

A Corpus-based Study of the Vocabulary Profile of High School English Textbooks in China

Min Yu^a and Willy A Renandya^b

^a sabrina7600@163.com, School of Foreign Languages, Tianshui Normal University, China

^b willy.renandya@nie.edu.sg, National Institute of Education, Nanyang Technological University, Singapore

* Corresponding author, willy.renandya@nie.edu.sg

APA Citation:

Yu, M., & Renandya, W. A. (2021). A Corpus-based Study of the Vocabulary Profile of High School English Textbooks in China. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 28-49.

<p>Received 22/04/2020</p> <p>Received in revised form 12/08/2020</p> <p>Accepted 07/09/2020</p>	<p>Abstract</p> <p>The study investigates the vocabulary profile of a set of English textbooks <i>New Senior English for China</i>, which is widely used for senior secondary education in China. It examines how the words required by the 2017 National English Curriculum Standard for General Senior Secondary Education in China are covered, repeated and distributed in the textbooks. The results show that the textbooks cover only about 80% of the lemmas required by the 2017 English Curriculum Standard. Among the lemmas covered in the textbooks, half of them are repeated less than five times in the textbooks. Most of the lemmas which recur more than five times in the textbooks have dispersion values above 0.5. Lemmas with dispersion values below 0.1 are mainly composed of theme-based words. Although the study indicates that some words are distributed favorably, the textbooks fail to provide sufficient coverage and repetition of the words required by the 2017 English Curriculum Standard. Therefore, extra exposure and repetition of these words are required for optimal learning.</p>
<p>Keywords L2 vocabulary; L2 curriculum; vocabulary coverage; textbooks</p>	

1. INTRODUCTION

English language teaching in mainland China (hereafter 'China') has been embraced enthusiastically by the public since English is not only a compulsory subject in the curriculum, but also a vital determination for university entrance exam and a promising factor for well-paid jobs (Adamson, 2004). As written in the English Language Curriculum Standard (MOE, 2018), students' good command of English does not only benefit their individual development, but also contributes to the country's future development in the long term.

Teaching and learning English in an EFL context such as China relies rather heavily on the textbooks since exposure to the target language for most learners is quite limited. For many teachers in China, they are the major, if not the only, pedagogic resource to hand. English language textbooks are also the most important source of English input to learners (Hu, 2002). The compilation of Textbooks in China is curriculum-oriented. English language textbooks are very much a "curricular artefact" and the 'main manifestation of the intended curriculum" (Adamson 2004, p. 6-8). Curriculum documents are the most important guidance for the selection of vocabulary, grammar, and other learning materials to include in textbooks.

Vocabulary knowledge is often regarded as a measure of students' language learning progress (Adamson, 2004). This is true when students are learning any language in China, no matter when they learn English, Japanese, Chinese or any other languages. The Chinese government, for example, prescribed the number of Chinese characters which were widely used in the mass media in mass literacy campaign in the 1950s. Similarly, in English language education, a specific number of English words is required to be taught and learned according to the English language curricula (ibid). Vocabulary teaching and learning are considered to be among the most important aspects in language teaching in China, which can be demonstrated by the fact that a wordlist containing core vocabulary words has always been included in every English curriculum in China.

The latest reform of the English language curricula for senior secondary education is the implementation of the 2017 National English Curriculum Standard for General Senior Secondary Education in China (referred as 2017 English Curriculum Standard) (MOE, 2018). This new

curriculum has included some major changes in the teaching aims and contents which need to be reflected in the new English language textbooks. This study focuses on the changes of the vocabulary requirements in this new curriculum standard. In particular, it examines how the words required in this new curriculum standard are covered, repeated and distributed in one set of widely used textbooks in high schools in China - *New Senior English for China* (Liu et al., 2007). This study seeks to address the following research questions:

1. To what extent does the set of textbooks *New Senior English for China* cover the vocabulary for compulsory learning as required by the 2017 English Curriculum Standard (MOE, 2018)?

2. What is the rate of the repetition and dispersion of these words in this set of textbooks?

2. LITERATURE REVIEW

2.1 Breadth and Coverage of Vocabulary

The term “breadth of vocabulary” is normally used to refer to the number of words known by the learners. It refers to ‘the number of words for which the person knows at least some of the significant aspects of meaning’ (Anderson & Freebody, 1981, p. 93). Breadth of vocabulary is frequently investigated with “coverage”, which means the percentage of vocabulary the learners know in a text. In the present study, breadth and coverage are used to refer to the number of English words required by the 2017 English Curriculum Standard which are covered in the English language textbooks.

