

## How Useful Can a College General English (GE) Text Book be for Preparation for a General English Proficiency Test (GEPT)?

Wenhua Hsu  
I-Shou University  
Taiwan

### Abstract

*This paper describes a preliminary attempt to survey the General English (GE) textbooks used in an EFL tertiary setting and their roles in new word learning and in preparation for an English proficiency test. The researcher compiled a corpus containing 25 GE textbooks, totaling 434,335 words as the source of analysis. Through the RANGE program with twenty-five 1,000-word-family lists derived from the British National Corpus (BNC) and the Corpus of Contemporary American English, the vocabulary levels of GE textbooks (set at 95% lexical coverage) and the lexical coverage of the intermediate General English Proficiency Test (GEPT) vocabulary in GE textbooks were measured. Results show that beyond the most frequent 3,000 word families, a GE textbook can supply students with 327–1,211 new word families. A GE textbook may be of some help in preparing learners for an intermediate GEPT by covering 30.49% to 65.02% of the vocabulary involved in the test. It is hoped that the indices examined in this study would help English teachers raise their awareness of vocabulary levels when choosing a textbook for classroom use.*

**Keywords:** lexical coverage, vocabulary levels, vocabulary threshold

### Introduction

In Taiwan, English education officially starts in the third year of elementary school, though private schools may begin English programs in the first grade. During Grades 1–9, English curricula aim to familiarize learners with basic grammar and the most commonly-used words. Referring to West's (1953) General Service List (GSL) of English words and English programs used by high schools in Asian countries, the Ministry of Education published a basic English word list containing 2,000 lexical items as a curricular guideline for elementary and junior high schools. At the senior high school level, the content of English courses is mainly humanities-based and geared towards the general interest of learners rather than to specific purposes. For admission to universities, senior high school graduates must take either the General Scholastic Ability Test (GSAT) or the Department Required Test (DRT). Both college entrance exams include an English test, which involves a vocabulary over the 4,000-word-family level (College Entrance Examination Center, 2017). The first 3,000 word families are generally presumed to be a benchmark of English proficiency for senior high school graduates.

At the tertiary level, English is a required language subject. English courses for general purposes are offered to non-English majors two to three hours per week in the freshman and sophomore years respectively. The curriculum design of General English is expected to broaden students' horizon so that they can meaningfully relate their academic study to other realms of understanding. Crucial to this goal is providing students with versatile academic content covering topics such as culture, nature, science and technology to achieve an all-encompassing development of knowledge. In a recent decade, one of the most widely-implemented efforts in college education has been the adoption of the graduation requirement of English proficiency. Because of academic demands for English abilities and the need for English mastery from the job market, many universities and even technical colleges set graduation benchmarks of English proficiency for their students. Among various English proficiency tests used in higher education institutions, the General English Proficiency Test (GEPT) is one of the English proficiency tests EFL learners in Taiwan are likely to encounter at some point in their studies or in their career. In contrast to internationally well-known English proficiency tests such as TOEFL, TOEIC and IELTS, the GEPT is a test that was developed by the Language Training & Testing Center (LTTC) (<http://www.lttc.ntu.edu.tw>), commissioned by Taiwan's Ministry of Education, for EFL learners at all levels of proficiency. The test is recognized by government agencies as a criterion for promotion.

It is also used by the Academia Sinica in Taiwan as a means of evaluating the English abilities of applicants for their scholarship programs, by private enterprises as a means of determining the English abilities of their employees, and by public and private schools as a criterion of admission, placement or graduation. Currently, four levels of the GEPT are regularly administered: elementary, intermediate, high-intermediate and advanced. The GEPT elementary level is presumed to be appropriate for students who have studied English through junior high school (Grades 7–9). The intermediate level is seen as suitable for senior high school graduates (Grades 10–12) or university freshmen. The high-intermediate level is thought to be suitable for university graduates majoring in English. The advanced level is considered rather difficult so that only one with knowledge of 10,000+ word families would be able to pass it. Each level is administered through a two-stage process. First, all examinees at each level take a listening and reading comprehension test. Those who pass the first stage are allowed to register for the second stage, i.e. the speaking and writing portions of the test. In a range of English proficiency levels for graduation benchmarks stipulated by some universities and colleges for their students, passing the GEPT intermediate-level first stage (listening and reading tests) is widely adopted.

