Investigating vocabulary coverage and load in an Indonesian EFL textbook series

Yurieke Nadiya Rahmat and Averil Coxhead

School of Linguistics and Applied Language Studies, Faculty of Humanities and Social Sciences,
Victoria University of Wellington, PO. BOx 600, Wellington 6140, New Zealand

ABSTRACT

This article reports a corpus-based study of a series of Indonesian EFL textbooks for senior high school students from grades 10 to 12 (16-18 years old), published by the Indonesian government. The textbooks were collected from the Indonesian bookkeeping information system and transferred into ready-to-analyse corpora using Range Programme (Heatley et al., 2002) to establish the vocabulary load and what words are available for learning. The most significant finding in this study showed that the textbooks require 3,000-4,000 word families to reach 95% coverage with some help needed to comprehend the textbooks and 5,000-6,000 word families to gain 98% coverage which means the students can cope independently to read the textbooks. The textbooks contain a large amount of high-frequency words at more than 80% and up to 7% of the text is made up of Indonesian words which play roles in learning objectives and pre-vocabulary teaching activities. The results of this study suggest the vocabulary features and implications for learning, present the pedagogical implications for teachers and textbook writers, and provide a springboard for future research.

Keywords: Indonesian EFL textbooks; Indonesian words; vocabulary; vocabulary load

INTRODUCTION

In Indonesia, English is the first foreign language officially taught in formal education institutions from kindergarten to university levels (Dardjowidjojo, 2002). It is compulsory for three years in both junior and senior high schools (Hamied, 2012; Sudarman & Chinokul, 2018). However, even after six-years learning English at school, Indonesian students' English proficiency remains low (Kirkpatrick, 2012). In particular, they have a low level of vocabulary knowledge in English (Nurweni & Read, 1999; Mustafa, 2019). One of the reasons behind this issue could be due to the lack of vocabulary exposure in learning resources, especially textbooks, which are the fundamental resources for lexical input in the classroom (see Häcker, 2008; Guerrettaz & Johnston, 2013).

Textbooks provide opportunities for learners to gain language exposure through reading written texts. Urquhart and Weir (2014) define reading as "the process of receiving and interpreting information encoded in language form via the medium of print" (p. 22). In other words, when learners are reading, they are encoding the meaning of words, phrases, sentences, and the whole texts to make sense of what has been written. To cope well in reading comprehension, learners need to be familiar with many words in reading texts. If learners have low levels of vocabulary knowledge, the process of understanding the texts cannot be fully achieved. That is, there is a close relationship between reading and vocabulary. The more

Corresponding Author
Email: yuriekenadiya@gmail.com

language learners read, the better their vocabulary would be. The more words they know, the better reading comprehension can be (Nation, 2013).

Several studies focused on a corpus-based analysis on vocabulary in textbooks from EFL (English as English as Foreign Language) contexts, such as Sun & Dang (2020) in China; Alsaif & Milton (2012) in Saudi Arabia; Nguyen (2020) in Vietnam; and Mukundan & Aziz (2009) in Malaysia; Aziez & Aziez (2018) in Indonesia (see literature review). However, to the best of our knowledge, none of these studies analyzed the textbooks in this current context which were from the Indonesian government and have been used by approximately 4.8 million senior-high school students ranging from grades 10 to 12 (16-18 years old) in Indonesia. These textbooks are mainly for general English purposes and the students are mostly at the beginner level of English and have a working vocabulary of less than 2,000 word families (see Nurweni & Read, 1999; Mustafa, 2019). Since the textbooks in this study were used by many students in Indonesia as English learning resources, it is highly needed to look deeply into this specific context regarding what words the students might encounter and how many words needed for learners to cope well with reading the textbooks.

This study will then focus on whether or not the textbooks are suitable for Indonesian learners in this context. This study aims to fulfil the mentioned gap through the investigation of what vocabulary learners will encounter while reading these Indonesian EFL textbooks and this will shed light on the extent to which these textbooks can support learning vocabulary in Indonesian context particularly and in other EFL context generally. To achieve this purpose, the focus of this study will be: (1) to investigate how many words that the Indonesian students need to know to comprehend the textbooks; (2) to know the coverage of high-, mid-, and low-frequency words, and supplementary lists (Indonesian words, proper nouns, marginal words, compound nouns, abbreviations, and other languages) over the textbooks; (3) to analyse the roles of Indonesian vocabulary in the textbooks.

Vocabulary load in reading

The concept of lexical thresholds in reading was developed by Laufer (1989). It refers to how many words in terms of percentage learners need to know for reading comprehension. She states that learners need to understand 95% of the total words in written texts for comprehension. If a text is made up of 500 words, learners need to know 475 words. Hu and Nation (2000) investigated thresholds further and found that 98% coverage is highly needed for better reading comprehension. It means learners are encouraged to know 490 out of 500 words in a text. Schmitt et al. (2011) and van Zeeland and Schmitt (2013) support that the 98% coverage becomes a desirable goal for learners to fully understand a text. Both the 95% and 98% thresholds emphasize how important is it for learners to know as much vocabulary as possible to help them understand written texts better (see Nation, 2013). Otherwise, they would find it difficult to comprehend the texts while reading.

