <https://doi.org/10.1080/09571736.2012.658221>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/10.3316/QRJ0902027>
- Cao, T. H. P. (2018). *Vocabulary in EFL Textbook: An Analysis of "Life A2-B1" Coursebook Used for Vietnamese Tertiary Students*. Proceedings of the 7th Vietnamese Young Researchers Conference in Education at Hanoi National University of Education, Vietnam (section 3, pp. 548-556). Hanoi National University of Education Publishing House.
- Dang, H. T. (2017). *An evaluation of Vocabulary used in the textbook "Tieng Anh 12" published by Vietnam Education Publisher in 2008*. Unpublished M.A. Thesis. Vietnam National University, Ha Noi.
- Dang, T. N. Y., & Webb, S. (2020). Vocabulary instruction and the good language teachers. In C. Griffiths, Z. T., & Brown, A. (Eds.) *Lessons from good language teachers* (pp. 203-218). Cambridge University Press.
- Hoang, V. V. (2010). The Current Situation and Issues of the Teaching of English in Vietnam. *Ritsumeikan Studies in Language and Culture*, 22(1), 7-18.

- Hoang, V. V. (2015). The Development of the Ten-Year English Textbook Series for Vietnamese Schools under the National Foreign Language 2020 Project: A Cross-Cultural Collaborative Experience. *VNU Journal of Science: Foreign Studies*, 31(3), 1-17.
- Hoang, V. V. (2016). Renovation in Curriculum Design and Textbook Development: An Effective Solution to Improving the Quality of English Teaching in Vietnamese Schools in the Context of Integration and Globalization. *VNU Journal of Science: Education Research*, 32(4), 9-20.
- Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Multilingual Matters.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15.
- Nation, I. S. P. (2001). *Learning vocabulary in another language* (Cambridge Applied Linguistics). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P., & Beglar, D. (2007) A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nguyen, C. D. (2020). Lexical features of reading passages in English textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal*, 52(3), 509-522. <https://doi.org/10.1177/0033688219895045>
- Nguyen, T. T. M., Marlina, R., & Cao, T. H. P. (2021). How well do ELT textbooks prepare students to use English in global contexts? An evaluation of the Vietnamese English textbooks from an English as an international language (EIL) perspective. *Asian Englishes*, 23(2), 184-200. <https://doi.org/10.1080/13488678.2020.1717794>
- Prime Minister of SRV (2008). *Teaching and Learning Foreign Languages in the National Education System, Period 2008-2020* (Issued under Decision No. 1400/QĐ-TTg signed by the Prime Minister of the Government of the Socialist Republic of Vietnam).
- Rahmat, Y. N., & Coxhead, A. (2021). Investigating vocabulary coverage and load in an Indonesian EFL textbook series. *Indonesian Journal of Applied Linguistics*, 10(3), 804-814. <https://doi.org/10.17509/ijal.v10i3.31768>
- Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93, 102279. <https://doi.org/10.1016/j.system.2020.102279>
- Vietnam MOET. (2010). *Pilot English Curriculum for Vietnamese Primary Schools* (Issued under Decision No. 3321/QĐ-BGDĐT by the Vietnam's Minister of the Ministry of Education and Training).
- Vietnam MOET. (2012a). *Pilot English Curriculum for Vietnamese Lower Secondary Schools* (Issued under Decision No. 01/QĐ-BGDĐT by the Vietnam's Minister of the Ministry of Education and Training).
- Vietnam MOET. (2012b). *Pilot English Curriculum for Vietnamese Upper Secondary Schools* (Issued under Decision No. 5290/QĐ-BGDĐT by the Vietnam's Minister of the Ministry of Education and Training).
- Vietnam MOET. (2020). *Circular No. 25/2020/TT-BGDĐT on Procedure to select textbooks in general schools*. <https://luatvietnam.vn/giao-duc/thong-tu-25-2020-tt-bgdtdt-quy-dinh-lua-chon-sgk-trong-co-so-giao-duc-pho-thong-189806-d1.html>
- Webb, S. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *The Canadian Modern Language Review*, 65, 441-70.
- Webb, S. A., & Chang, A. C. S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113-126. <https://doi.org/10.1177/0033688212439367>
- Webb, S., & Nation, P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1-9.
- Webb, S., & Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, 47(2), 300-322. <https://doi.org/10.1002/tesq.70>