



UNIVERSITY OF LEEDS

This is a repository copy of *Vocabulary in high-school EFL textbooks: Texts and learner knowledge*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/161721/>

Version: Accepted Version

Article:

Sun, Y and Dang, TNY orcid.org/0000-0002-3189-7776 (2020) Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93. 102279. ISSN 0346-251X

<https://doi.org/10.1016/j.system.2020.102279>

© 2020 Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Vocabulary in high school EFL textbooks: Texts and learner knowledge

Abstract

High school education is an important stage of foreign language education. This study examined the vocabulary in a 273,094-word corpus of high school EFL textbooks in China and measured the vocabulary knowledge of 265 high school students who used these textbooks. The corpus analysis showed that 3,000 word-families and 9,000 word-families were needed to reach 95% and 98% of the whole textbook corpus, respectively. However, vocabulary size needed for comprehension of each textbook varied greatly and did not always correspond to the textbook levels. Additionally, while the most frequent 1,000 words were fairly well-represented in the textbooks, the 2nd and 3rd most frequent 1,000 word-families were not. Scores on Webb, Sasao, and Ballance's (2017) Updated Vocabulary Levels Test revealed that the majority of the students either had mastered only the most frequent 1,000 words or had not mastered any 1,000-word levels. Pedagogical implications for high school EFL textbook writers, program managers, and teachers are discussed.

Key words: vocabulary; corpus; textbooks; high school; high frequency words, test, coverage

1. Introduction

For second language (L2) vocabulary learning to happen, learners need to be exposed to a large amount of input (Webb & Nation, 2017). However, in many English as a Foreign Language (EFL) contexts, the amount of input is fairly limited, and instructed input, especially the

language in EFL textbooks, is likely to be the major source for vocabulary learning (Alsaifa & Milton, 2012; Jordan & Gray, 2019). To optimize learning, the vocabulary in EFL textbooks should be carefully selected so that learners can understand the content of the textbooks and at the same time pay attention to the words that are most useful for them. In recognition of this need, several studies have investigated the vocabulary in EFL textbooks. Most of them focused on global textbooks produced by international publishers for L2 learners all over the world (Eldridge & Neufeld, 2009; Hsu, 2009; Matsuoka & Hirsh, 2010; O'Loughlin, 2012). Very few studies have examined the vocabulary in local high school textbooks, which are approved by the Ministry of Education in a specific country. This is surprising given that local textbooks appear to outnumber global textbooks in EFL contexts (Hughes, 2019) and high school education is an important stage of foreign language education (Zhou, 2010). Moreover, earlier studies investigating vocabulary in EFL textbooks often relied solely on the information from corpus-based analysis. Considering corpus-based information in relation to the vocabulary knowledge of students who actually use these textbooks would provide better insights into the vocabulary load of the textbooks for their users.

In recognition of these gaps, this study aims to (a) investigate the vocabulary load in a set of high school EFL textbooks used in China, (b) examine the occurrences of high frequency words in these textbooks, and (c) measure the vocabulary knowledge of students who used this set of textbooks. The high school where this study was conducted had similar features as many high schools in China, and it had typical features of EFL contexts listed by Webb and Nation (2017). The students had limited contact with English outside the classroom and their average time of learning English at school was merely around an hour each day. The EFL textbooks were selected and prescribed by the Ministry of Education and teachers had little control of the

textbooks being used. Additionally, the English language teaching was test-oriented which prepared students to attend the College Entrance Examination at the end of the high school. Those textbooks were the major sources of input in English class because they contained the words tested in the College Entrance Examination. Given the typical features of the context examined in the present study, it is expected that the study would provide useful insights into the extent to which EFL textbooks can help to facilitate the learning of the most useful words in the Chinese EFL context in particular and in many other EFL contexts in general.

2. Literature review

2.1. EFL learners' knowledge of high-frequency words

There are a huge number of words in English. Learning all of them may be challenging for most EFL learners given that they can acquire about 400 word-families per year (Webb & Chang, 2012). A word-family (*employ*) includes a base form (*employ*), its inflections (*employed, employs, employing*), and closely related derivations (*employee, employees, employer, employers, employment, employable, unemployable, employability, unemployed, unemployment*). Vocabulary research has found that most English words (e.g., *desertification, cymbal, triathlon*) appear very infrequently whereas a small number of words (3,000 word-families) (e.g., *early, keep, bring*) occur very frequently and accounted for 75%-90% of the words in a range of spoken and written discourse (Dang & Webb, 2020). The words in the former group are called mid and low-frequency words while those in the latter are high-frequency words (Nation, 2013; Schmitt & Schmitt, 2014). The larger the number of words that learners know in a text, the better they can comprehend the text (Schmitt, Jiang, & Grabe, 2011). Given the great coverage of high-frequency words in texts, knowledge of these words would enable learners to recognize a

considerable percentage of words in various kinds of discourses (e.g., movies, television programs, newspapers, and general conversation), which would then improve their comprehension quickly. For these reasons, it is essential for EFL learners to have a solid knowledge of high-frequency words before learning words at lower frequency levels (Nation, 2013; Schmitt & Schmitt, 2014; Webb & Chang, 2012).

Most studies measuring EFL learners' knowledge of high-frequency words have been carried out with university students (e.g., Dang, 2019a; Sakata, 2019; Webb & Chang, 2012). Only five studies have been done with high school students. Studies with learners in Denmark (Henriksen & Danelund, 2015; Stæhr, 2008), and Spain (Olmos, 2009) consistently reported that most students had not mastered the most frequent 2,000 words, and a reasonable number of them had not mastered even the most frequent 1,000 words. In contrast, Nguyen's (2020) study with Vietnamese EFL high-school students revealed that as a group the participants had reached the mastery level of the most frequent 2,000 words but not the most frequent 3,000 words. As learners' vocabulary knowledge may vary according to contexts, studies with high school students in other EFL contexts such as China, a country with an enormous number of EFL learners, would provide further insights into the vocabulary knowledge of high-school EFL students. Moreover, except for Nguyen (2020), earlier research with high school students used the Vocabulary Levels Test (Nation, 1990; Schmitt, Schmitt & Clapham, 2001) to measure vocabulary knowledge. This test was based on West's (1953) General Service List, which does not represent high-frequency vocabulary as well as Nation's (2012) BNC/COCA lists (Dang & Webb, 2016a; Dang, Webb, & Coxhead, 2020). A study using the Updated Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017), which was based on Nation's (2012) BNC/COCA lists, would provide a better assessment of EFL learners' knowledge of high-frequency words.

2.2. Vocabulary in EFL textbooks

There are three common approaches toward examining vocabulary in EFL textbooks. The first approach estimates the number of words needed to reach certain lexical coverage points. Lexical coverage is “the percentage of running words in the text known by the learners” (Nation, 2006, p.61). As research has found a close relationship between vocabulary knowledge and comprehension (van-Zeeland and Schmitt, 2013), examining the vocabulary in EFL textbooks from the perspective of lexical coverage is useful because it would indicate the extent to which learners can understand the textbooks. Although the level of comprehension increases according to lexical coverage (Schmitt, Jiang & Grabe, 2011), 95% is commonly used as the lexical coverage cut-off point to indicate acceptable or reasonable comprehension of texts while 98% coverage is used to indicate very good or ideal comprehension (van-Zeeland and Schmitt, 2013). Matsuoka and Hirsh (2010) found that knowledge of the most frequent 2,000 words, academic words, and other assumed known words covered 95.5% of the words in the *New Headway Student’s Book Upper-Intermediate* (Soars & Soars, 2005). Hsu’s (2009) study of university EFL textbooks revealed that the vocabulary sizes needed to reach 95% coverage of these textbooks varied from 2,500 to 13,000 word-families, and the book levels claimed by the publishers did not always correspond to the lexical demands of the textbooks. Nguyen (2020) reported that 5,000 word-families were needed to reach 95% coverage of the reading passages in the high-school EFL textbooks in Vietnam.