According to the LTTC, GEPT scores can be aligned with the Common European Framework of Reference for Languages (CEFR), which maps out language abilities on a scale of levels ranging from A1 for beginners to C2 for those who have mastered a language. Earning the certificate of the GEPT intermediate level (having to pass both stages of the test) is equivalent to the B1 threshold level under the CEFR, a score of 57–86 for the TOEFL Internet-based test, 550–780 for TOEIC and 4.5–5.0 for IELTS. For English-majoring students, the English proficiency graduation requirement is generally set at the GEPT high-intermediate level, equal to the B2 vantage level on the CEFR, 87–109 for TOEFL iBT, 785–990 for TOEIC and 5.5–6.0 for IELTS. The vocabulary size involved in the reading tests at the GEPT elementary, intermediate and high-intermediate levels is 2,263 words, 4,947 words (including the elementary level 2,263 words) and more than 8,000 words (both the elementary and intermediate vocabulary inclusive) respectively. The vocabulary lists for different levels of GEPT are available at its website (<https://www.gept.org.tw/>).

In Taiwan, English is not an official language. After taking required English courses in the first two years of college, one may learn new English words at a decreasing rate and may even almost stop learning. As far as non-English majors are concerned, general English (GE) courses may be regarded as a transition between senior high school and college English proficiency benchmarks. If students continue to take optional English for Academic Purposes (EAP) or English for Specific Purposes (ESP) courses in the third and fourth years, then GE courses may also be viewed as a launch pad for further English programs. GE textbooks used in the freshman and sophomore years may therefore play an important role in enhancing English abilities. In view of their importance, this research addressed two research questions.

1. If a senior high school graduate has a vocabulary size of the most frequent 3,000 word families, how many new words may one learn from a GE textbook?
2. To what extent does a GE textbook cover the GEPT intermediate vocabulary (the minimum English ability required by most universities in Taiwan)? In other words, how useful is a GE textbook for the intermediate GEPT?

## ***Literature Review***

### **Lexical Coverage**

In light of the potential role of GE courses in the current EFL context, vocabulary goals should be considered first in choosing and preparing teaching materials. Breadth of vocabulary has been identified as one of the most important indicators of reading proficiency and language abilities (Hu & Nation, 2000; Laufer & Sim, 1985; Qian, 2002), since a rich vocabulary makes language tasks (listening, speaking, reading, writing and translation) easier to perform. The limited vocabulary of EFL learners is a major source of difficulty in reading an English text. According to Nation (2006), lexical/text coverage refers to “the percentage of running words in the text known by the readers” (p. 61). It may be the best gauge of whether a text is likely to be adequately understood. Technically, it is calculated as “the number of the words known in a text, multiplied by 100 and then divided by the total number of running words, i.e. tokens in the text” (Nation, 2001, p. 145). The assumption made behind lexical coverage is that there is a lexical knowledge threshold which marks the boundary between having and not having sufficient vocabulary knowledge for adequate comprehension. Native English-speaking children view a vocabulary load of two unknown words per hundred words (i.e. 98% lexical coverage) as difficult reading (Carver, 1994).

That is, one needs to know sufficiently different words (types) to account for 98% of the running words (tokens) in a text. Some researchers regard one unknown word in every twenty words (95% coverage) as the necessary level beneath which readers are not expected to read an authentic text successfully (Laufer, 1989; Read, 2000; Schmitt & McCarthy, 1997). In other words, if more than one word is unknown in every twenty words (<95% lexical coverage), learners would face a daunting amount of dictionary work, namely, looking up new words roughly every two lines. The notion behind this proposition is that learners depend on vocabulary as their first resource. Successful comprehension involves much more than being able to decode the vocabulary in a text, but a lack of familiarity with more than 5% of the running words in a text can make reading a formidable task.

Laufer (1989) proposed that for learners to be able to gain reasonable comprehension of written text, it is necessary to know at least 95% of the words. In a more recent study, Laufer and Ravenhorst-Kalovski (2010) suggest two lexical coverage points, 95% and 98%, which represent the possible lower and upper boundaries over which learners are likely to gain a minimally acceptable comprehension and an optimal comprehension. If at least 95% lexical text coverage is needed for unassisted comprehension, then the researcher would like to apply this assumption to English language testing, since learners cannot resort to dictionaries or consult teachers while doing a test. When students are unassisted on a test, vocabulary size may be one of the predictors of test scores. As the density of unknown words increases, reading and listening comprehension drops.