Nation (2006) showed how many words were needed to gain 95% and 98% in different texts (see Table 1) using the British National Corpus (BNC) word levels. For example, he estimated that if students read novels, they will need 4,000 word families to gain 95% coverage but 9,000 word families to gain 98%. It should be noted that both 95% and 98% coverage in Table 1 include proper nouns. Table 1 also shows findings from Webb and Macalister (2013), who analysed the lexical load of three different corpora (e.g. graded readers, children's school journal, and the Wellington written corpus) and found that graded readers require learners to have 2,000 word families to reach 95% and 3,000 word families to reach 98% coverage.

Table 1

The Vocabulary Size Needed to Reach 95% and 98% Coverage (Plus 1 – 6% Proper Nouns) of Different Texts (Adapted from Nation, 2006; Webb & Macalister, 2013)

Texts	95% coverage	98% coverage
Graded readers	2,000 word families	3,000 word families
Spoken English	3,000 word families	7,000 word families
Novels	4,000 word families	9,000 word families
Newspapers	4,000 word families	8,000 word families
Children's movies	4,000 word families	6,000 word families
Children's school journals	5,000 word families	10,000 word families
Wellington written corpus	5,000 word families	10,000 word families

Both the children's school journal and Wellington written corpus require 5,000 word families to reach 95% and 10,000 word families to reach 98%. These findings suggest that graded readers are the most suitable reading texts for L2 learners because they are written using a controlled vocabulary with a few words outside the most frequent 1st 2,000 (Webb & Macalister, 2013). Many words in graded readers are repeated often in the graded readers, many more than ten times, which can support students' vocabulary acquisition (Webb & Macalister, 2013). Note that Table 1 aims

to show the examples of how many words needed to reach 95% and 98% in a variety of texts. However, this is not the focus of this study to match the textbooks used in this study with texts from Nation (2006) and Webb and Macalister (2013) and thus this table will be referred to in the pedagogical implications section.

While a 3% difference between 95% and 98% may seem small, it represents a large number of word families. A word family consists of a headword and its derivations and inflections (see Nation, 2013), for example, the word family for *comprehend* includes *comprehends*, *comprehended*, *comprehending*, and *comprehension.* These typical words can be encountered by learners and their understanding of the prefixes and suffixes may improve as they become more proficient in the language. It is then needed to make a decision on using word families when investigating how many words our students have known and need to know (Nation, 2013). Hence, this study used the word families in counting the vocabulary. Now, let us turn to the concept of frequency and the kinds of words learners might encounter in texts, based on Nation (2006).

Nation's 2006 word categories (high, mid, and low-frequency words) and supplementary lists
The concept of word categories (high, mid and low) by Nation (2006) is needed to help decide how much vocabulary needs to be learned and what words learners encounter in a corpus. This can be applied in any contexts, including ESL and EFL contexts. The first word category is high-frequency words, which are the most frequent and dominant in any texts (Nation, 2006; Vilkaitė-Lozdienė & Schmitt, 2020). They consist of the 1st and 2nd 1000 word families (e.g. *agree, at, both, children,* and *better*) and these words cover over 80% of the words in most texts (Vilkaitė-Lozdienė & Schmitt, 2020). They include function words such as *some, because, three, at,* and *in* (Nation, 2013). Because of the dominance and frequency of high frequency words, Nation (2013) argues that it is necessary for all language learners to learn such words and should require attention from both teachers and students. Macalister (2014) and Nation (2013) also stated that adapting four stands of meaning-focused input (MFI), meaning focused output (MFO), language-focused learning (LFL), and fluency development (FL) in textbooks, learning materials, and curriculum design may be useful for learning a language. While MFI is related to comprehensible input through listening and reading, MFO is based on knowledge in which the learners gain from speaking and writing activities (Nation, 2013). The LFL emphasizes on learning language items, for instance, grammar and parts of speech (Nation, 2013). The remaining strand is FL which involves

regular fluency practice for learners which leads to better language comprehension (Nation, 2013).

Mid-frequency words occur less frequently than high frequency words, but they are, 'frequent enough to be a sensible learning goal after the high-frequency words' (Nation, 2013, p. 25). Mid-frequency includes the 3rd 1000 to 9th 1,000 word families (Nation, 2013). Mid-frequency vocabulary covers more frequency bands than high frequency vocabulary, and learners will encounter these words less often than high frequency vocabulary. Some examples of mid-frequency words include *approve, comprehend*, and *earthquake*. Low-frequency words consist of words that appear rarely, have only small coverage in any texts, and include frequency bands from the 9th 1,000 and beyond (Nation, 2013; Schmitt & Schmitt, 2012). Some examples of words in these lists include technical vocabulary in specific subject areas such as *ecotourism, proboscis, transitive, regency*. These three categories provide guidance on which words learners need to learn first (high frequency), and to understand the role of frequency in vocabulary for language learning. Supplementary lists such as proper nouns and other lists are included for the analysis (see methodology section).