The second line of research on EFL textbooks counts the number of high-frequency words appearing in textbooks. This line of research is important because it helps to indicate the extent to which the words that are useful for EFL learners are represented in textbooks. Eldridge and Neufeld (2009) found that only 1,400 out of the most frequent 2,000 word-families were

represented in the *Success* coursebook series (McKinley & Hastings, 2007). Similarly, O'Loughlin (2012) reported that only 1,435 out of the most frequent 2,000 word-families appeared in the *New English File* textbooks (Oxenden & Latham-Koenig, 2006; Oxenden, Latham-Koenig, & Seligson, 2004, 2005). Alsaif and Milton (2012) found that 1,690 out of the most frequent 2,000 word-families occurred in Year 6-Year 12 EFL textbooks in Saudi Arabia.

The third line of research examined the repetition of words in textbooks. This line of research is useful because repetition is a key factor for vocabulary learning (Webb & Nation, 2017). That is, the more often a word is encountered, the more likely it is learned. Matsuoka and Hirsh (2010) examined the occurrences of West's (1953) 2nd most frequent 1,000 General Service List words in *New Headway Student's Book Upper-Intermediate* (Soars & Soars, 2005). The findings showed that 39.7% of these words did not appear in the textbooks, 20.1% occurred only once, and only 18.7% occurred 5 or more times. In contrast, 1,005 low-frequency words appeared in the textbooks, 66.4% of them occurred only once and only 7.2% occurred 5 or more times. Nguyen (2020) investigated the occurrences of words outside the most frequent 2,000 word-families in the reading texts in high school EFL textbooks in Vietnam and found that 95.85% of these words appeared only 1-5 times.

Looking beyond the EFL contexts, several studies have measured the vocabulary knowledge of high school students in English speaking countries (e.g., Coxhead, Nation, & Sim, 2015; Luxton, Fry, & Coxhead, 2017) and international school contexts (Coxhead & Boutorwick, 2018) and examined the vocabulary loads of the textbooks used in these contexts (Coxhead & Boutorwick, 2018; Coxhead, Stevens, & Tinkle, 2010; Greene & Coxhead, 2015). However, the textbooks examined in these studies were subject-specific textbooks written in English (e.g., science textbooks, English literature textbooks, and mathematic textbooks) rather than English

language learning textbooks. Given the different focuses of these studies and the present study, they were not reviewed in detail in this article.

Taken together, previous studies have provided useful information about the vocabulary in EFL textbooks. However, there are several areas that deserve attention from further research. First, none of the previous studies examined textbooks from all three perspectives: (a) vocabulary load, (b) number of high-frequency words occurring in textbooks, and (c) repetitions of words in textbooks. Second, previous studies often relied solely on the analysis of textbook corpora. Only Nguyen (2020) examined the lexical demands of textbooks in relation to the textbook users' vocabulary knowledge, but he did not focus on each grade level. Therefore, it is unclear from his study the extent to which the textbooks matched the vocabulary level of the students at a certain grade level. Importantly, Nguyen only examined vocabulary in reading texts, not all the texts in the textbooks. Third, except for Nguyen (2020), previous research used the 2,000 word-families from West's (1953) General Service List (GSL) to represent high-frequency vocabulary. Subsequent studies (Dang & Webb, 2016a; Dang, Webb, & Coxhead, 2020) have found that Nation's (2012) BNC/COCA lists better represent high-frequency words than the GSL. Additionally, Schmitt and Schmitt (2014) suggest that high-frequency words should be expanded to the 3rd 1,000-word level. Third, previous research either examined the repetition of words from the 2nd 1,000 word level or low-frequency words. It is also important to investigate the repetition of the 1st, 2nd, and 3rd 1,000 word levels because the relative value of these words for EFL learners in terms of lexical coverage decreases significantly according to the frequency levels (Webb et al., 2017). Last but not least, most previous studies focused on global textbooks, only two (Alsaif & Milton, 2012; Nguyen, 2020) examined vocabulary in local high school EFL textbooks. This is surprising since local textbooks tend to outnumber global textbooks in EFL

contexts (Hughes, 2019) and high school education is indispensable in foreign language education (Zhou, 2010).

2.3. The present study

The literature review indicates the importance of (a) investigating the vocabulary in local high school EFL textbooks from various aspects and (b) adopting the Updated Vocabulary Levels Test (Webb et al., 2017) to measure the vocabulary knowledge of students who use these textbooks. To address this need, the present study measured knowledge of high-frequency words of the students who used these textbooks. It also examined (a) the vocabulary load, (b) the number of the 1st, 2nd and 3rd most frequent words and (c) repetition of these words and words outside the most frequent 3,000 words in the high school EFL textbooks published by the Yilin Press in China. The Yilin textbooks have been widely adopted in China for EFL teaching. Compared with other high school EFL textbooks in China, the Yilin textbooks are relatively new and more catering to the current EFL teaching conditions in China (Yilin Education, 2014).

2.4. Research questions

1. To what extent do high school students using the Yilin high school EFL textbooks know the most frequent 3,000 word-families?
2. How many word-families are needed to reach 95% and 98% coverage of the Yilin high school EFL textbooks?
3. To what extent are the most frequent 3,000 word-families and words at lower frequency levels encountered in these textbooks?

This study is significant because better attention to these issues can allow textbook publishers to improve their work, allow teachers to better design their own supplemental materials, and better advice selection of outside-the-classroom learning resources.

3. Methodology

3.1. Participants

A total of 265 EFL students at a high school in the east of China participated in the study (Table 1). All participants used the Yilin textbooks in their English lessons. They were recruited on a voluntary basis. By the time the research was conducted, these students were in the middle of their second semester. As the participants had not finished their high school study at that time, their proficiency level was likely to be at Level 4 of the China's Standards of English, which is relevant to lower intermediate level (National Education Examinations Authority, 2018).

[TABLE 1 NEAR HERE]

3.2. Corpus

A textbook corpus of 273,094 words was developed from 11 senior high school EFL textbooks published by the Yilin Press in 2010. Materials in the textbooks include both the written texts in the students' books and the transcripts of listening activities. These materials were collected in either PDF or Microsoft Word format. Then, they were converted into text files so that they can be analyzed with Heatley, Nation, and Coxhead's (2002) RANGE programme. RANGE is a program which classifies the words of texts into different frequency levels of the word lists used with it. The textbook corpus has three sub-corpora responding to three grade levels at high-school: Senior 1, Senior 2, and Senior 3. The Senior 1 and Senior 2 sub-corpora each consisted of four textbooks while the Senior 3 sub-corpus was made up of three textbooks (Table 2).