#### *Lexical Threshold*

A well-educated adult native speaker of English has a vocabulary of around 20,000 words (Nation, 2001). Past studies have shown that the minimal vocabulary size needed for reading authentic texts starts at a low of 5,000 words and ranges up to 10,000 words for reading university textbooks (Hirsh & Nation, 1992; Laufer, 1989). In studies on the lexical thresholds of a variety of texts, Nation (2006) noted that if 98% lexical text coverage is needed for unassisted comprehension, a vocabulary of 8,000 to 9,000 word families is needed for comprehension of written text and a vocabulary of 6,000–7,000 for spoken text.

Chujo (2004) collected a small corpus of TOEFL and TOEIC preparation tests to gauge vocabulary levels. Set at the text coverage of 95% and measured by the British National Corpus (BNC) 1<sup>st</sup>–14<sup>th</sup> 1,000 word family lists, TOEFL involves more vocabulary than TOEIC (6,000–6,500 vocabulary level for TOEFL versus 4,500–5,000 vocabulary level for TOEIC). This means those with knowledge of the first 6,000–6,500 word families are more likely to achieve 95% lexical coverage in TOEFL than others with knowledge of fewer than 6,000 word families. Similarly, to get above average scores in TOEIC, one needs to command a vocabulary of at least 4,500–5,000 word families.

Compared to English native speakers' 20,000 word families, the requirement for a vocabulary size of 5,000–6,000 word families in the current EFL context may be a more feasible goal for college teachers in assisting their students to meet vocabulary thresholds.

### **Research Method**

#### *Textbook Selection*

Since this study aimed to create a corpus of GE reading textbooks widely used in universities in Taiwan to form the basis of analysis, the criteria for inclusion of books in the corpus were based on the popularity of GE textbooks according to sales data from eight bookstores. Among college GE textbooks ranging from low-intermediate to advanced levels, intermediate and high-intermediate GE textbooks are the most commonly-used. A total of thirteen intermediate and twelve high-intermediate textbooks were chosen based on popularity. Excluding exercises and supplementary readings, the main articles in each book were scanned into computer files, manually typed for some texts with illustrations, and proofread to ensure text completeness and spelling correctness. One factor to be noted here is that the actual vocabulary size may be inflated and text coverage may shrink if proper nouns are included. After removing proper nouns, the resulting corpus contained a total of 434,335 tokens, as shown in Table 1.

**Table 1: The composition of a corpus of English Reading textbooks for general purposes**

	GE Textbooks	Level claimed by the publisher	Number of Texts	Tokens
1	What a World 3	intermediate	16	16,716
2	In Context-Steps to academic Reading	intermediate	36	24,305
3	Active Skills for Reading 3	intermediate	32	19,798
4	Active Skills for Reading 4	high-intermediate	32	25,962
5	Cause & Effect-reading 3	intermediate	24	17,975
6	Concepts & Comments-reading 4	high-intermediate	20	16,383
7	Hot Topics 2	intermediate	50	17,384
8	Hot Topics 3	high-intermediate	40	23,403
9	Interactions 2 for reading	intermediate	10	10,574
10	Mosaic Reading 1	high-intermediate	20	17,628
11	NorthStar –high intermediate	high-intermediate	20	16,321
12	Reading Challenge 3	intermediate	20	7,957
13	Reading for Real-high-intermediate	high-intermediate	12	9,185
14	Reading for Success 3	intermediate	32	19,566
15	Reading for the Real World 2	intermediate	24	16,587
16	Reading for the Real World 3	high-intermediate	24	16,414
17	Reading Matters 3	high-intermediate	50	44,499
18	Select Readings-intermediate	intermediate	14	11,204
19	Select Readings-upper-intermediate	high-intermediate	12	16,277
20	Weaving It Together 3	intermediate	16	11,991
21	Weaving It Together 4	high-intermediate	18	16,807
22	World Class Reading 2	intermediate	14	9,968
23	World Class Reading 3	high-intermediate	14	15,811
24	Concepts for Today 4	intermediate	12	10,629
25	Topic for Today 5	high-intermediate	24	20,991
Excluding proper nouns, the total tokens=				434,335

### *The instrument*

In order to measure the vocabulary level of a text, we ran the RANGE program (Heatley, Nation&Coxhead, 2004) equipped with twenty-five 1,000-word-family lists to calculate the number of different word families or types used in the text. Using the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), Nation (2012) and his colleagues compiled 25,000 word families and ranked them from the 1<sup>st</sup> to 25<sup>th</sup> 1000-word-family lists according to their frequency, range and dispersion in the corpora. The rationale behind the ranked word lists is that higher-frequency and wider-range words are more likely to be encountered and thus acquired before lower-frequency and narrower-range words (Nation, 2006). To compare GE textbooks, the operational measures therefore involved vocabulary size, vocabulary levels (along the BNC/COCA word-frequency scale) and lexical text coverage.