Vocabulary analysis in EFL textbooks at secondary school level
Studies on vocabulary found that high school students need to know many words in English to cope with their textbooks and that the books often included words outside the much needed high frequency bands. In Saudi Arabia, Alsaif and Milton (2012) investigated 22 EFL school textbooks from Year 6 to Year 12 and found that the textbooks covered 84% of the most common 1st 2,000 word families and 55% of the most frequent 1st 5,000 word families. In Hong Kong, Tang (2009) investigated six EFL secondary school textbooks and revealed that these textbooks only contained around half of the most frequent 2,000 word families. In Malaysia, Mukundan and Aziz (2009) examined five English secondary school textbooks and reported that the most frequent 2,000 word families were not introduced and repeated frequently. Sun and Dang (2020) investigated a 273,094-word corpus of high-school EFL textbooks in China and found that the textbooks covered the most frequent 1st 1,000 word families well, but not the 2nd 1,000 and 3rd 1,000 word families. They also found that students need to know 3,000 word families to reach 95% and 9,000 word families to gain 98%. Sadly, knowledge of 3,000 word families is not commonly reported in the literature, and 9,000 word families is well beyond what might be expected at high school level in an L2 context. Aziez and Aziez (2018) analyzed six Indonesian EFL textbooks from both junior and senior high schools published by private companies and found

that Indonesian students need 3,000 word families to comprehend 95%, and more to achieve 98%. These previous studies provide valuable information regarding vocabulary in the EFL textbooks. However to the best of our knowledge, very little is known about secondary-school textbooks in the Indonesian context, especially published by Indonesian government. This is the primary aim of the present study.

The vocabulary size of students in EFL contexts
Several studies found that the vocabulary size of students in EFL contexts are relatively limited. For instance, Nurweni and Read (1999) tested 324 first-year university students in Indonesia by using three vocabulary measures such as an translation test, a word-associates test, and an interview, which are based on the General Service List (West, 1953) and the University Word List (Xue & Nation, 1984). This study found that the students had a low level of vocabulary knowledge, with a worrying lack of knowledge of the 1st 2,000 word families. Another example is that Mustafa (2019) investigated vocabulary size of 1775 students from both junior and senior high-schools in Indonesia who had studied English for at least three years. Using the Nation and Beglar (2007) Vocabulary Size Test (VST), he reported that the participants recognised 72% of the 1st 1,000 word families of English, 54% of the 2nd 1,000, and 49% of the 3rd 1,000.

Low levels of vocabulary knowledge are also found in other EFL contexts. For example, in Taiwan, Webb and Chang (2012) measured the vocabulary knowledge of 166 EFL learners from a vocational senior high school with a minimum of five years of learning English using a bilingual version of the Vocabulary Levels Test (VLT) (Schmitt et al., 2001) and found that 47% of the learners had mastered the 1st 1,000 word families and 16% had mastered the 2,000 word families level over five years. In Hong Kong, Tang (2007) measured the vocabulary size of a total of 449 participants from both primary and junior secondary schools who studied English as a compulsory subject by using L_Lex of the Lingua Vocabulary Tests v2.01 (Meara et al., 2001) and the Vocabulary Levels Test (Nation, 1990). Tang's (2007) study showed that a majority of these learners lacked knowledge of the 1st 3,000 word families.

In China, Sun and Dang (2020) used Webb, Sasao and Ballance's (2017) Updated Vocabulary Levels Test to measure the vocabulary knowledge of EFL high school students. They found that the majority of students did not master the most frequent 1st 1,000 word families. Matthews and Cheng (2015) reported that 167 learners who studied at a large university in China showed poor word recognition in the 1st 2,000 high-frequency words when taking an IELTS listening test. The present study does not include testing of vocabulary knowledge, but it does aim to find out how many words learners in Indonesia who use the textbook series need to know to cope with reading them.

METHOD
This study used a corpus based-analysis which involved the investigation of corpora from a set of three Indonesian EFL textbooks series for senior high school students from grades 10 to 12. The textbooks were firstly published by Indonesian Ministry of Education and Culture in 2013 as English learning and teaching resources based on Indonesian National Curriculum (Kurikulum 13), but this study used the newest revised version from 2015 to 2017 to gain the up-to-date data for corpus analysis. The textbooks were used as English learning and teaching resources based on Kurikulum 13, an Indonesian National Curriculum. A total of 88,554 running words was used in this study for the investigation (see Table 2). Textbooks 1 and 3 have a similar number of words, and Textbook 2 is roughly 10,000 words shorter than the other two.