[TABLE 2 NEAR HERE]

3.3. Measuring the receptive vocabulary knowledge

Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT) was administered to the Senior 1, Senior 2, and Senior 3 students in April 2019. The UVLT has five levels measuring knowledge of the words at the 1st, 2nd, 3rd, 4th and 5th most frequent 1,000-word levels. Each level consists of 30 items. Each item has three meanings and six words. Test takers have to match the meaning with the relevant words (see Figure 1). The test had a pencil-and-paper format and was delivered as part of the participants' English lessons. Instructions were given in the participants' L1 to ensure that they understood the steps of completing the test.

[FIGURE 1 NEAR HERE]

3.4. Analyzing the vocabulary in the Yilin textbook corpus

To determine the vocabulary size needed to reach 95% and 98% coverage of the textbook corpus, the corpus and its sub-corpora were run through RANGE with Nation's (2012) BNC/COCA 25 lists of 1,000 words each, which represent the 25,000 most frequently-used English words. Misspelt items in the corpus (e.g., *writting*, *weigt*, *decause*) were identified and corrected (e.g., *writing*, *weight*, *because*) so that they were listed in the relevant BNC/COCA lists. Words that did not appear in the BNC/COCA 25,000 word lists were classified by RANGE as proper nouns (baseword list 31), marginal words (baseword list 32), compounds (baseword 33), abbreviation (baseword 34), or *Not in the lists* words. A number of proper nouns (e.g., *xiaoyong*, *Shangri-la*, *Liwei*) were included in *Not in the lists*. These items were added to the list of proper nouns. The coverage of each 1,000-word level plus proper nouns and marginal words (e.g., *ah*, *phew*) were then added up together until the cumulative coverage reached the 95% and 98% cut-off points. Previous research (e.g., Dang, 2019b; Dang & Webb, 2014; Nation, 2006;

Rodgers & Webb, 2011) considered proper nouns and marginal words have little learning burden and added them to the cumulative coverage.

To examine the extent to which the most frequent 3,000 word-families and words at lower frequency encountered in the textbooks, word-families in the 1st, 2nd, and 3rd 1,000 BNC/COCA lists and words at lower 1,000 word levels were categorized into five bands based on their frequency in the textbook corpus and its sub-corpora: (a) 0-1 occurrence, (b) 7 or more occurrences, (c) 10 or more occurrences, (d) 15 or more encounters, and (e) 20 or more encounters. Although there is no frequency threshold for vocabulary learning to happen, words that are not encountered or encountered only once are unlikely to be learned, 7 or more encounters are likely to be needed for deliberate learning (Webb & Nation, 2017), 10 or more encounters are likely to be needed for incidental learning from reading (Webb, 2007; Pellicer-Sánchez & Schmitt, 2010), and 15 or more encounters are likely to be needed for incidental learning from listening (van Zeeland & Schmitt, 2013). Uchihara, Webb, and Yanagisawa's (2019) meta-analysis of research on incidental learning found that the influence of frequency on incidental vocabulary learning may remain significant up to about 20 encounters. As the textbooks examined in the present study include both spoken and written texts and aim to help students learn vocabulary both incidentally and deliberately, applying a range of frequency cut-off points would provide a better idea of the extent to which the textbooks facilitate vocabulary learning.

4. Results

4.1. The vocabulary knowledge of learners

In answer to Research Question 1 about the extent to which high-school students using the Yilin

high school EFL textbooks know the most frequent 3,000 word-families, Table 3 presents the means and standard deviations of the participants' scores on the UVLT. Irrespective of the groups, there was a decrease in the participants' scores from the 1st to the 2nd and 3rd most frequent 1,000 words. Results of a series of one-way repeated measures ANOVA revealed that these differences were always significant at $p < 0.05$. This indicates that the participants were likely to know more words at higher frequency levels than those at lower frequency levels.

[TABLE 3 NEAR HERE]

To master a level of the UVLT, test-takers need to achieve at least 29 out of 30 correct answers in that level (Webb et al., 2017). Applying this threshold, as a group, the Senior 1 students had not mastered any levels of the UVLT; Senior 2 students only mastered the most frequent 1,000 words; Senior 3 students mastered the most frequent 1,000 words and nearly mastered the 2nd most frequent 1,000 words.

The last row of Table 3 shows that of the three groups, Senior 3 students had the highest overall scores in the UVLT. Next came the Senior 2 students. Senior 1 students had the lowest overall scores. One-way between groups ANOVA with Tukey HSD post-hoc test was used to examine whether there was statistically significant difference in the overall UVLT scores of the three groups. As the assumption of homogeneity of variances was violated, Welsh test was run. The results revealed significant differences at the $p < 0.05$ level in the overall UVLT scores of the three groups: $F(2,262) = 158.11, p < 0.001$. When the scores of each group at each 1,000-word level were compared, at the 2nd and 3rd most frequent 1,000 words, Senior 1 students had a significantly lower mean score than Senior 2 students, who had a significantly lower mean score than Senior 3 students. However, at the 1st 1,000-word level, while the mean scores of the Senior 2 students and Senior 3 students were significantly higher than that of Senior 1, there was no

significant difference in the mean scores of Senior 2 and Senior 3. This may be because both Senior 2 and Senior 3 students had mastered the most frequent 1,000 words while Senior 1 students had not.

Table 4 presents the percentage of students mastering each 1,000-word level. Only about 5% of Senior 3 students (5 students) had mastered the 3rd most frequent 1,000 words and none of the Senior 1 and Senior 2 students had mastered this level. As for the 2nd most frequent 1,000 words, while more than 60% of Senior 3 students had achieved mastery of this level, only two Senior 1 students (2.15%) and one Senior 2 students (1.23%) had. In fact, most Senior 1 and Senior 2 students either only mastered the 1st 1,000-word level or had not mastered any level. It should be noted that more than 50% of the Senior 1 students and 17.28% of the Senior 2 students still had not mastered even the most frequent 1,000 words.

[TABLE 4 NEAR HERE]

4.2. The vocabulary size needed to reach 95% and 98% coverage of the Yilin textbooks

Research Question 2 asks about the number of word-families needed to reach 95% and 98% coverage of the Yilin high school EFL textbooks. Together with proper nouns and marginal words, 3,000 word-families and 9,000 word-families are needed to reach 95% and 98% coverage of the whole corpus (see the second column of Table 5). As for textbooks in each grade, the next three columns of Table 5 show that 3,000 word-families plus proper nouns and marginal words are also needed to reach 95% coverage. However, the vocabulary size needed to reach 98% coverage of textbooks at each grade varied: 9,000 word-families (Senior 1), 11,000 word-families (Senior 2), and 8,000 word-families (Senior 3).

[TABLE 5 NEAR HERE]

A similar pattern was seen when the spoken and written components of the textbooks were analyzed separately. Table 6 shows that irrespective of the types of text and grades, 3,000 word-families plus proper nouns and marginal words were always needed to achieve 95% coverage. However, spoken texts always required a larger vocabulary size to reach 98% coverage than written texts. Also, no matter whether the spoken or written component was examined, Senior 1 textbooks always required a larger vocabulary size to reach 98% coverage than Senior 2 and Senior 3 textbooks.