According to Webster’s Third New International Dictionary, the English language contains approximately 114,000 word families excluding proper names (Nation, 2001). Educated monolingual native speakers know approximately 20,000 word families (Nation, 2001; Schmitt, 2010). However, acquiring 20,000 word families is a very ambitious goal for second or foreign language learners. Researchers have found that words are not created equal: some are more frequently used, and therefore more useful. These words appear much more frequently and account for a large proportion in daily language use. These words should be given a higher priority in the L2 classroom learning since knowing them enables the learners to know a large part of the text and therefore, promotes

comprehension in reading and communication. A classic list of such words is the General Service List of English Words, which contains 2,000 word families and accounts for some 80% of words found in most written texts (West, 1953). Another list generated from academic texts is the Academic Word List (Coxhead, 1998), which contains 570 headwords and accounts for 10% in written academic texts in the fields of arts, commerce, law and science. These 570 headwords are outside of A General Service List of English Words. These two sets of wordlists are often used as a guide for the choice of vocabulary words in basic English curricula.

In China, English language curricula of different levels of education typically list words which students should master in order to pass the relevant English language tests. Therefore, numerous studies on vocabulary breadth and coverage of English language textbooks in China have examined how words required by curriculum documents are covered in textbooks. Most of these studies show that the selection of vocabulary in textbooks is not always well-aligned with the contents of the curriculum. This of course could have a negative impact on the students since the curriculum documents in China are important guidance for drafting high stake exams such as College Entrance exam, CET band 4 and Band 6 etc. Wang and Xu (2013), for example, found that the four most widely used sets of textbooks did not adequately cover the vocabulary required for College English Test (CET) Band 4 and Band 6 in China, with some of the textbooks covering only about 60% of the words mandated by the curriculum. Liu and Zhang (2015) and Zhou (2012) found the textbooks they studied covered, averagely, only about 75% of the lemmas required by the College English Curriculum Requirements. At secondary education level, similar problems were found. Xie (2010) found that both of the two English language textbooks failed to cover all the lemmas required by the curriculum standards, only covering 84.5% of them, among which 14% recurred only once or twice. Zhou and Li (2013) found the two sets of textbooks they studied covered only about 75% lemmas required by the curriculum standard (MOE, 2003).

Insufficient coverage of words required by curriculum documents has been a common phenomenon in English language textbooks used in China. However, there is so far a lack of research on the words required by 2017 English Curriculum Standard, the latest version of English language curriculum, which may be in effect and a guidance of English

teaching and learning in general senior secondary education in the coming 10 years. To find out how the words required by 2017 English Curriculum Standard are presented in English language textbooks is important to teachers, learners and other relevant people.

2.2 Repetition and Dispersion of Vocabulary in Textbooks

Repetition refers to the number of times the target words are seen by the learners. The term “repetition” is used interchangeably with “number of encounters” or “occurrence”, “exposures”, “recurrence”, “frequency” in the present study.

Ebbinghaus’ study on human memory showed that newly learned knowledge need be reviewed before it is forgotten, and such a review should be done at increasingly longer intervals (Ebbinghaus, 1913, as cited in Gu, 2005). According Ebbinghaus’ forgetting curve principle, people forget what they have learned quickly after their initial learning, and the forgetting speed gradually slows down with frequent reviews. Therefore, ELT experts (e.g., Nation, 2001) suggest that multiple encounters of words are needed for vocabulary learning and retention. They generally agree that the more often words are met, the more likely they will be learned and retained (Horst et al., 1998; Pigada & Schmitt, 2006). No consensus has been reached on the specific number of encounters which can be considered adequate for the learning and retention of words. Some researchers, such as Waring and Takaki (2003), claimed that eight encounters are enough, but others maintained that at least ten exposures are necessary (Pigada & Schmitt, 2006; Webb, 2007). Horst et al. (1998) concluded eight to twelve encounters are essential, while others are more optimistic, believing that six encounters can already equip learners with much word knowledge (Rott, 1999).

A number of studies on the vocabulary load in textbooks reveals inadequate repetition of target words (Matsuoka & Hirsh, 2010; Nordlund, 2016). Studies on vocabulary repetition in English language textbooks used in China have found similar results. Liu and Zhang (2015) examined four textbooks and found that 23% to 30% lemmas from the College English Curriculum Requirements appeared only once. A similar study was carried out by Zhou (2012). He found more than half of the word families required by the College English Curriculum Requirements appeared four times or less in the four textbooks he studied. Zhou and Li

(2013) studied how the words required by the curriculum standard (MOE, 2003) were repeated in the set of textbooks – *New Senior English for China*. They found 36% of these lemmas recurred four times or less in the textbooks.

Though repetition is crucial for the vocabulary learning, it alone is inadequate to depict an accurate vocabulary profile in textbooks. If words are repeated at a high frequency, but such frequency happens within a specific section of the corpus, the results may still be dubitable or even misleading (Gries, 2010). An important distinction about repetition is between massed and spaced repetition. Spaced repetition, which means the words appear with longer intervals, is reported to be more effective in vocabulary learning than massed repetition, where the concentration of repetitions of a word is in only one part of a text, or the repeated attention to a word is limited in a continuous period of time (Matsuoka & Hirsh, 2010; Nation, 2001; Nation & Webb, 2011). The different types of repetition (i.e., spaced or massed) can be measured using a statistic called “dispersion value”. This statistic shows how evenly, or unevenly words are distributed across the corpus (Oakes, 1998; Gardner & Davies, 2014). The dispersion value ranges from 0 to 1, with 0.9 or 1 suggesting even distribution and 0 or 0.1 suggesting uneven distribution (Katz, 1996).