### *Base word lists*

In addition to the BNC/COCA 1<sup>st</sup>–25<sup>th</sup> 1,000 word families, which are already incorporated in the RANGE program, another base word lists was compiled for this research. It was the GEPT intermediate word list with elementary vocabulary inclusive, totaling 4,947 words. Through RANGE, the 4,947 words were made into 3,788 word families following Bauer and Nation's (1993) six-level scale, which includes all the affixes from levels 2 to 6. Word families are regarded as an important counting unit in terms of the learning burden (Nagy et al, 1989). The concept of a word family is used to represent a group of words whose meaning can be inferred when the meaning of the base form in the group is known to a learner. Therefore, comprehending regularly inflected or derived members of a word family does not require much effort if learners know the base word and if they have the knowledge of basic word building processes (Bauer & Nation, 1993).

The notion is useful when we are concerned with vocabulary size and lexical coverage in a reading textbook. Subsequently, to calculate lexical coverage and to assess the vocabulary levels of the twenty-five GE textbooks, the RANGE program was run to compare a text against certain base word lists to see what words in the text are and are not in the word lists, and to see what percentage of the vocabulary items in the text are covered by the lists, namely lexical coverage.

## Results and Discussion

RQ1. If a senior high school graduate has a vocabulary size of the most frequent 3,000 word families, how many new words may one learn from a GE textbook?

**Table 2: Text coverage by BNC/COCA 3000 across 25 GE textbooks**

GE Textbook	BNC/COCA 3000				Not in the BNC/COCA 3000				Total	
	tokens	%	types	%	tokens	%	types	%	tokens	types
Active Skills for Reading 3	17,957	90.70%	2,166	59.15%	1,841	9.30%	1,496	40.85%	19,798	3,662
Active Skills for Reading 4	22,971	88.48%	2,711	52.80%	2,991	11.52%	<b>2,423</b>	47.20%	25,962	5,134
Cause & Effect 3	16,898	94.01%	1,922	73.45%	1,077	5.99%	695	26.55%	17,975	2,617
Concepts & Comments 4	14,727	89.89%	1,819	62.83%	1,656	10.11%	1,076	37.17%	16,383	2,895
Concepts for Today 4	9,503	89.41%	1,672	60.90%	1,126	10.59%	1,074	39.10%	10,629	2,746
Hot Topics 2	16,158	92.95%	1,983	65.49%	1,226	7.05%	1,045	34.51%	17,384	3,028
Hot Topics 3	20,707	88.48%	2,338	56.11%	2,696	11.52%	1,828	43.89%	23,403	4,166
Interactions 2	9,661	91.37%	1,492	68.24%	913	8.63%	694	31.76%	10,574	2,186
Mosaic Reading 1	15,833	89.82%	2,196	56.16%	1,795	10.18%	1,715	43.84%	17,628	3,911
NorthStar-high intermediate	14,997	91.89%	2,097	60.86%	1,324	8.11%	1,349	39.14%	16,321	3,446
Reading Challenge 3	7,147	89.82%	1,405	68.22%	810	10.18%	<b>654</b>	31.78%	7,957	2,059
Reading for Real high-intermediate	8,512	92.67%	1,568	65.70%	673	7.33%	818	34.30%	9,185	2,386
Reading for Success 3	17,249	88.16%	2,127	53.94%	2,317	11.84%	1,816	46.06%	19,566	3,943
Reading for the Real World 2	14,291	86.16%	2,087	57.63%	2,296	13.84%	1,534	42.37%	16,587	3,621
Reading for the Real World 3	13,834	84.28%	1,924	54.85%	2,580	15.72%	1,584	45.15%	16,414	3,508
Reading Matters 3	38,679	86.92%	3,250	45.79%	5,820	13.08%	3,848	54.21%	44,499	7,098
Select Readings- intermediate	10,518	93.88%	1,624	68.28%	686	6.12%	755	31.72%	11,204	2,379
Select Readings-high intermediate	14,737	90.54%	2,009	57.77%	1,540	9.46%	1,468	42.23%	16,277	3,477
In context	22,793	93.78%	2,274	66.93%	1,512	6.22%	1,123	33.07%	24,305	3,397
Topics for Today 5	18193	86.67%	2,241	50.65%	2,798	13.33%	2,184	49.35%	20,991	4,425
Weaving It Together 4	14,703	87.48%	2,145	54.25%	2,104	12.52%	1,809	45.75%	16,807	3,954
Weaving It Together 3	11,216	93.54%	2,300	70.67%	775	6.46%	954	29.33%	11,991	3,254
What a World 3	15,248	91.22%	1,921	63.98%	1,468	8.78%	1,081	36.02%	16,716	3,002
World Class Readings 2	9,229	92.59%	1,630	68.50%	739	7.41%	749	31.50%	9,968	2,379
World Class Readings 3	13,609	86.07%	1,950	53.52%	2,202	13.93%	1,694	46.48%	15,811	3,644