[TABLE 6 NEAR HERE]

Similarly, there is a variation in the vocabulary size needed to reach 95% and 98% of each textbook (see the Appendix). While most textbooks required 3,000 word-families to reach the 95% coverage, Yilin 2 and Yilin 7 required a larger vocabulary size (4,000 word-families). Moreover, the vocabulary size needed to reach 98% of each textbook ranged from 5,000 word-families (Yilin 4) to 6,000 word-families (Yilin 5, Yilin 9), 8,000 word-families (Yilin 3), 9,000 word-families (Yilin 1, Yilin 11), 10,000 word-families (Yilin 7) and more than 25,000 word-families (Yilin 2, Yilin 6, Yilin 8, Yilin 10). It is important to note that no matter whether the spoken component and the written component was examined together or separately, the vocabulary sizes needed for comprehension of these textbooks did not increase according to the grade level or sequence of textbooks but fairly random.

4.3. Occurrences of the most frequent 3,000 words and words at lower frequency levels in the textbook corpus

In answer to Research Question 3 about the extent to which the most frequent 3,000 word-families and words at lower frequency levels encountered in these textbooks, the first column of

Table 7 shows that more than 98% of the most frequent 1,000 words appeared in the textbook corpus while 86.7% of the 2nd most frequent 1,000 words and 62.8% of the 3rd most frequent 1,000 words occurred in the textbook corpus. It means that 505 items from the 2nd and 3rd most frequent 1,000 words (e.g., *illustrate, somewhat, soup, contract, enormous, establish, capture, case, crucial, demonstrate, essential, foundation, fulfill, overall, proportion, significant*) were absent from the textbook corpus. In contrast, the whole corpus contained 1,008 words that are outside the most frequent 3,000 word-families. A reasonable number of them are words at very low frequency level such as *desertification, cymbal, triathlon, and confetti* (12th 1,000 word level), *concubine* (13th 1,000 word level), *betide* (16th 1,000 word level), and *salicyclic* (17th 1,000 word level).

A similar trend was seen when the vocabulary in textbooks at each grade level was examined. As shown in the last three columns of Table 7, around 93% of the most frequent 1,000 words appeared in Senior 1, Senior 2 and Senior 3 textbooks whereas the percentage of the 2nd and 3rd most frequent 1,000 words was just about 58%-71% and 30%-47%, respectively. Textbooks at each grade level also contained from 342 to 585 words outside the most frequent 3,000 words.

[TABLE 7 NEAR HERE]

Let us consider the percentage of the most frequent 3,000 words and words at lower frequency levels that Senior 1, Senior 2, and Senior 3 students would re-encounter in the textbooks. As shown in the first three columns of Table 8, the percentage of the 1st 1,000 words that students did not encounter or encountered only once in the textbooks decreased according to the grade levels. By the end of the Senior 1 level, students had completed the Senior 1 textbooks; therefore, they encountered 9.3% of the 1st 1,000 words for 0-1 time. However, by the end of

Senior 2 level, students had completed the Senior 1 and Senior 2 textbooks. As a result, the percentage of the 1st 1,000 words encountered 0-1 time went down to 3.1%. By the end of the Senior 3 level, students had completed all Senior 1, 2 and 3 textbooks. This helped to reduce the percentage of the 1st 1,000 words encountered for 0-1 time to only 1.6%. As the percentage of the 1st 1,000 words encountered 0-1 time decreased, the percentage of the 1st 1,000 words encountered multiple times increased steadily as the students progress further in their study. The percentage of word encountered 7 or more times rose from 73.4% (Senior 1) to 87.7% (Senior 2) and 92.5% (Senior 3). Across the three grade levels, there was also an increase in the percentage of words encountered 10 times or more (from 66.2% to 89.6%), 15 times or more (53.3% to 84%) and 20 times or more (47.3% to 75.9%).

As for the 2nd and 3rd 1,000 most frequent words, the next six columns in Table 8 show that there was an increase in the percentage of word-families encountered multiple times. Despite this fact, by the end of Senior 3, the percentage of 2nd and 3rd most frequent 1,000 words that students would either encounter once or not encounter at all was still 15.5% and 41.1%, respectively. Additionally, only 60.6% of the 2nd most frequent 1,000 words and 36.5% of the 3rd most frequent 1,000 words would be encountered 7 or more times. The percentages of the 2nd and 3rd most frequent 1,000 words that occurred 10 or more times (50.2%, 27%), 15 or more times (37.5%, 18.8%) and 20 or more times (29.6%, 13.2%) were even lower. The last three columns of Table 8 show that the textbooks contained a substantial number of words outside the most frequent 3,000 word-families. By the end of Senior 3, students would encounter about 17% of these words once, 29% of the words 7 or more times, 18% ten or more times, nearly 10% fifteen or more times, and about 6% twenty or more times.

[TABLE 8 NEAR HERE]

5. Discussion

Textbooks play a significant role in the English language learning and teaching in EFL contexts because they help to bring structure for programs, standardize instructions, and save teachers' time (Richards, 2001; Jordan & Gray, 2019; Hughes, 2019). However, textbooks also have certain limitations (Richards, 2001; Jordan & Gray, 2019). Drawing on the information from corpora and learners, the present study examined the vocabulary in local high school EFL textbooks from multiple perspectives: vocabulary load, the occurrences of high-frequency words, and repetitions of words in textbooks. Therefore, it provides solid evidence to support the claim about the strengths and limitations of EFL textbooks. It revealed that the Yilin textbooks represented the most frequent 1,000 words fairly well. Once completing the whole set of Senior 1, Senior 2, and Senior 3 textbooks, students would encounter 98% of the most frequent 1,000 words. In fact, they would encounter nearly 93% of these words 7 or more times, 90% of these words 10 or more times, 84% fifteen or more times, and nearly 76% twenty or more times. The large number of the most 1,000 words appearing in the textbooks and the high percentage of these words occurring multiple times would create a good condition for the learning of the most frequent 1,000 words. This is really meaningful because these words have the greatest value for EFL learners (Dang & Webb, 2016a, 2016b). The good representation of the most frequent 1,000 words in the Yilin textbooks also helps to explain why as a whole group, the Senior 1 students had not mastered the most frequent 1,000 words but the Senior 2 and Senior 3 students had.

While Yilin textbooks seem to create good conditions for the learning of the most frequent 1,000 words, they have certain problems. First, the corpus-based analysis indicated that the participants would need a vocabulary size of 3,000 word-families to reach 95% coverage of the textbooks, which indicates acceptable or reasonable comprehension. To achieve a higher

level of comprehension, they would need larger vocabulary sizes (8,000-11,000 word-families). However, the UVLT scores revealed that very few students (5 out of 265 students) had mastered the most frequent 3,000 word-families. In fact, the majority of them either had mastered only the most frequent 1,000 word-families or had not mastered any 1,000-word levels. It means that regardless of the book levels, the Yilin textbooks may contain about 2,000-3,000 words families that are new to the majority of the participants. In other words, the participants are expected to cope with about 2,000-3,000 new word-families per year. This number is much larger than the number of words that EFL learners are likely to learn per year (400 word-families) (Webb & Chang, 2012), which suggests that the Yilin textbooks may be too demanding for the students in terms of vocabulary load.

Second, while the debate about authenticity in EFL textbooks is beyond the scope of this article, a common consensus is that the language presented in the textbooks should reflect the language used in natural communicative situations to some extent so that they can prepare learners for real communication (Clavel-Arroitia & Fuster-Márquez, 2014; Römer, 2004). However, the present study found that the spoken texts in the Yilin textbooks required a larger vocabulary size to reach the 98% coverage than the written texts. This contrasts with the findings of previous studies which examined the occurrences of words in various kinds of spoken discourse (e.g., general conversation, television programs, movies, songs, academic speech) and written discourse (e.g. novels, newspapers, academic writing) (Dang & Webb, 2014, 2020; Nation, 2006). These studies consistently showed that a smaller vocabulary size is needed to reach 98% coverage of spoken texts than written texts. It means in natural communicative situations, spoken texts are always less demanding than written texts in terms of lexical coverage.