Both repetition and dispersion rates should be taken into consideration in vocabulary acquisition. The latter however is severely under-studied. To date, only a few studies have examined vocabulary dispersion in textbooks. Alcaraz-Mármol (2015) conducted an experiment to compare the predictive power of dispersion and frequency of words in one EFL coursebook for middle school students in Spain. She calculated the number of occurrences of 30 target words in the textbook. Gries’ Deviation of Proportions (DP) was also calculated for all the target words according to their dispersion in the textbook. She found that the dispersion value, which was represented by the DP in her study, was a better predictor compared with mere repetition.

Zhou (2012) analysed the dispersion value of the words required by the College English Language Curriculum Standard in English language textbooks used in China. He found that the dispersion value of more than 90% word families is fairly high. Word families with low dispersion values were mostly theme-related proper nouns. Based on these results, the

author concluded these textbooks provided favorable vocabulary learning opportunities for students in terms of the vocabulary dispersion.

In sum, most studies on vocabulary coverage in textbooks revealed that textbooks generally fail to cover the target words satisfactorily, and the repetition of target words in textbooks is inadequate for effective learning and retention. However, most such studies are based on previous curriculum published in 2003, or curriculum for college education level. There is so far a lack of research on the words required by 2017 English Curriculum Standard. In addition, very few such studies have taken both repetition and style of repetition into consideration. The present study therefore seeks to examine the coverage, repetition, and the repetition style of the words required by the 2017 English Curriculum Standard and to also provide clarity on past research studies that often reported mixed and inconsistent findings.

3. METHODOLOGY

3.1 The Textbooks

This study investigated the vocabulary profile of a set of English language textbooks *New Senior English for China*, published by People's Education Press (PEP). The choice of the textbooks for the present study was motivated by two reasons. First, this set of textbooks was published by People's Education Press (PEP) in Beijing, a publishing company under the direct jurisdiction of the Ministry of Education of China (Zhang, 2004). Textbooks published by PEP usually enjoy the reputation as the authority and considered to be of a higher quality in China. Second, this set of textbooks is widely used in senior secondary schools in China. According to a study by Gu, Zhang and Xiao (2011), 67.8% of the participants in their study reported that they learned English with this set of textbooks when they were in senior secondary schools.

TABLE 1

Descriptive Statistics of the Vocabulary in New Senior English for China

Text File	Tokens in Text	Word Types	Standardized TTR
Overall	305038	12894	40.60
BOOK 1.txt	22961	2925	40.97

BOOK 2.txt	23596	2894	39.79
BOOK 3.txt	24479	3178	40.74
BOOK 4.txt	24386	3124	40.29
BOOK 5.txt	26218	3410	41.21
BOOK 6.txt	25483	3208	40.78
BOOK 7.txt	29811	3650	40.91
BOOK 8.txt	28895	3881	39.25

Each textbook of the set contains 5 units. The content is theme based. Every unit focuses on one topic. Topics include friendships, travelling, science and technology, and so on, mainly presented through a reading text. Tasks and activities are designed to test the comprehension of the reading text, enhance the vocabulary knowledge, and the speaking, writing and listening skills about the topic.

The eight textbooks were scanned and saved as PDF files. ABBYY Fine Reader 12 was used to convert the PDF files into word files. Non-word contents, such as pictures and graphs, were deleted from the corpus. The final version was saved as a txt file. Altogether, there were eight text files, one file per textbook. The overall descriptive statistics of the corpus generated by WordSmith 7.0 are presented in Table 1. The textbooks from Book 1 to Book 8 contain similar numbers of tokens and word types, except that the last two textbooks, Book 7 and Book 8, contain apparently many more tokens and word types. This is a reasonable arrangement since the last two textbooks are at the last stage of secondary school education and should be more difficult in the content compared with the other six textbooks.

3.2 Wordlist in the 2017 English Curriculum Standard

As shown in Table 2, the basic vocabulary requirement for senior secondary education is to learn 500 new words¹. Students should master these words by learning the first five English textbooks, from Book 1 to Book 5. Students should master extra 1,000 to 1,100 new words in order to meet the requirement for college entry exam. Book 1 to Book 8 are designed to prepare students for this exam. In order to reach an

¹ “Words” in this Section refers to the original words listed in and required by the English Curriculum Standards, without being lemmatised or familized. For example, “exciting” and “excited” are listed as two different words. “Advice” and “advise” are also treated as two different words.

advanced level or if students are keen on furthering their English learning, they may choose to continue learning of Book 9 to Book 11, or choose to learn more modules such as second foreign language other than English, or Business English, Tourism English, and English Literature (MOE, 2018, pp 9-10).