Table 2
A Summary of the Indonesian EFL Textbooks

Textbooks	Chapters	Pages	Running words
Textbook 1	15	204	32,867
Textbook 2	15	164	22,962
Textbook 3	16	239	32,725
Total			88,554

Preparing the textbooks for analysis
Electronic copies were downloaded from Indonesian Bookkepping Information System by Indonesian Ministry of Education and Culture that can be found in a website (http://buku.kemdikbud.go.id/). We included all the words that occur in the textbooks in order to acknowledge all the words in the textbooks and make them into ready-to-analyze corpora. Also, all proper nouns and other word lists of marginal words (e.g. *um* and *ah*), compound nouns, abbreviations, Indonesian words and a small number of other languages (French and Dutch) are examined along with vocabulary load. To prepare the textbooks for the analysis, we followed these steps (see Coxhead et al., 2020):
- Every page of each book was scanned and converted into a text format file (.txt). Any pages that could not be converted were typed manually.

- All words were carefully checked and edited to identify and correct any mistakes or unknown words. For instance, website sources and phonetic symbols were excluded from the text files, and misspelled words were corrected, e.g. *priviledge (privilege)* and *cigarete (cigarette).*
- Words from languages other than English were identified and grouped into new base-word lists using the Range programme (Heatley et al. 2002). This step included: Indonesian words e.g. *gembira, pasangan, menentukan* and proper nouns (e.g. *Surabaya, Malang, Habibie*), French (e.g. *repondez*) and Dutch (e.g. *eens, imprerator*).
- Hyphenated words were replaced with spaces in the texts so that the hyphenated words were separated and could be counted into single lexical items by Range (Heatley et al., 2002).

Data analysis

Range programme (Heatley et al., 2002) was used to analyze the vocabulary load of the Indonesian textbooks in this study. This programme consists of several base word lists such as the frequency-based 25,000 word families (Nation, 2013) from British National Corpus (BNC)/ Corpus of Contemporary American English (COCA), and supplementary lists for proper nouns, marginal words (e.g. *ah, ouch, xoxo*), compound words (e.g. *afterbirth, backpack*), abbreviations (e.g. *ATM, ASEAN*), and the Indonesian, French and Dutch words, which are relevant to inform how much and what vocabulary appears in these EFL textbooks. Also, Range programme can be found on Nation's website (https://www.wgtn.ac.nz/lals/about/staff/paul-nation#vocab-programs).

FINDINGS AND DISCUSSION
Vocabulary load of the textbooks

Vocabulary load of the textbooks can be seen in Table 3. It shows the cumulative coverage of textbooks with and without the supplementary lists in regards to 95% and 98% coverages. It is important to see the gap between the coverage of with and without the supplementary lists in order to give valuable information how supplementary lists play an important role in reaching 95% and 98% coverages. With the addition of supplementary lists, Textbooks 1 and 2 reach 95% at 3,000 word families and 98% at 5,000 word families. Textbook 3 has a higher vocabulary load than Textbooks 1 and 2, reaching 95% at 4,000 word families plus supplementary lists and 98% at 6,000 word families plus supplementary lists (see columns 3, 5, and 7). This means that 5,000 to 6,000 word families are needed to achieve better reading comprehension at 98% (Hu & Nation, 2000; Schmitt et al., 2011), which seems to be relatively demanding for Indonesian students who mostly have a working vocabulary knowledge of fewer than 2,000 word families (Nurweni & Read, 1999; Mustafa, 2019), even if they are reading their textbooks intensively and meeting these words repeatedly. It is also important to see that none of the textbooks reach either 95% or 98% without the supplementary lists (see columns 2, 4, and 6). Also, the obvious difference in the textbook coverages in Table 3 resulted in the different sets of the coverage of high, mid, and low-frequency words (see Table 3 and 4).

These results suggest that Indonesian students will encounter many unfamiliar words when they read these textbooks and doing so could negatively impact their reading comprehension. This finding is similar to Sun and Dang (2020) on the challenging nature of vocabulary in EFL high-school textbooks in China. Nation (2013) states that learners who do not know many words are unlikely to be good readers of English.

Table 3
Vocabulary Load of the Textbooks, With and Without the Supplementary Word Lists Comparison

Word Lists	Textbook 1 %		Textbook 2 %		Textbook 3 %	
	No suppl. lists	With suppl. lists	No suppl. lists	With suppl. lists	No suppl. lists	With suppl. lists
1st 1,000	73.72	83.74	77.43	84.35	70.67	80.27
2nd 1,000	81.24	91.26	85.11	92.46	79.29	88.89
3rd 1,000	85.41	**95.43**	88.63	**95.98**	85.03	94.63
4th 1,000	86.98	97.00	90.14	97.49	86.57	**96.17**
5th 1,000	88.21	**98.23**	90.99	**98.34**	87.85	97.45
6th 1,000	88.72	98.74	91.53	98.88	88.57	**98.17**
7th 1,000	89.24	98.26	91.84	99.19	89.20	98.80
8th 1,000	89.41	98.74	92.08	99.43	89.46	99.06
9th 1,000	89.49	99.51	92.28	99.63	89.88	99.48
10th 1,000	89.64	99.66	92.36	99.71	90.07	99.67
11th – 25th 1,000	89.98	100	92.62	99.97	90.40	100

The coverage of high-, mid-, and low-frequency words

Table 4 shows the coverage of high-, mid- and low-frequency words in the three textbooks. The proportion of the 1st 2,000 words (high-frequency words) across the textbooks is roughly at 80%, which is similar to Alsaif and Milton's (2012) study and other kinds of written texts (Vilkaitė-Lozdienė & Schmitt, 2020). Mid-frequency vocabulary covers just over 7% in Textbook 2 up to 10.59% in Textbook 3. Low frequency words have the lowest coverage across the textbooks, as expected.