Third, for learning to happen, tasks and materials should be slightly beyond students' current language proficiency level but should not be too hard, otherwise they may demotivate students and hinder learning (Krashen, 1981; Vygotsky, 1978). In other words, to scaffold learners' vocabulary learning, the vocabulary loads of school textbooks should "differ systematically between grade levels and between higher and lower tracks" (Meurers, Bryant, Wagner, Chinkina, & Trautwein, 2018, p.518). That is, the Senior 1 textbooks should be less demanding than the Senior 2 textbooks, which should be less demanding than the Senior 3 textbooks in terms of lexical coverage. Likewise, the vocabulary size needed to reach 95% and 98% coverage of each textbook should increase gradually from Yilin 1 to Yilin 11. However, regardless of the set of textbooks, 3,000 word-families were required to reach the 95% coverage. To achieve 98%, larger vocabulary sizes were needed in the case of Senior 1 textbooks (9,000 word-families) and Senior 2 textbooks (11,000 word-families) than in the case of Senior 3 textbooks (8,000 word-families). Analysis of the vocabulary load of each textbook also showed that the vocabulary size needed for comprehension of each textbook is fairly random.

Fourth, although vocabulary researchers (Nation, 2013; Schmitt & Schmitt, 2014; Webb & Chang, 2012) have suggested that knowledge of the most frequent 3,000 words is crucial for EFL learners, the present study found that the 2nd and 3rd 1,000 word-families were not well represented in the textbooks. As revealed in the corpus analysis, by the end of Senior 3, students had completed the whole sets of Senior 1, Senior 2, and Senior 3 textbooks. Yet there were still 505 word-families from the 2nd and 3rd most frequent 1,000 word-families that they did not encounter. In fact, 16% of the 2nd most frequent 1,000 words and 41% of the 3rd most frequent 1,000 words were either never encountered by the students in the textbooks or encountered only once. The percentage of the 2nd and 3rd most frequent 1,000 words encountered multiple times by

the students were fairly modest: 60% and 37% (7 or more times), 50%-27% (10 or more times), 38%-19% (15 or more times), 30%-13% (20 times or more). In contrast, the textbooks included a considerable number of lower frequency words (1,008 word-families). 17% of these words only occurred once. Although a number of low-frequency words occurred multiple times, which may create conditions for learning, given the students' insufficient knowledge of the 2nd and 3rd 1,000 words, the learning time should be spent on the 2nd and 3rd 1,000 words rather than words at lower frequency level. Moreover, presenting too many low-frequency words while students had not mastered the most frequent 3,000 words would distract them from the most frequent 3,000 words and at the same time create learning burdens and hinder students' further vocabulary development (Nguyen, 2020). The poor representation of the 2nd and 3rd 1,000 word-families in the Yilin textbooks helps to explain why most participants, even Senior 3 students, had not mastered the most frequent 3,000 words. The findings of this study are in line with research in other contexts which reported that high-frequency words were not well represented in global EFL textbooks (Eldridge & Neufeld, 2009; O'Loughlin, 2012) and high-school EFL textbooks (Alsaif & Milton, 2012) and that the levels of university EFL textbooks did not always coincide with the lexical demands of the textbooks (Hsu, 2009).

The present study found that the vocabulary knowledge of the participants grows grade by grade, which suggests that instruction may have some impact on the vocabulary growth. Despite this fact, after 9 years of formal English instruction, most Senior 3 students had mastered either the most frequent 2,000 words or the most frequent 1,000 words. The most concerning point is that several students still had insufficient knowledge of the most frequent 1,000 words. The findings of the present study are consistent with previous studies with high-school students in Denmark (Henriksen & Danelund, 2015; Stæhr, 2008) and Spain (Olmos, 2009) as well as

those with university EFL students (e.g., Dang, 2019a; Webb & Chang, 2012). However, the findings of the present study are slightly different from Nguyen's (2020) findings with Vietnamese high school EFL learners. One possible reason for this difference is that Nguyen did not report the percentage of individual learners who had mastered each 1,000-word level. As the vocabulary level varies according to individual (Milton, 2009), there are chances that as the whole group Nguyen's participants had reached the mastery level of the most frequent 2,000 words, but there were still a number of students having not mastered these words.

Although various factors may lead to EFL learners' insufficient knowledge of the most frequent 3,000 words, one possible reason may be the design of EFL textbooks in terms of vocabulary. As shown in the case of the Yilin textbooks, the 2nd and 3rd most frequent 1,000 words were not well represented in the textbooks. Additionally, the textbooks appear to be too demanding for most of the participants in terms of lexical coverage, and there was a lack of consistency between the lexical demands of textbooks and the book levels. Sakata, Tagashira, and Mochizuki (2014, as cited in Sakata, 2019, p. 2) found that EFL learners were likely to learn words occurring frequently in textbooks irrespective of their frequency in natural language use whereas they were less likely to learn words which have high-frequency in natural language use but do not appear often in textbooks. Dang and Webb (2020) found that textbooks are among the factors that have the strongest influence on EFL teachers' selection of words for instruction. Considering the significant role of textbooks in EFL learning and teaching, it is fair to say that the textbook is probably one of the major contributors to EFL learners' insufficient knowledge of the most frequent 3,000 words.

The present study has several limitations. First, it examined the textbooks and learners in a specific context. Although these textbooks and learners share many features of those in other

EFL contexts, caution should be taken when generalizing the findings of this study. Second, this study only examined single words while knowledge of multi-words is also important for EFL learners. Further research with textbooks and learners in other EFL contexts which investigates both single words and multi-words would provide further insights into the vocabulary in high school EFL textbooks.

6. Pedagogical implications

6.1. Recommendations for textbook writers and English language program managers/designers

Yilin textbook writers in particular and EFL textbook writers in general should be mindful when selecting the vocabulary to include in textbooks for high school students to optimize the learning of the most frequent 3,000 words.

First, the most frequent 3,000 words are crucial for EFL learners to perform various kinds of spoken and written tasks. However, the present study shows that the Yilin textbooks did not well represent the 2nd and 3rd most frequent 1,000 words whereas they included a considerable number of low-frequency words. Therefore, Yilin textbook writers in particular and EFL high school textbook writers in general should try to minimize the number of words outside the most frequent 3,000 words and maximize the number of the most frequent 3,000 words and their repetitions in textbooks for high school students. They could analyze the materials with VocabProfile (Cobb, n.d), an online corpus-based tool which has a similar function as RANGE (see Dang (in press) for step-by-step instructions on how to use VocabProfile to analyze texts). If those texts are too lexically demanding and insufficiently represent the most frequent 3,000 words, textbook writers should discard these texts. Another option is to adapt these texts by replacing the low-frequent words with their synonyms/antonyms that are the most frequent 3,000

words (see Webb and Nation (2008) for further information about how to adapt vocabulary in materials).