The wordlist in 2017 English Curriculum Standard includes 3,000 words approximately. The wordlist contains three sub-wordlists. The first sub-list contains 500 words, which form the basic vocabulary requirement for all students at the senior secondary education. The second sub-list contains about 1,000 words, which are required to be learned by students who will take the college entrance exam. The third sub-list includes about 1,500 words, which are required to be learned by students at the primary and junior secondary education.

TABLE 2

Integrated Vocabulary Requirements in English Curriculum Standards from Primary School Education to Senior Secondary Education (2011, 2017)

	English Curriculum Standards (2011, 2017)	PEP Textbooks
	Advanced level requirement 1,000 new words, accumulatively reaching 4,000 to 4,200 words, without Word List.	<i>New Senior English for China</i> Book 1 - 11
Senior Secondary School (Grade 10 - 12)	College Entry Requirement 1,000 to 1,100 new words, accumulatively 3,000 to 3,200 words, with Word List provided.	<i>New Senior English for China</i> Book 1 - 8
	Basic requirement for senior secondary education 500 new words approximately, accumulatively 2,000 to 2,100 words, with Word List provided.	<i>New Senior English for China</i> Book 1- 5
Junior Secondary School (Grade 7 - 9)	LEVEL 3, LEVEL 4 and LEVEL 5 Basic requirement for Grade 9 (Junior 3), vocabulary requirement: 800 to 900 new words, accumulatively reaching 1,500 to 1,600 words, including those required at Level 1 and 2.	
Primary School (Grade 1 - 6)	LEVEL 1 and LEVEL 2 700 words	

A close analysis reveals that the words required by the 2017 English Curriculum Standard contains 325 lemmas from AWL (Coxhead, 1998), 1499 from GSL (West, 1953), and 75% of the lemmas required by the 2017 English Curriculum Standard are from the first three thousand words in the BNC-COCA frequency lists (Nation, 2017), which demonstrated that the words required by the 2017 English Curriculum Standard were carefully selected according to their frequency in use.

TABLE 3

Adapted Wordlists from 2017 English Curriculum Standard

List	Description	Number of Lemmas
Lemma List One	Lemmas required to be mastered at Primary School and Junior Secondary education	1,437
Lemma List Two	Lemmas required to be mastered at senior secondary education (Compulsory)	486
Lemma List Three	Lemmas required to be mastered at senior secondary education for college entry (selective compulsory, only for students who will take college entrance exam)	985
Lemma List Four	Accumulated vocabulary requirement upon the completion of senior secondary education, including vocabulary required by Primary School and Junior high school (Lemma List One), Compulsory word requirement for senior secondary education (Lemma List Two), and college entry vocabulary requirement (Lemma List Three)	2,908
Lemma List Five	Overall vocabulary which should be learned at senior secondary education (Lemma List Two + Lemma List Three)	1,471

(Adapted from English Curriculum Standard, 2018)

It is worth noting that the words in the wordlist in the 2017 English Curriculum Standard have not been lemmatized or familized. A lemmatized wordlist refers to one which includes both the headword and the inflected words that can be formed from the head word (Nation, 2001). A familized wordlist refers to one which includes both the headword and the inflected and derivational words that can be formed from the head word (Nation, 2001). For example, the words “excited” and “exciting” are considered as two different words in the Wordlist of the 2017 English Curriculum Standard, while the headword of the family “excite” was not in the wordlist. The original wordlist in the 2017 English

Curriculum Standard was converted into a lemma wordlist for the present study through WordSmith Tool (7.0 version). The lemma list compiled by Yasumasa Someya (1998) was uploaded to WordList program as the “lemma list” according to the instruction to WordSmith Tools. The original wordlist required by the 2017 English Curriculum Standard was uploaded to WordList program as the “text”. The lemma list of for the original words required by the 2017 English Curriculum Standard was generated in such way. The lemma list of the vocabulary required by the 2017 Curriculum Standard was split into five sub-lemma lists for the purpose of the current study, as shown in Table 3.

Lemma List One included the lemmas which students should master in their study at primary schools and junior secondary education. This list contains 1,437 lemmas which have been previously studied by students at senior secondary education. Lemma List Two included 486 lemmas which are the basic word requirement for students at senior secondary education. They are for compulsory learning of senior secondary education in China, corresponding to the learning of Book 1 to Book 5. Lemma List Three included 985 lemmas that students need to study in order to meet the college entry exam requirement. Students who would like to pursue higher education after their senior secondary education should master these words in addition to the lemmas in Lemma List One and Two. Lemma List Four was the combination of Lemma List One, Two and Three. It includes all the lemmas students are supposed to master upon their completion of their senior secondary education, including the lemmas required at primary schools, junior and senior secondary education. Lemma List Five is the combination of Lemma List Two and Three, which includes the lemmas students should learn at senior secondary education only, including both the basic vocabulary requirement and those required for college entry.