As can be seen from Table 2, the categories “BNC/COCA 3000” and “not in the BNC/COCA 3000” present how much of the vocabulary in a GE textbook will be familiar to learners and how many words learners may not know, assuming they have a vocabulary of the most frequent 3,000 word families. For example, 17,957 of the running words in *Active Skills for Reading 3* were in the list BNC/COCA 3000, whereas 1,841 tokens were not. These 1,841 occurrences involving 1,496 different word types would possibly be college freshmen’s new words. By and large, if college freshmen read all the texts of a GE course book within an academic year, beyond the 3,000-word-family level, they would come across 654-2,423 new words, appearing in different types. When transforming to word families, the ratio of word types to word families is approximately 2 to 1 according to Table 2. More accurate ratios of the word type against word family range from 1.42 to 2.01. The continuum from containing 654 to 2,423 new word types (roughly 327 to 1,211 word families) implies that the vocabulary learning goals subject to the selection of a GE textbook are quite wide-ranging. Learning vocabulary at a pace of 327 new word families per academic year did not seem likely to catch up with the 5,000-word-family threshold as proposed by Laufer (1997), if vocabulary size is increased at such a rate per year and continues over four years of college study. Conversely, learning 1,211 new word families from one textbook inclined the researcher to speculate on the possibility in the current EFL context.

If Nation's (2001) estimate that native speakers read about 10-12 books per year to acquire 1,000 word families is correct, then students may easily forget most of the 1,211 new words from one textbook due to only one or two repetitions. In view of this, at the onset of syllabus design, lexical breadth and vocabulary size contained in a textbook cannot be ignored and should be taken into account in the first place. RQ2. To what extent does a GE textbook cover the GEPT intermediate vocabulary (the minimum English ability required by most universities in Taiwan)? In other words, how useful is a GE textbook for the intermediate GEPT?

Research Question 2 can be further put in another way, "how well does a GE textbook prepare for the intermediate GEPT?" Table 3 below gives us some indication of the task facing an EFL learner when preparing to learn the vocabulary necessary for access to the intermediate GEPT. A total of 2,463 word families from the textbook *Reading Matter 3* occurred in the intermediate GEPT 3,788-word-family list. This GE textbook, which was claimed to be suitable for students at the high-intermediate level of English proficiency (see Table 1), would possibly prepare learners for 65.02 % (2,463/3,788) of the vocabulary they would meet in an intermediate GEPT. The vocabulary used in *Reading Challenge 3* for low-intermediate learners covers 30.49 % of the intermediate GEPT vocabulary (1,155/3,788).