Second, while the present study suggests that high-school textbooks should offer a better representation of the most frequent 3000 words, it does not mean that all of these words should be included in a single textbook. In fact, it would be challenging to include all 3rd 1,000 words in a textbook given that by nature they do not occur as frequently as the most frequent 2,000 words. Moreover, research has shown that spaced learning is likely to result in better retention than massed learning (Webb & Nation, 2017). Therefore, while textbooks at all levels should focus on the most frequent 3,000 words, a systematic approach should be taken by textbook writers when determining which high-frequency words to include in each textbook. Textbooks at the early grade should include a large proportion of the 1st 1,000 words and a small proportion of the 2nd and 3rd 1,000 words to draw learners' attention to the most frequent 1,000 words. As learners progress further, the proportion of the 1st 1000 words in textbooks could decrease gradually while the proportion of the 2nd and 3rd 1000 words should increase gradually. This sequence would help learners learn the most frequent words first and that knowledge would support and motivate learners when learning less frequent words (Dang & Webb, 2016a; 2016b). That said, it is not desirable to include all of the most frequent 3,000 words in a single textbook; however, the whole set of textbooks should offer students multiple exposures to most (if not all) of the most frequent 3,000 words given the importance of these words for communication.

As for English language program managers/designers, they should not be biased toward the book levels claimed by publishers and should be more cautious when selecting textbooks for students in their programs.

6.2. Recommendations for EFL teachers

We acknowledge that in many EFL contexts such as that in the present study, teachers have little control of the textbooks being used. Therefore, when waiting for changes from textbook writers and program managers/designers, it is important for teachers to be aware of the limitations of the textbooks. Rather than simply using the prescribed textbooks, they should critically evaluate the vocabulary in textbooks to see if it matches their students' levels. They may use the UVLT to diagnose their students' knowledge of the most frequent 3,000 words at the beginning of each grade level. Also, they should analyze the vocabulary in the textbooks to see if the textbooks suit their students' levels. If not, they should design supplementary materials and activities to bridge the gap between the vocabulary load of the textbooks and their students' current vocabulary knowledge. Teachers using the Yilin textbooks can refer to the Appendix for the information about the vocabulary load of each textbook. Those using other textbooks can follow Dang's (in press) instructions to analyze their teaching materials with VocabProfile. Additionally, teachers can consult Nation (2007), Coxhead (2014), and Webb and Nation (2017) for ideas about designing materials and activities to better support learners' vocabulary development.

Apart from helping learners learn vocabulary at school, teachers should encourage them to make good use of graded readers and English language television programs outside the classroom. These are excellent resources for meaning-focused input for EFL learners (Dang, 2019b; Nation & Waring, 2019; Peters, Heynen, & Puimege, 2016; Rodgers & Webb, 2011). Graded readers are relevant resources for students with the vocabulary knowledge lower than the most frequent 3,000 words. These materials are specifically designed to provide repeated encounters with the most frequent 3,000 words and thus provide these students with a great deal

of opportunities to develop their knowledge of these words (see Nation and Waring, 2019 for detailed instruction on how to organize an intensive reading program). Once students have mastered the most frequent 3,000 words, they could make use of English language television programs. Extensive viewing television programs would allow them to be exposed to a large amount of authentic input. It also helps them to consolidate and expand their knowledge of the most frequent 3,000 words and learn words at lower frequency levels (see Webb, 2015 for further discussion of how to organize extensive viewing activities).

7. Conclusion

This study investigated the vocabulary load of the EFL textbooks and the vocabulary knowledge of students using these textbooks in a high school in China. The results revealed that most students had insufficient knowledge of the most frequent 3,000 words. Given the students' vocabulary level, the textbooks appear to be too demanding for them in terms of lexical coverage. This study also found that the 2nd and 3rd most frequent 1,000 words were not well represented in the textbooks, which indicated that these textbooks may provide little help with facilitating these students' development of these words. As the specific context examined in this study has similar features as high schools in many other EFL contexts, this study provides useful information for those working in similar contexts. It suggests that textbook writers should carefully check the vocabulary loads of texts when including them in textbooks, and program managers should take more caution when selecting textbooks for their students. Meanwhile, teachers should critically evaluate the suitability of prescribed textbooks for their students and provide them with supplementary materials to better facilitate their vocabulary development.

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/10.1080/09571736.2012.658221>
- Clavel-Arroitia, B., & Fuster-Márquez, M. (2014). The authenticity of real texts in advanced English language textbooks. *ELT Journal*, 68(2), 124–134. <https://doi.org/10.1093/elt/cct060>
- Cobb, T. (n.d.). *Lextutor*. Retrieved from <http://www.lexutor.ca>
- Coxhead, A. (2014). *New ways in teaching vocabulary, revised*. Alexandria, Virginia: TESOL Inc.
- Coxhead, A, Nation, I. S. P., & Sim, D. (2015). Vocabulary size and native speaker secondary school students. *New Zealand Journal of Educational Studies*. <https://doi.org/10.1007/s40841-015-0002-3>
- Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, Maths and Science. *TESOL Quarterly*, 52(3), 588–610. <https://doi.org/10.1002/tesq.450>
- Coxhead, A., Stevens, L., & Tinkle, J. (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics*, 16(2), 37–52.
- Dang, T. N. Y. (2019a). High-frequency words in academic spoken English: Corpora and learners. *ELT Journal*. <https://doi.org/10.1093/elt/ccz057>.

Dang, T. N. Y. (2019b). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly*. <https://doi.org/10.1002/tesq.552>.

Dang, T. N. Y. (in press). Selecting and adapting texts for listening and reading activities. In V. Viana (Ed.), *New ways in teaching with corpora*. Alexandria: TESOL Press.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>.

Dang, T. N. Y., & Webb, S. (2016a). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>.

Dang, T. N. Y., & Webb, S. (2016b). Making an essential word list for beginners. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.

Dang, T. N. Y., & Webb, S. (2020). Vocabulary instruction and the good language teachers. In C. Griffiths, Z. Tajeddin, & A. Brown (Eds.), *Lessons from good language teachers* (pp. 203–218). Cambridge: Cambridge University Press.

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020.). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*. <https://doi.org/10.1177/1362168820911189>

Eldridge, J., & Neufeld, S. (2009). The graded reader is dead, long live the electronic reader. *The Reading Matrix*, 9(2), 224–244.

Greene, J. W., & Coxhead, A. (2015). *Academic vocabulary for middle school students: Research-based lists and strategies for key content areas*. Baltimore: Brookes.

Heatley, A., Nation, I. S. P. and Coxhead, A. (2002). *Range computer programme*. Retrieved from: <https://www.victoria.ac.nz/lals/about/staff/paul-nation>

Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In P. Pietilä, K. Doró, & R. Pipalová (Eds.), *Lexical issues in L2 writing* (pp. 1–27). Newcastle upon Tyne: Cambridge Scholars Publishing.

Hughes, S. H. (2019). Coursebooks: Is there more than meets the eye? *ELT Journal*, 73(4), 447–455. <https://doi.org/10.1093/elt/ccz040>

Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage. *Electronic Journal of Foreign Language Teaching*, 6(1), 42–62.

Jordan, G., & Gray, H. (2019). We need to talk about coursebooks. *ELT Journal*, 73(4), 438–446. <https://doi.org/10.1093/elt/ccz038>

Krashen, S. (1981) *SLA and Second Language Learning*. Oxford: Pergamon

Luxton, J., Fry, J., & Coxhead, A. (2017). Exploring the knowledge and development of academic English vocabulary of students in NZ secondary schools. *SET*, 17(1), 12–22. <https://doi.org/10.18296/set.0071>

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22(1), 56–70.