3.3 Procedure

To address the first research question, the WordList program in WordSmith 7.0 was operated. The five lemma wordlists generated were uploaded as the match lists separately. The lemma list compiled by Yasumasa Someya (1998) was applied to join the inflections of the same lemmas in the self-constructed corpus of the textbooks automatically. The self-constructed corpus of the textbooks was uploaded as the

selected files to make a word list. The matched wordlist contained the words which existed in both the textbooks and the wordlist required by the 2017 English Curriculum Standard.

To address the second research question, two steps were carried out. The first step was to refer to the results of previous step to find out the frequency of occurrence of each word in the corpus. The second step was to calculate the dispersion rate of the words in the corpus. The program “Concord” in WordSmith was applied for such a purpose. The corpus of textbooks was uploaded to Concord program of WordSmith 7.0. The inflection wordlist was uploaded as the “search-words” file in the Concord program, which computed the dispersion rates of the inflections of lemmas. The dispersion rates of the inflections were merged manually in order to get an overall dispersion rate of the lemma heads.

4. RESULTS

The WordList program in WordSmith 7.0 was utilized to address the first research question. The five lemma lists were uploaded into the program so that they could be matched against the researcher-constructed corpus of the textbooks.

TABLE 4

Descriptive Statistics of Vocabulary Coverage

	Total Lemmas	Books 1-5		Books 1-8	
Lemma List One	1,437	1,346	93.67%	1,384	96.31%
Lemma List Two	486	384	79.01%	447	91.98%
Lemma List Three	985			770	78.17%
Lemma List Four	2,908			2,601	89.44%
Lemma List Five	1,471			1,217	82.74%

As shown in Table 4, the eight books (Books 1 to 8), cover 2,601 lemmas out of the total 2,908 lemmas in Lemma List Four, covering 89.44% of the lemmas required by the 2017 English Curriculum Standard. A closer analysis revealed that among the 307 uncovered lemmas, 53 lemmas have been studied by the students because they are required by

the English Language curricula at the primary and junior secondary schools (MOE, 2011). Examples of these words include DRY, EVENING, COLOUR, SCHOOLBAG, TROUSERS, WHEEL, etc. Although these words are not covered in the textbooks for senior secondary education, they are assumed to be known words to the students. There are 254 uncovered lemmas which are new words for students and should be studied and mastered through learning Book 1 to Book 8 at the senior secondary education stage, accounting for 17.26% of the total 1,471 lemmas which should be learned at this stage. These important but missing lemmas include ADDICT, INVESTMENT, LABOR, FIND, IMPLY, WAGE, etc. Among these 254 uncovered lemmas, 39 of them are from Lemma List Two, whose lemmas are set as the compulsory and basic lemmas and should be mastered by students in senior secondary schools. These lemmas include ALARM, IMAGE, HONOR, MEDIA, ROMANTIC, POSTPONE, STRATEGY, etc. The remaining 215 uncovered words are from Lemma List Three, which contains the extra lemmas students should master to meet the vocabulary requirement of the college entry exam, such as APPETITE, ASSIGN, COOPERATE, GENERATE, HUMBLE, NEGOTIATE, WISDOM, etc.

To address the second research question, two steps were carried out. The first step was to compute the rate of repetition of the target lemmas in the textbooks with the WordList program in WordSmith 7.0. The second step was to calculate the dispersion rate of the lemmas in the researcher-constructed corpus. Following Zhou (2012), the present study only considered the dispersion values of the lemmas which recurred five times or above in the corpus. The reason is that repetition of five times is considered to be the threshold needed for successful acquisition of the target words (Nation, 2001; Zhou, 2012). The program "Concord" in WordSmith was utilized for this purpose.

Table 5 shows the overall repetition rates of the lemmas required by the 2017 English Curriculum Standard in the textbooks. Lemma List Two contains 486 lemmas which should be covered by its corresponding textbooks, which are Book 1 to Book 5. There are 384 of these lemmas which are covered in the first five books of the set of textbooks. The word which appears most frequently in the textbooks is CHINA, 104 times in the five textbooks. The word PARTNER ranks the second and appears 100 times in the textbooks. Words appearing dozens of times in the textbooks include ACCIDENT, COMPANY, IMPROVE, TOPIC, etc. Nearly half of these 384 lemmas are repeated four times or less in the

textbooks, accounting for 44% of them, including BATTLE, BRANCH, PRESSUE, REMIND, BENEFIT, SOLUTION, etc.