Table 4
The Coverage of Textbooks (%) Based on High-, Mid-, and Low-Frequency Words

Word levels	Textbook 1		Textbook 2		Textbook 3	
	Token %	Cumulative %	Token %	Cumulative %	Token %	Cumulative %
1st 2,000	81.24	81.24	85.11	85.11	79.29	79.29
3rd – 9th 1,000	8.25	89.49	7.17	92.28	10.59	89.88
10th – 25th 1,000	0.49	89.98	0.34	92.62	0.52	90.40

Here is a sample of text in Textbook 1 (Extract 1, Grade 10), from Chapter 1 *Talking about self* (p. 4). In this sample, the words are marked as regular text for high-frequency words, **bold** for mid-frequency words, and *italics* for low-frequency words. Words which are not high, mid or low frequency are underlined (e.g. proper nouns).

I guess I would better tell you something about myself first. I am 16 years old and I attend <u>Thomas Edison</u> High School here in <u>Minneapolis, Minnesota, USA</u>. I have two brothers and two half-sisters, and I am the middle child. My father died a few years ago so my mother runs the house and the family business. My father was a *barista*. I have lots of hobbies. I like music – mostly classical music and fold music? But I do not play an instrument. I like sports, especially **tennis** and **basketball**. At school I am in the basketball team and I spend most of my extra – **curricular** time playing **basketball**. I am into animals very much. My sister and I have three dogs. They need lots of attention as you can imagine. My favourite subjects at school are art and **geography**. I think I would like to become a park **ranger** when I **graduate**, perhaps work for National Parks Service. I do not like reading but I love drawing and painting. How about you? Please drop me a line, <u>Alia</u>! Cannot wait to hear from you!

Excerpt 1. An example of high-, mid-, and low-frequency words in a reading text from Textbook 1

This example illustrates the way that high-frequency words dominate written texts (Nation, 2013) and shows how valuable high-frequency words are for language learners (Nation, 2013: Vilkaitė-Lozdienė & Schmitt, 2020). One reason for the high coverage of the high-frequency word families across the textbooks is the occurrence and repetition of function words in the 1st 1,000 word families, such as *the, to, of, you, a, and, in,* and *or,* which are more than a thousand times in the textbooks. This finding is in line with Nation (2013), who states that function words usually represent the most frequent high-frequency words in reading texts. Other high-frequency words, for example, *ask, because, information, friend,* and *happened* appear more than ten times in the textbooks. This repetition means these words have a greater chance to be learnt by students (see Brown et al., 2008), especially when they are repeated more than ten times (Webb, 2007).

The coverage of mid-frequency words (3rd – 9th 1,000) is the second largest in the textbooks. Textbook 3 contains mid-frequency items that occur more than ten times such as *description, bullying, grammar, punctuation, noun, leopard,* and *tsunami.* Twenty mid-frequency words are shared in the three textbooks. They are related to the textbook itself (e.g. *text, task, column,* and *paragraph*), describing language (*e.g. verb, noun, structure, phrase, passive, grammar,* and *pronouns*), and reading topics (e.g. *bullying*). Of these twenty shared items between all three textbooks, words such as *text, verb, source,* and *structure* are repeated more than ten times in each textbook (Table 5). Repetition means learners might learn those words. A question remains about how useful these words are for everyday language use in English, even if they are very useful for navigating the textbook for learners and teachers.

Lastly, low-frequency words have the lowest coverage over the textbooks at well under 1%. For example, Textbook 1 contains low-frequency technical terms such as *thermodynamics* and *aerodynamics*, which occur only once. Excerpt 1 shows one low-frequency word (*barista*), which also appears once. *Prepositions* appears one or two times across the textbooks. This result is in line with Nation (2013) who points out that words beyond 9th 1,000 may occur infrequently. These words are of questionable value in high school textbooks for second language learners.