**Table 3: Text coverage and word families covered by the intermediate GEPT vocabulary across 25 GE textbooks**

Textbook	% Coverage (in tokens)	Occurrences of the intermediate GEPT word families	% preparation for the intermediate GEPT vocabulary
Reading Matters 3	91.85 %	2,463	65.02 %
Active Skills for reading 4	93.05 %	2,119	55.94 %
Topics for Today 5	91.82 %	1,871	49.39 %
Weaving It Together 4	92.85 %	1,852	48.89 %
Mosaic Reading 1	93.07 %	1,847	48.76 %
Reading for Success 3	93.30 %	1,811	47.81 %
Hot Topics 3	93.25 %	1,799	47.49 %
NorthStar-high intermediate	94.76 %	1,766	46.62 %
Select Readings-upper intermediate	94.71 %	1,707	45.06 %
Active Skills for Reading 3	94.85 %	1,707	45.06 %
Reading for the Real World 2	93.08 %	1,694	44.72 %
World Class Readings 3	91.85 %	1,688	44.56 %
In Context	96.82 %	1,606	42.40 %
Reading for the Real World 3	92.30 %	1,588	41.92 %
What a World 3	93.93 %	1,492	39.39 %
Hot Topics 2	95.54 %	1,475	38.94 %
Concepts & Comments 4	95.46 %	1,467	38.73 %
Concepts for Today 4	94.17 %	1,416	37.38 %
Cause & Effect 3	96.79 %	1,363	35.98 %
World Class Reading 2	96.65 %	1,362	35.96 %
Weaving It Together 3	96.11 %	1,326	35.00 %
Select Readings-intermediate	96.77 %	1,316	34.74 %
Reading for Real-high intermediate	94.77 %	1,295	34.19 %
Interactions 2	95.29 %	1,204	31.78 %
Reading Challenge 3	95.16 %	1,155	30.49 %

Note: % preparation for the intermediate GEPT vocabulary is calculated as the number of word families occurring in the intermediate GEPT vocabulary divided by the intermediate GEPT 3,788 word families.

As far as the difficulty of a GE textbook measured by the coverage of the intermediate GEPT vocabulary is concerned, *In Context* is the least difficult (96.82 % text coverage) among the 25 textbooks. When learners reach the vocabulary level of the intermediate GEPT 3,788 word families, the textbook *In Context* would become less challenging to them. Their vocabulary knowledge of 3,788 word families would enable them to understand 96.82 % of the total words of *In Context*.

They would be able to gain adequate comprehension of the texts from *In Context* in relative ease compared to the other textbooks. On the contrary, *Topics for Today 5* is the most difficult, with only 91.82 % of the words in the texts being known to learners with knowledge of 3,788 word families from the intermediate GEPT vocabulary. As aforementioned, there are twenty-five 1,000-word-family bands created from the BNC and COCA in the RANGE program. Thus, the ranked BNC/COCA twenty-five 1,000-word-family lists were used to identify the graduations among the diverse vocabulary levels contained within the GE textbooks. The text coverage of each 1,000-word-family band in the target textbook was calculated and accumulated by counting the number of 1,000-word-family bands needed until the total coverage reached 95 %. Running *Select Readings-intermediation* the RANGE program with base word lists, we can see that the vocabulary level of *Select Readings-intermediate* was 3,000 word families with its cumulative coverage reaching 95.34 %, if we accept the assumption of 95 % coverage as the minimum level for being able to successfully guess meanings from context and gain reasonable comprehension.

**Table 4: Vocabulary levels of Select Readings-intermediate and Reading Matters 3**

BNC/COCA 1,000-word family lists	Select Readings-intermediate tokens/ coverage	Reading Matters 3 tokens/ coverage
1 <sup>st</sup> 1,000	8,766/ 85.90 %	22070/ 77.16 %
2 <sup>nd</sup> 1,000	757/ 7.41 %	2377/ 8.64 %
3 <sup>rd</sup> 1,000	207/ 2.03 %	804/ 2.92 %
4 <sup>th</sup> 1,000		67/ 2.4 %
5 <sup>th</sup> 1,000		392/ 1.42 %
6 <sup>th</sup> 1,000		218/ 0.79 %
7 <sup>th</sup> 1,000		146/ 0.53 %
8 <sup>th</sup> 1,000		181/ 0.65 %
9 <sup>th</sup> 1,000		94/ 0.34 %
10 <sup>th</sup> 1,000		65/ 0.23 %
		By this level, cumulated coverage =95.08 %

Table 4 shows that *Reading Matters 3*'s vocabulary distribution among the ranked BNC/COCA word-family lists is more diverse than *Select Readings-intermediate*. We can see a striking difference in vocabulary size between these two books (one for intermediate learners and the other for high-intermediate learners claimed by the publishers), ranging from 2,500—3,000 to 9,000—9,500 word families. This suggests that using *Reading Matters 3* would result in learners working on a lot of vocabulary, some of which they would meet only once or would meet a few times throughout that textbook. Below is a complete list of vocabulary levels across the 25 GE textbooks concerning their vocabulary distribution among the BNC/COCA 1<sup>st</sup>-25<sup>th</sup> 1,000 with the accumulated coverage reaching 95 %.