TABLE 5*Descriptive Statistics of Vocabulary Repetition Rates*

		Lemma List Two	Lemma List Three	Lemma List Five
Book 1-5	≤ 4 Times	169	44%	
	5-15 Times	160	42%	
	≥16 Times	55	14%	
Book 1-8	≤ 4 Times		420	55%
	5-15 Times		271	35%
	≥16 Times		78	10%
			550	45%
			477	39%
			190	16%

Similarly, lemmas in Lemma List Three should be covered by textbooks from Book 1 to Book 8. The research shows that 770 lemmas (out of 985) are covered in these eight textbooks. Among these 770 lemmas, words which appear most frequently include THINK, FOUND, PARAGRAPH, HOUSE, and SAY. They appear more than 100 times in the eight textbooks. Words appearing dozens of times in the textbooks include PARK, EVIDENCE, CLONE, MEAN, PREDICT, GIFT, GRAVITY, DEBATE, THEORY, etc. More than half of them (55%) appear four times or less in the textbooks, such as ABANDON, BRILLIANT, CONFUCIUS, EFFICIENT, CONCEPT, EMPLOY, WEAPON.

Table 5 also shows the overall repetition rate of the lemmas which should be learned through Book 1 to Book 8. Among the total 1,217 lemmas which appear in the textbooks, there are 45% of them which are repeated four times or less in the textbooks.

As for the dispersion value of the lemmas required by the curriculum standard in the textbooks, following Zhou (2012), the present study focused on the analysis of the dispersion values of the lemmas which recur five times or above in the textbooks. The reason is that if the lemmas are not repeated adequately, there is little point to look into the distribution of them in the texts, which is represented through the dispersion values in this study.

TABLE 6*Summary of Lemmas Repeated Five Times or Above*

		Lemma List Two	Lemma List Three	Lemma List Five
Book 1-5	5-15 Times	160		
	≥16 Times	55		
		Total: 215 lemmas		
Book 1-8	5-15 Times		271	477
	≥16 Times		78	190
			Total: 349 lemmas	Total: 667 lemmas

As Table 6 shows, the dispersion values of three groups of words are calculated, 215 lemmas from Lemma List Two in Book 1 to Book 5, 349 lemmas from Lemma List Three in Book 1 to Book 8, and 667 lemmas in Book 1 to Book 8.

TABLE 7*Descriptive Statistics of Vocabulary Dispersion*

		Lemma List Two	Lemma List Three	Lemma List Five		
Book 1-5		215 lemmas				
	0.1 <	6	6			
	≥ 0.5	119	119			
	≥ 0.8	20	20			
Book 1-8			349 lemmas	667 lemmas		
	0.1 <		6	2%	38	6%
	≥ 0.5		228	65%	404	61%
	≥ 0.8		30	9%	66	10%

As Table 7 shows, 55% of the lemmas which appear five times or above from Lemma List Two have dispersion values above 0.5 in Book 1 to Book 5. Samples of such words are TIP, TITLE, PARTNER, COUNTRY, COMPANY, ACCIDENT, IMPROVE, etc.

In Book 1 to Book 8, 65% of the lemmas which appear five times or above from Lemma List Three have dispersion values above 0.5. In

addition, the lemmas with dispersion values above 0.5 account for 61% of the total lemmas which recur five times or above in Lemma List Five. Around 10% of the lemmas have dispersion values at 0.8 or above.

5. DISCUSSION

The findings of the present study reveal that although there is a rather substantial change in the two versions of English Language Curriculum Standards published in 2003 and 2017, the current English textbooks present a similar coverage of the words required by these two standards. The textbooks cover 84.5% of the lemmas required by the 2003 English Curriculum Standard (Xie, 2010) and 82.74% of the lemmas required by the 2017 English Curriculum Standard according to the present study.

TABLE 8

Comparison of the Findings in Xie's Study and the Present Study

Studies	Total Number of Lemmas Required by Curriculum Standards for Senior Secondary Education	Number of Lemmas Required for College Entry Exam	Lemmas Covered in <i>New Senior English for China</i>	Coverage Rate
Xie (2010)	3,398 (MOE, 2003, as cited in Xie, 2010)	920 (MOE, 2003, as cited in Xie, 2010)	569	61.8%
The Present Study	2,908 (MOE, 2018)	985 (MOE, 2018)	770	79.01%

As for the coverage of words in some of the sub-wordlists, the present study shows different results from Xie's (2010) study. As Table 8 shows, the 2003 English Curriculum Standard required 920 lemmas for college entry exam while 2017 English Curriculum Standard required 985 lemmas, an increase of 65 lemmas. The comparison of Xie's study (2010) and the present study shows that the same set of textbooks *New Senior English for China* covered 770 lemmas required by the 2017 English Curriculum Standard for college entry exam, which is 201 lemmas more

than that in the 2003 English Curriculum Standard. More words required by the 2017 English Curriculum Standard for college entry exam are covered in the textbooks compared with the 2003 English Curriculum Standard.