Indonesian words, proper nouns and other lexical items in the textbooks

Outside the frequency bands, there is a range of other vocabulary in the textbooks. Table 6 shows the proportion of Indonesian words, proper nouns, transparent compounds, and other lexical items in the textbooks. The total coverage of all these supplementary lists is highest in Textbook 1 at 10.02%, followed by 9.60% in Textbook 3. The

most striking feature of Table 6 is the percentage of Bahasa Indonesian words at 7.21% in Textbook 1, 4.46% in Textbook 2, and 7.28% in Textbook 3. Indonesian proper nouns also occur in the textbooks, for example in *the Battle of Surabaya* (Chapter 9), *B.J Habibie* (Chapter 10), and *Cut Nyak Dien* (Chapter 11). These words are familiar to Indonesian students and the content is potentially interesting because the concepts and places are culturally and socially relevant. Having Indonesian content is likely to motivate students in reading as they are familiar with the given contexts and can use their background knowledge to facilitate comprehension (see Nation, 2013).

Table 5
The 20 Most Frequent Shared Mid-Frequency Words in the Three Textbooks by Frequency

Words	Textbook 1	Textbook 2	Textbook 3	Total
text	137	46	249	432
task	141	2	250	393
verb	85	20	60	165
paragraph	74	5	71	150
noun	93	2	37	132
bullying	1	40	60	101
source	35	39	15	89
structure	26	16	36	78
verbs	21	7	48	76
description	22	3	40	65
phrase	28	2	17	47
grammar	18	1	27	46
etc	17	8	17	42
passive	8	13	8	29
pronoun	20	6	1	27
respond	5	3	19	27
texts	6	4	16	26
column	15	1	9	25
paragraphs	8	1	16	25
appropriate	3	1	17	21

Table 6
The Coverage of Supplementary Lists across the Textbooks

Word Lists	Textbook 1	Textbook 2	Textbook 3
Indonesian words	7.21	4.46	7.28
Proper nouns	1.65	1.58	0.84
Marginal words	0.43	0.54	0.96
Transparent compounds	0.58	0.65	0.44
Abbreviations	0.14	0.12	0.09
Other languages	0.00	0.03	0.00
TOTAL	10.02	7.35	9.60

The proper nouns cover the next highest proportion of texts of the supplementary lists across the textbooks. The coverage is similar in Textbook 1 and Textbook 2 at around 1.6%, but drops to 0.84% in Textbook 3. Examples of proper nouns include *Seattle, Frank,* and *Peterson* in Textbook 3. Marginal words (e.g. *oh, wow,* and *buzz*), compound words (e.g. *classmates, headlines,* and *online),* abbreviations *(e.g. Voice of Asia (VOA), Best Friend Forever (BFF)* and *Influenza Virus (N5H1)),* and words from other languages (mainly French and Dutch) (e.g. *eens, ik, voor,* and *imperator)* cover less than 1% of the textbooks. The French and Dutch words appear in a reading passage in Textbook 2 (p. 159) about Ki Hajar Dewantara, an Indonesian educational activist during colonialism. The next section returns to Indonesian words in the textbooks so that we may understand more about their role.

The role of Indonesian words in the textbooks
The analysis of the Indonesian words in the textbooks found that they occur mostly in the learning objectives and vocabulary exercises. The learning objective sections aim to facilitate students' understanding of what goals they need to achieve in every lesson. These sections account for roughly 50 Indonesian words in Textbook 1 and 2 and 100 Indonesian words in Textbook 3. Several Indonesian words which look similar in written and spoken forms to English words in this section occur frequently, such as *teks (*text*), struktur (*structure*), pasif (*passive*), kolom (*column*), paragraf* (paragraph*), komunikasi (*communication*), dan (*and*), sosial (*social*)* and *konteks (*contexts*)* (see Table 5). This shows us that many Indonesian words in these textbooks seem to have direct translations from English, which may help the students to recognize particular English words faster because they are cognates.

Indonesian words also occur as the translation of English words in matching exercises in pre-teaching activities. The purpose of these exercises is to increase the chance to remember these words when learners encounter them in reading texts. In each exercise, 10 words at maximum are introduced to the students. For example, words such as *sound, run, attend, reserved, distant, stadium, commuter, approval, proceed,* and *unforeseen* appear in vocabulary exercises. These words include only high-frequency items (e.g. *sound, run, attend,* and *reserved*) and also mid-frequency words (*distant, stadium, commuter, approval, proceed,* and *unforeseen).* This means that the target words in the vocabulary sections provide learners with the knowledge of vocabulary which is beyond the 1st 2,000 word families. The learners may benefit from such exercises as they may have a variety of vocabulary and thus increase the opportunity for retrieval when the words are repeated frequently. However, the target words only appear in a particular chapter but do not reoccur in other chapters. Students may first remember these words but forget them easily because they do not encounter them repeatedly in other reading texts (see Nation, 2013).

PEDAGOGICAL IMPLICATIONS
This section highlights the pedagogical implications of this study. It begins with vocabulary features and their implications in learning, and pedagogical implications for teachers and textbook writers.