McKinley, S., & Hastings, B. (2007). *Longman English Success*. New Jersey: Pearson Longman

Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology, 110*(4), 518–543. <https://doi.org/10.1037/edu0000225>

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: 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/10.3138/cmlr.63.1.59>

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 1-12. <https://doi.org/10.2167/illt039.0>

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nation, I. S. P., & Waring, R. (2019). *Teaching extensive reading in another language*. London: Routledge.

- National Education Examinations Authority. (2018). China's standards of English. [Online]. [Accessed 15 July]. Available from: <http://www.neea.edu.cn/html1/report/19011/3260-1.htm>
- Nguyen, C. D. (2020). Lexical features of reading passages in English textbooks for Vietnamese high-school students: Do they foster both content and vocabulary knowledge? *RELC*. <https://doi.org/10.1177/0033688219895045>
- Olmos, C. (2009). An assessment of the vocabulary knowledge of students in the final year of secondary education. Is their vocabulary extensive enough? *International Journal of English Studies, Special Issue*, 73–90.
- O'Loughlin, R. (2012). Tuning into vocabulary frequency in coursebooks. *RELC Journal*, 43(2), 255–269. <https://doi.org/10.1177/0033688212450640>
- Oxenden C, Latham-Koenig C, and Seligson P (2004a) *New English File (Elementary Student's Book)*. Oxford: Oxford University Press.
- Oxenden C, Latham-Koenig C, and Seligson P (2005) *New English File (Pre-intermediate Student's Book)*. Oxford: Oxford University Press.
- Oxenden C, Latham-Koenig C (2006). *New English File (Intermediate Student's Book)*. Oxford: Oxford University Press.
- Pellicer-Sánchez, A. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22(1), 31-55.

Peters, E., Heynen, E., & Puimege, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148.

<https://doi.org/10.1016/j.system.2016.10.002>

Richards, J. C. (2001). The role of textbooks in a language program. *RELC Guidelines*, 23(2), 12–16.

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*. <https://doi.org/https://doi.org/10.5054/tq.2011.268062>

Römer, U. (2004). Comparing real and idea language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 152–193). Amsterdam: John Benjamins.

Sakata, N. (2019). Profiling vocabulary for proficiency development: Effects of input and general frequency on L2 learning. *System*, 87, 1–12.

<https://doi.org/10.1016/j.system.2019.102167>

Sakata, N., Tagashira, K., & Mochizuki, M. (2014, December). Vocabulary in high school textbooks and its influence on students. In Paper presented at the JACET Vocabulary SIG Annual Conference, Tokyo, Japan.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(i), 26–43.

<https://doi.org/10.1111/j.1540-4781.2011.01146.x>

Schmitt, N., & Schmitt, D. (2014). 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., 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>

Soars, L., & Soars, J. (2005). *New headway student's book (Upper–Intermediate)*. Oxford: Oxford University Press.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>

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>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

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. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). New York, NY: Routledge.

Webb, S. & Nation, I. S. P (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1-10.

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, I. S. P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

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: Longman, Green.

Yilin Education. (2014). *About us-Bureau of basic education*. Retrieved from: <http://edu.yilin.com>

Zhou, Y. (2010). English language learning strategy use by Chinese senior high school students. *English Language Teaching*, 3(4), 152–158.

Tables

Table 1. Information of the participants in the present study (N=265)

Grade*	Number of participants	Age	Number of years of English study	Textbooks to be used in each grade
Senior 1	91	16	7	Yilin 1-4
Senior 2	81	17	8	Yilin 5-8
Senior 3	93	18	9	Yilin 9-11

* *In the Chinese educational system, high school has three years. Senior 1, Senior 2, and Senior 3 students are first year, second year, and third year high school students, respectively.*

Table 2. Component of the Yilin textbook corpus (273,094 words)

Senior 1 (85,037 words)		Senior 2 (103,462 words)		Senior 3 (84,595 words)	
Textbook	Number of words	Textbook	Number of words	Textbook	Number of words
Yilin 1	19,949	Yilin 5	20,745	Yilin 9	27,868
Yilin 2	19,599	Yilin 6	27,438	Yilin 10	29,027
Yilin 3	23,545	Yilin 7	27,860	Yilin 11	27,700
Yilin 4	21,944	Yilin 8	27,419		

Table 3. Mean UVLT scores with standard deviations of each group of participants

UVLT level	Senior 1 (n=91)	Senior 2 (n=81)	Senior 3 (n=93)
1 st most frequent 1,000 words	26.04 (6.82)	29.20 (2.08)	29.77 (0.59)
2 nd most frequent 1,000 words	16.19 (7.69)	22.94 (3.81)	28.38 (2.64)
3 rd most frequent 1,000 words	8.93 (6.09)	16.96 (4.90)	23.90 (3.59)
Overall	51.16 (17.25)	69.10 (9.15)	82.05 (5.84)

*Standard deviations in parentheses

Table 4. The number and percentage of students in each group mastering each level of the UVLT

Groups	Yet to master the 1 st 1,000 word level	Mastered the 1 st 1,000 word level	Mastered the 2 nd 1,000 word level	Mastered the 3 rd 1,000 word level
Senior 1 (n =91)	49 (53.85 %)	40 (43.96%)	2 (2.20%)	0 (0%)
Senior 2 (n =81)	14 (17.28%)	66 (81.48%)	1(1.23%)	0 (0%)
Senior 3 (n = 93)	2 (2.15%)	29 (31.18%)	57 (61.29%)	5 (5.38%)

*the percentage of students in parentheses

Table 5. Lexical coverage of the textbook corpus

Vocabulary level (+proper nouns & marginal words)	Whole corpus	Senior 1	Senior 2	Senior 3
1,000	84.28	86.0	84.06	82.81
2,000	92.18	93.19	91.57	91.9
3,000	95.54	95.86	95.32	95.47
4,000	96.63	96.8	96.47	96.65
5,000	97.36	97.49	97.22	97.4
6,000	97.62	97.73	97.46	97.71
7,000	97.82	97.86	97.64	97.99
8,000	97.93	97.95	97.76	98.11
9,000	98.02	98.05	97.87	98.17
10,000	98.1	98.14	97.96	98.24
11,000	98.15	98.17	98.04	98.27
12,000	98.18	98.22	98.09	98.28
13,000	98.23	98.23	98.12	98.38
14,000	98.24	98.24	98.13	98.4
15,000	98.25	98.27	98.13	98.4
16,000	98.27	98.28	98.14	98.47
17,000	98.27	98.28	98.14	98.47
18,000	98.27	98.28	98.14	98.47
19,000	98.27	98.28	98.15	98.47
20,000	98.28	98.28	98.17	98.47
21,000	98.28	98.28	98.17	98.47
22,000	98.28	98.28	98.17	98.48
23,000	98.28	98.28	98.17	98.48
24,000	98.28	98.28	98.17	98.48
25,000	98.28	98.28	98.17	98.48

Table 6. Lexical coverage of spoken and written components of the textbook corpus