In sum, the textbooks provide about 80% coverage of the lemmas required by the curriculum standard for senior high schools in China. Given the limited exposure to English outside the classroom, we feel that it may not be easy for the students to meet the uncovered 20% of the lemmas required by the curriculum. Therefore, supplementary wordlists may be necessary to compensate for this gap. We compiled two supplementary wordlists that contain all the uncovered lemmas. Appendix 1 contains 102 lemmas not covered in Books 1 to 5; while Appendix 2 provides a list of 215 lemmas needed for college entrance examinations, but are not listed in the textbooks. We hope that teachers may find the wordlists useful to supplement their teaching of vocabulary and future textbook writers may also use the wordlists as a useful reference for their current or future coursebook projects. The two appendices can be viewed or downloaded here (<https://tinyurl.com/u45yd8y>).

As for the repetition rate, the results of the present study are in line with the findings of most previous studies on vocabulary repetition rates, which show inadequate repetitions of the target words in textbooks (Liu & Zhang, 2015; Matsuoka & Hirsh, 2010; Nordlund, 2016; Zhou, 2012). Similar to the present study, Zhou and Li (2013) for example also analysed the vocabulary repetition in *New Senior English for China*. They found 36% of the lemmas required by the curriculum standard (MOE, 2003) recurred four times or less while the present study found 45% of the lemmas which recurred four times or less. This shows that the textbooks *New Senior English for China* represents the words required by the 2003 Curriculum Standard better than the words required by the 2017 Curriculum Standard. Although there is no consensus as to the question of how many times words should be met in order to be learned, very few, if any, researchers or pedagogical practitioners believe that four encounters or fewer is sufficient. Nearly half of the words required by the 2017 Curriculum Standard (MOE, 2018) appeared four times or less in the textbooks. Such inadequate recurrence of vocabulary, we believe, could negatively affect vocabulary learning outcomes.

As for the dispersion values, according to Katz (1996), when the dispersion value falls below 0.1, the words are considered to be distributed very unevenly. The unevenly distributed words in the present study are 6% of the total words which recurred five times or above in the textbooks. Such dispersion values are similar to the findings reported in Zhou (2012), in which he found that on average about 6.5% of the word families required by the 2003 Curriculum Standard for College English have dispersion values below 0.1 in the four sets of textbooks he studied. He also found the word families with dispersion values lower than 0.1 in his study were proper names, or theme-based words, or “easy” words which students have studied previously in primary schools or high schools. Similarly, in the present study, the 38 lemmas with dispersion values below 0.1 also include many theme-based words such as AMBULANCE, ALCOHOL, BLOOD, CLONE, VOLCANO, CAPSULE, CARBON, TYPHOON, and WRIST. In addition, most lemmas which recur five times or above have dispersion values above 0.5. About 10% of them have dispersion values at 0.8 or above, which is considered to be a reliable benchmark to distinguish the evenly distributed words from the less evenly distributed ones (Gardner & Davies, 2014). In sum, the lemmas required by the 2017 English Curriculum Standard which recur five times or above in the textbooks are generally distributed evenly in the textbooks.

6. CONCLUSIONS AND RECOMMENDATIONS

Overall, the present study reveals inadequate coverage and repetition of the words required by the 2017 English Curriculum Standard in the widely used English textbooks *New Senior English for China*, though the dispersion rate shows words with adequate frequency are well distributed in the textbooks. The selection and arrangement of content and vocabulary of the set of textbooks fail to provide favorable learning opportunities of the target words stipulated in the 2017 English Curriculum Standard.

One pedagogical implication is that teachers need to expose their students to those words not covered in the textbooks. The two supplementary wordlists referred to in the previous section may come in handy for teachers to use as a reference point. However, as Xie (2010) pointed out, a mere listing of the words may not be adequate in

promoting vocabulary learning and retention. Teachers may for example group these words thematically and then intentionally include these words in their lesson planning and delivery. Thematically relevant reading materials can also be used to teach these words during the reading lessons.

Another option is to initiate an extensive reading project for students, which could substantially increase exposure to the target language. Extensive reading provides students with ample opportunities to encounter words a number of times in meaningful contexts. Research shows that students who read in quantity over a period of time acquire substantial language learning benefits, including improved reading fluency and increased vocabulary knowledge (Ng et al., 2019). For optimal vocabulary learning benefits, teachers are well-advised to heed Nation and Waring's (1997) recommendation, i.e., combining direct and deliberate vocabulary instruction with a carefully planned extensive reading programme.

Though the computing process of vocabulary coverage, repetition and dispersion was carefully designed and implemented to ensure revealing and reliable findings, future study could be further improved from the following perspectives. Firstly, vocabulary repetition research could consider sense repetition as well instead of just the form repetition. Secondly, the actual time intervals between two encounters could be considered when computing the repetition rate and dispersion value. Interview or survey of the teachers and learners may help to understand how much time is actually spent on each unit and textbook. Corpus could be interpreted in consideration of the time intervals of teaching and the forgetting curve (Ebbinghaus, 1913, as cited in Gu, 2005) to investigate how the encounters of words are spread throughout the teaching process and whether such distribution of encounters accords with the human memory mechanism.

ABOUT THE AUTHORS

Min Yu: an English lecturer at Tianshui Normal University, Gansu, China. Her research interests include vocabulary acquisition, second language reading, extensive reading, and corpus linguistics. She has published some articles in the area of Applied Linguistics and a few English language textbooks used by Chinese students.