Vocabulary features of the textbooks are shown in Table 7 to provide recommendations for students' needs. The table also highlights the importance of building knowledge of high-frequency words in English, and this research shows that the Indonesian textbooks contain a large proportion of these words. It is important for language teachers to firstly measure students' knowledge of high-frequency vocabulary and up to the 5th 1,000 word families by using Webb et al. (2017) Updated Vocabulary Levels Test. This would help teachers to understand their students' vocabulary levels, and to make full use of the textbooks to facilitate students' vocabulary learning.

Table 7
Vocabulary Features in Indonesian EFL Textbooks and Implications for Learning

Word category	Features of these words	Possible implications of what the learners need
High-frequency words	- Include function words (e.g. *in, on,* and *at*) and many content words (e.g. *represent, adoption,* and *production*); - Cover almost 80% of written texts (Vilkaitė-Lozdienė & Schmitt, 2020); - Are likely to be repeated frequently.	- Learners need to have more exposure and direct learning of these words especially at beginner level. The words should be repeated frequently with a focus on their form, meaning, and use (see Nation, 2013) so that it provides learning opportunities for students to be familiar with such words; - Teachers can apply vocabulary learning strategies such as the use of word cards or online word card programmes for direct learning. Repeated activities can trigger students to be familiar with the target words.
Mid-frequency words	- Appear less frequently than high-frequency words; - Are related to textbooks, language description, and reading topics in these textbooks.	- Learners need to be exposed to the target mid-frequency words in readings with a great chance of repetition. - Teachers can create learning opportunities to learn mid-frequency words by providing helpful reading texts and encouraging deliberate learning (see Nation, 2013).
Low-frequency words	- Contain technical terms - Occur less frequently.	- Learners can learn several strategies to deal with low-frequency words in reading texts. For example, guessing the meaning from the contexts
Indonesian words	Appear in learning objectives and pre-vocabulary teaching activities.	- Learners need to be aware of what to achieve before learning and the meaning of words from English to Indonesian. - Indonesian words can be used in form of translation to help building vocabulary knowledge in English, especially for beginner students.
Proper nouns	Consist of familiar names, people, places, and topics.	- Learners need more exposure to familiar contexts in readings about topics, people and places which are known by the students to help them understand the reading texts easily due to their background knowledge (see Nation, 2013).

Teachers could also augment the textbooks with graded readers, for instance, to help with high-frequency words, as can be seen at https://erfoundation.org/wordpress/. Graded readers are suitable for Indonesian students who have low vocabulary knowledge because they are written with controlled vocabulary mostly made up of high frequency words (Webb & Macalister, 2013; see also Table 1) and have the greatest repetitions of word families. Many words can be encountered more than ten times in graded readers (Webb & Macalister, 2013). Textbook writers can use the Range programme to identify the vocabulary in texts. For example, writers can focus on high frequency words that are worth learning, and consider carefully what mid-frequency words might be worth learning. They can also ensure there are repetitions of target words in textbooks. Writers and teachers need to consider adapting Nation's Four Strands of meaning-focused input, meaning-focused output, language-focused learning, and fluency development in textbooks, learning materials and curriculum design (see Nation, 2013; Macalister, 2014). Macalister (2014) states that all four strands should be considered in language learning as a whole unit, not in every activity. This balanced idea would strengthen students' vocabulary acquisition that may lead to the process of retention of vocabulary (Nation, 2013).

The implications are also based on the inclusion of Indonesian words in the textbooks. The use of Indonesian words is advantageous for learners at the beginner levels. EFL teachers can make use of the words found in pre-vocabulary teaching exercises to introduce the meaning of English words, which would help students build their vocabulary knowledge by focusing on form, meaning, and use (see Nation, 2013). Textbook writers need to be aware of the coverage of Indonesian words in the textbooks. When it is high, it may decrease the chances of students being exposed to English words. Thoughtful consideration is needed when integrating Indonesian words in the textbooks. Lastly, the use of proper nouns which are familiar to student contexts are useful in reading texts. They can create interest and allow students to make use of their background knowledge so that it is easier for them to understand the texts due to their familiarity with the topics (see Nation, 2013).

LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE RESEARCH

This research is a small-scale study of a textbook series that focus on a limited number of textbooks from one single publisher. It is difficult to generalize beyond this focus. The analysis of this series cannot represent a complete investigation of whether and how EFL textbooks in Indonesia support students' vocabulary acquisition. This study does not specifically categorize the analysis of the vocabulary based on reading texts, instructions, and exercises so that more focused in terms of their occurrence in the textbooks should be considered. Also, this study does not investigate how teachers and students use textbooks to enhance vocabulary teaching and learning. Teachers, for example, may choose several lessons to use in the class and skip more texts due to limited time, which in turn may reduce the chance for students to encounter words in context and gain repetitions needed for memory enhancement. Recommendations for future research include conducting a large-scale study of more textbooks in Indonesia, as a replication of this study, and to gain a wider view of textbooks and vocabulary. Other avenues could be investigating in depth how teachers and students use textbooks for vocabulary learning, analysing multiword units in the textbooks and testing knowledge of collocations (words that occur together, e.g. *equal rights*, *high standards*), using Nguyen and Webb's (2017) test.