Vocabulary level (+Proper nouns & Marginal words)	Whole corpus		Senior 1		Senior 2		Senior 3	
	spoken texts	written texts	spoken texts	written texts	spoken texts	written texts	spoken texts	written texts
1,000	85.9	83.81	87.65	84.99	85.52	84.12	84.91	82.37
2,000	92.99	92.14	93.88	92.52	92.52	91.8	92.87	92.12
3,000	95.85	95.75	95.73	95.52	95.90	95.78	95.87	95.89
4,000	96.93	96.8	96.51	96.45	97.13	96.87	97.00	97.02
5,000	97.51	97.56	97.03	97.19	97.65	97.73	97.72	97.68
6,000	97.68	97.85	97.18	97.45	97.79	98.02	97.96	98.01
7,000	97.79	98.07	97.20	97.63	97.92	98.24	98.11	98.27
8,000	97.86	98.19	97.26	97.73	98.00	98.37	98.18	98.41
9,000	97.96	98.27	97.38	97.81	98.13	98.40	98.22	98.48
10,000	97.99	98.37	97.41	97.93	98.15	98.52	98.25	98.55
11,000	98.04	98.42	97.43	97.97	98.25	98.60	98.26	98.59
12,000	98.07	98.45	97.51	98.01	98.28	98.65	98.26	98.60
13,000	98.11	98.49	97.51	98.02	98.28	98.70	98.38	98.67
14,000	98.12	98.5	97.52	98.03	98.28	98.71	98.40	98.68
15,000	98.13	98.51	97.56	98.06	98.28	98.71	98.40	98.68
16,000	98.16	98.53	97.56	98.07	98.29	98.71	98.49	98.73
17,000	98.16	98.53	97.56	98.07	98.29	98.72	98.49	98.73
18,000	98.16	98.53	97.56	98.07	98.29	98.73	98.49	98.73
19,000	98.17	98.53	97.56	98.07	98.31	98.74	98.49	98.73
20,000	98.18	98.54	97.56	98.07	98.33	98.75	98.49	98.73
21,000	98.18	98.54	97.56	98.07	98.33	98.75	98.49	98.73
22,000	98.18	98.54	97.56	98.07	98.33	98.75	98.5	98.74
23,000	98.18	98.54	97.57	98.07	98.33	98.75	98.5	98.74
24,000	98.18	98.54	97.57	98.07	98.34	98.75	98.5	98.74
25,000	98.18	98.54	97.57	98.07	98.34	98.75	98.5	98.74

Table 7. Number and percentage of word-families appearing in the high school EFL textbook corpus

BNC/COCA word level	Whole corpus	Senior 1	Senior 2	Senior 3
1 st 1000	985 (98.5%)	927 (92.7%)	935 (93.5%)	933 (93.3%)
2 nd 1000	867 (86.7%)	583 (58.3%)	712 (71.2%)	687 (68.7%)
3 rd 1000	628 (62.8%)	304 (30.4%)	447 (44.7%)	465 (46.5%)
Outsides the most frequent 3,000 words	1,008	342	505	585

*the percentage in parentheses

Table 8. Percentage of the most frequent 3,000 words and words at lower frequency levels re-encountered by Senior 1, Senior 2, and Senior 3 students in the textbooks (%)

Number of occurrences	1 st 1,000 words			2 nd 1,000 words			3 rd 1,000 words			Outside the most frequent 3000 words		
	Senior	Senior	Senior	Senior	Senior	Senior	Senior	Senior	Senior	Senior	Senior	Senior
	1	2	3	1	2	3	1	2	3	1	2	3
0-1	9.3	3.1	1.6	47.2	26.2	15.5	73.7	51.5	41.1	26.90*	20.51*	16.77*
7 or more	73.4	87.7	92.5	25.8	48.7	60.6	8.4	26.7	36.5	24.56	26.17	29.27
10 or more	66.2	83.4	89.6	18.6	37.9	50.2	5.9	18.6	27.0	14.04	16.27	18.06
15 or more	53.3	74.4	84.0	11.5	26.6	37.5	3.3	11.8	18.8	9.94	10.33	9.62
20 or more	47.3	66.3	75.9	7.5	20.3	29.6	2.0	8.9	13.2	5.85	7.36	6.45

*these words occurred once in the corpus

Figure

Figure 1. Example of an item from the Update Vocabulary Levels Test (Webb et al., 2017)

	box	brother	horse	hour	house	plan
family member		✓				
sixty minutes				✓		
Way of doing things						✓

Appendix. Lexical coverage of each textbook

Vocabulary level (+proper nouns & marginal words)	Yilin 1	Yilin 2	Yilin 3	Yilin 4	Yilin 5	Yilin 6	Yilin 7	Yilin 8	Yilin 9	Yilin 10	Yilin 11
1,000	86	84.92	85.09	85.16	84.81	85.29	82.61	83.72	82.58	81.85	84.03
2,000	93.19	91.89	93.03	93.11	92.6	92.4	90.98	90.55	91.17	91.63	92.9
3,000	95.86	94.5	95.74	96.42	95.64	95.64	94.95	95.11	95.18	95.45	95.77
4,000	96.8	95.2	96.81	97.45	97.13	96.28	96.3	96.31	96.71	96.49	96.75
5,000	97.49	95.92	97.56	98.14	97.82	97.18	97.05	96.97	97.89	97.08	97.23
6,000	97.73	96.19	97.8	98.29	98.09	97.43	97.18	97.32	98.3	97.22	97.62
7,000	97.86	96.37	97.98	98.39	98.21	97.59	97.49	97.44	98.78	97.37	97.84
8,000	97.95	96.51	98.04	98.46	98.27	97.71	97.58	97.62	98.88	97.47	97.99
9,000	98.05	96.55	98.13	98.6	98.27	97.73	97.96	97.64	98.94	97.52	98.05
10,000	98.14	96.62	98.21	98.74	98.46	97.82	98.02	97.67	99.04	97.58	98.09
11,000	98.17	96.64	98.22	98.81	98.54	97.85	98.03	97.87	99.06	97.61	98.14
12,000	98.22	96.65	98.24	98.89	98.61	97.86	98.11	97.90	99.06	97.61	98.16
13,000	98.23	96.65	98.25	98.9	98.66	97.87	98.18	97.90	99.29	97.67	98.16
14,000	98.24	96.67	98.26	98.92	98.66	97.87	98.19	97.92	99.33	97.67	98.17
15,000	98.27	96.78	98.26	98.94	98.66	97.87	98.19	97.93	99.33	97.67	98.17
16,000	98.28	96.78	98.27	98.94	98.66	97.87	98.2	97.94	99.37	97.83	98.17
17,000	98.28	96.78	98.27	98.94	98.67	97.87	98.21	97.94	99.37	97.83	98.17
18,000	98.28	96.78	98.27	98.94	98.68	97.87	98.21	97.94	99.37	97.83	98.17
19,000	98.28	96.78	98.27	98.94	98.68	97.88	98.21	97.96	99.38	97.83	98.17
20,000	98.28	96.78	98.27	98.94	98.68	97.94	98.21	97.96	99.39	97.83	98.17
21,000	98.28	96.79	98.27	98.94	98.68	97.94	98.21	97.96	99.39	97.83	98.18
22,000	98.28	96.79	98.27	98.94	98.68	97.94	98.21	97.96	99.4	97.83	98.2
23,000	98.28	96.79	98.27	98.94	98.68	97.94	98.21	97.96	99.4	97.83	98.2
24,000	98.28	96.79	98.27	98.94	98.68	97.94	98.21	97.97	99.4	97.83	98.2
25,000	98.28	96.79	98.27	98.94	98.68	97.94	98.21	97.97	99.4	97.83	98.2