Willy A. Renandya: a language teacher educator. He is currently teaching applied linguistics courses at the National Institute of Education, Singapore. He has taught in many parts of Asia and is a frequent keynote speaker at ELT conferences. He maintains an active language teacher professional development forum called Teacher Voices: <https://www.facebook.com/groups/teachervoices/>.

FUNDING

Tianshui Normal University provided financial support for part of this research under project number 0113-1011301810.

REFERENCES

- Adamson, B. (2004). *China's English: A history of English in Chinese education*. Hong Kong University Press.
- Alcaraz-Mármol, G. G. (2015). Dispersion and frequency: Is there any difference as regards their relation to L2 vocabulary gains? *International Journal of English Studies*, 15(2), 1-16.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). International Reading Association.
- Coxhead, A. J. (1998). *An academic word list (English Language Institute Publication No. 18)*. Victoria University
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Gries, S. T. (2010). Useful statistics for corpus linguistics. *A mosaic of corpus linguistics: Selected approaches*, 66, 269-291.
- Gu, P. Y. (2005). *Vocabulary learning strategies in the Chinese EFL context*. Marshall Cavendish Academic
- Gu, X. D., Zhang, Z. C., & Xiao, W. (2011). An exploration of the current senior English teaching and learning in the process of the new curriculum reform. *Curriculum, Teaching Material and Method*, 31(2), 68-73.
- Horst, M., Cobb T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Hu, G. (2002). English language teaching in the People's Republic of China. In R., Silver, G., Hu, & M., lino, *English language education in China*,

- Japan, and Singapore*. Graduate Programmes and Research Office, National Institute of Education, Nanyang Technological University.
- Katz, S. M. (1996). Distribution of content words and phrases in text and language modelling. *Natural Language Engineering*, 2(1), 15-59.
- Liu, D., Gong, Y., Zheng, W., Brooks, D., & Aldhamland, M. (2007). (2nd ed.). *New Senior English for China (Books 1-8)*. People's Education Press.
- Liu, Y. H., & Zhang, L. J. (2015). A corpus-based study of lexical coverage and density in college English textbooks. *Foreign Language Education in China (Quarterly)*, 8(1), 42-50. Retrieved March 21, 2018, from www.cnki.cn.
- 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.
- MOE. (2003). *Putong gaozhong yingyu kecheng biaozhun*. [English curriculum standards for general senior secondary education.] People's Education Press.
- MOE. (2011). *Yiwu Jiaoyu yingyu kecheng biaozhun*. [English curriculum standards for Compulsory education.] Beijing Normal University Press.
- MOE. (2018). *Putong gaozhong yingyu kecheng biaozhun (2017 nian ban)*. [English curriculum standards for general senior secondary education, 2017.] People's Education Press.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I.S.P. (2017). *The BNC/COCA Level 6 word family lists (Version 1.0.0)* [Data file]. Available from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.
- Nation, P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning
- Ng, Q.R., Renandya, W.A., & Chong, M.Y.C. (2019). Extensive reading: Theory, research and implementation. *TEFLIN Journal*, 30(2), 171-186.
- Nordlund, M. M. (2016). EFL textbooks for young learners: a comparative analysis of vocabulary. *Education Inquiry (Co-Action Publishing)*, 7(1), 47-68. <https://doi.org/10.3402/edui.v7.27764>

- Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh University Press
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1–28.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition, 21*(1), 589–619.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Someya, Y. (1998). *English-language lemma list* [Data file]. Retrieved April 30, 2020, from http://www.lexically.net/wordsmith/version5/wordsmith_chinese/index.html?proc_stop_lists.htm
- Wang, T. S., & Xu, Y. Y. (2013). The distribution of word families in college English textbooks. *Computer-assisted Foreign Language Education, 9*(153), 10-15. Retrieved May 14, 2018, from www.cnki.cn.
- Waring R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130–60.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*, 46–65.
- West, M. (1953). *A general service list of English words*. Longman, Green and Co.
- Xie, J. C. (2010). A corpus analysis of vocabulary in EFL course books for high school. *Foreign Language Learning Theory and Practice, 1*, 55-61. Retrieved March 21, 2018, from www.cnki.cn.
- Zhang, L. J., (2004). Extending the reach of middle school EFL teachers in the People's Republic of China. In W. K., Ho, & R., Wong. *English language teaching in East Asia today: Changing policies and practices*. Eastern Universities Press
- Zhou, L. J., & Li, Q. S. (2013). A comparative study of the vocabulary in secondary junior middle school education in China. *English Teachers, 1*(1), 66-70. Retrieved March 20, 2018, from www.cnki.cn.
- Zhou, Q. (2012). *A corpus-based study on the vocabulary in China's contemporary college English coursebooks* [Unpublished doctoral dissertation]. Shanghai Normal University].