CONCLUSION

This current study analysed the nature of vocabulary in three EFL textbooks used by high-school students in Indonesia and published by the Indonesian Ministry of Education and Culture by using the Range program (Heatley et al., 2002) with the 25 1,000 word family lists of BNC/COCA. The findings suggest that the textbooks are relatively challenging for the students because they require the learners to know a large amount of vocabulary (5,000-6,000 to achieve 98% coverage). The textbooks contain a large number of words from supplementary lists, especially Indonesian words which mostly occurred in learning objectives and vocabulary exercises. With similar results as high-school textbooks in other EFL contexts such as Sun & Dang, (2020), Alsaif & Milton (2012), and Nguyen (2020), this current study gives valuable information on the nature of textbooks that EFL learners are exposed to.

ACKNOWLEDGEMENTS

REFERENCES
Alsaif, 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. *The Language Learning Journal*, *40*(1), 21-33.

https://doi.org/10.1080/09571736.2012.65822 1

Aziez, F., & Aziez, F. (2018). The vocabulary input of Indonesia's English textbooks and national examination texts for junior and senior high schools. *TESOL International Journal*, *13*(3), 66-67.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Coxhead, A., Rahmat, Y., & Yang, L. (2020). Academic single and multiword vocabulary in EFL textbooks: Case studies from Indonesia and China. *The TESOLANZ Journal, 28,* 75-88.

Dardjowidjojo, S. (2002). Bahasa asing sebagai bahasa pengantar dalam sistem pendidikan. *Bahasa dan Sastra, 22*(1), 48-65.

Guerrettaz, A. M., & Johnston, B. (2013). Materials in the classroom ecology. *The Modern Language Journal*, *97*(3), 779-796. https://doi.org/10.1111/j.1540-4781.2013.12027.x

Häcker, M. (2008). Eleven pets and 20 ways to express one's opinion: The vocabulary learners of German acquire at English secondary schools. *Language Learning Journal*, *36*(2), 215-226. https://doi.org/10.1080/09571730802393183

Hamied, A. F. (2012). English in multicultural and multilingual Indonesian education. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (pp. 69-78). Springer.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). The Range programme. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation/nation.aspx.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 23,* 403- 430.

Kirkpatrick, A. (2012). English as an international language in Asia: Implications for language education. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (pp. 29-44). Springer.

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.

Macalister, J. (2014). Teaching reading: Research into practice. *Language Teaching*, *47*(3), 387-397. https://doi.org/10.1017/S026144481400007X

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1-13. https://doi.org/10.1016/j.system.2015.04.015

Meara, P., Milton, J., & Lorenzo-Dus, N. (2001). *LEX_Manual*. University of Wales Swansea.

Mukundan, J., & Aziz, A. (2009). Loading and distribution of the 2000 high frequency words in Malaysian English language textbooks for Form 1 to Form 5. *Pertanika Journal of Social Sciences and Humanities*, *17*(2), 141-152.

Mustafa, F. (2019). English vocabulary size of Indonesian high school graduates: Curriculum expectation and reality. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *3*(2), 357-371. http://dx.doi.org/10.21093/ijeltal.v3i2.278

Nation, I. S. P., & Beglar, D. (2007). A Vocabulary Size Test. *The Language Teacher 31* (7), 9-13.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Newbury House.

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/10.1353/cml.2006.0049

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nguyen, C.-D. (2020). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal,* 1-10. https://doi.org/10.1177/0033688219895045

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298-320. https://doi.org/10.1177/1362168816639619

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, *18*(2), 161-175. https://doi.org/10.1016/S0889-4906(98)00005-2

Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4) 484-503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words unknown in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels

Test. *Language testing*, *18*(1), 55-88. https://doi.org/10.1177/026553220101800103

Sudarman, S., & Chinokul, S. (2018). The English vocabulary size and level of English Department students at Kutai Kartanegara University. *ETERNAL (English, Teaching, Learning, and Research Journal)*, *4*(1), 1-15. https://doi.org/10.24252/Eternal.V41.2018.A1

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, *93*, 1-13. https://doi.org/10.1016/j.system.2020.102279

Tang, E. (2007). An exploratory study of the English vocabulary size of Hong Kong primary and junior secondary school students. *The Journal of Asia TEFL*, *4*(1), 125-144.

Tang, E. (2009). Studying lexical input from English textbooks for primary and junior secondary students in Hong Kong. *The Journal of Asia TEFL*, *6*(3), 141-155.

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: high-, mid-, and low-frequency words. In S. Webb (Eds.), *The Routledge handbook of vocabulary studies* (pp. 81-96). Routledge.

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. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65. https://doi.org/10.1093/applin/aml048

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

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test. *ITL - International Journal of Applied Linguistics*, *168*(1), 33-69. https://doi.org/10.1075/itl.168.1.02web

West, M. (1953). *A General Service List of English Words*. Longman, Green and Co.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, *3*(2), 215-229.