**Table 5: Vocabulary levels of the GE Textbooks measured by the BNC/COCA 1<sup>st</sup>-25<sup>th</sup> 1,000 at the accumulated coverage reaching 95 %**

GE Textbooks	Vocabulary Levels	Textbook levels claimed by the publishers
Reading Matters3	9000-9500	High-intermediate
World Class Reading3	9000	High-intermediate
Topic for Today5	7000-7500	High-intermediate
Reading for the Real World3	6000-6500	High-intermediate
Active Skills for Reading4	6000	High-intermediate
Mosaic Reading1	6000	High-intermediate
Reading for Success3	6000	Intermediate
Weaving It Together4	6000	High-intermediate
Reading for the Real World2	5500-6000	Intermediate
Hot Topics3	5500-6000	High-intermediate
What a Word3	5000-5500	Intermediate
Concepts for Today4	4500-5000	Intermediate
NorthStar-high intermediate	4500-5000	High-intermediate
Active Skills for reading3	4000-4500	Intermediate
Reading Challenge3	4000-4500	Intermediate
Reading for Real—high intermediate	4000-4500	High-intermediate
Select Readings—upper intermediate	4000-4500	High-intermediate
Concepts & Comments4	4000	High-intermediate
Interactions2	4000	Intermediate
Hot Topics2	3500-4000	Intermediate
Weaving It Together3	3500-4000	Intermediate
World Class Reading2	3500-4000	Intermediate
Cause & Effect3	3000-3500	Intermediate
In Context	3000	Intermediate
Select Readings-intermediate	2500-3000	Intermediate

There are two apparent contradictions in Table 5 about the book levels claimed by the publishers and the vocabulary levels measured by the BNC/COCA 1<sup>st</sup>-25<sup>th</sup> 1,000 word families. Two high-intermediate level GE textbooks, *Reading Matters 3* and *World Class Reading 3*, involved a vocabulary level surpassing 9,000 word families and a wider dispersion of vocabulary along the BNC/COCA word-frequency scale. In contrast, three high-intermediate textbooks *Concepts & Comments 4*, *Select Reading-upper intermediate* and *Reading for Real-high-intermediate* only comprised a vocabulary at the 4,000 and 4,500 word-family levels. Similarly, the vocabulary levels of intermediate textbooks ranged from 2,500 to 6,000 word families. This implies that textbook authors may not have the same standard in the choice of words while writing their teaching materials for the publishers. Some vocabulary presumed difficult by some authors may be considered easy by others.

### Conclusion

The present corpus-based study on GE textbooks has discussed the role that a GE textbook may play in vocabulary learning in two ways. Beyond the 3,000-word-family level, an intermediate to high-intermediate GE textbook can supply students with 327 to 1,211 new word families. This figure gives us a clue about the importance of considering the appropriateness of book levels when choosing college GE textbooks, especially when students' vocabulary size has reached a certain level. If learners' vocabulary is small, they may not enjoy reading due to the enormous vocabulary load. However, if learners' vocabulary is large, learning gains may be small because there are few new words available in the text to learn. As a consequence, more thought should be given to the selection of a GE textbook. Not knowing students' vocabulary capacity in advance, a college professor may choose GE textbooks based on their intuition or publishers' claims. In this research, the book levels claimed by the publishers did not seem to coincide with their bona fide vocabulary levels in a few instances. The aim of this study has been to generate that attention. An intermediate to high-intermediate GE textbook may be useful in preparing learners for the intermediate GEPT, with coverage from 30.49 % to 65.02 % of the vocabulary involved in the test. If students intend to gain a high score on the intermediate GEPT so as to achieve the English graduation benchmarks, they would need to make a determined and continuous effort to expand their vocabulary during the college years. To compensate for insufficient number of class hours, teachers can give reading assignments through the implementation of extensive reading and thereby increase students' reading input. If well-planned, extensive reading may be a feasible approach to bridging the vocabulary gap. Furthermore, students also need to do intensive reading and learn how to use words accurately to deepen their lexical knowledge.

The current textbook corpus was limited by the factor that some GE textbooks may undergo constant revisions, updates or replacement with new articles. Additionally, there are always new book arrivals in the market. As a consequence, the results obtained are not conclusive as a rule of thumb for future selection of the textbooks with the same titles. Further research concerning students' perception towards language textbooks may be worth investigating but are beyond the current focus. It may be fruitful in triangulating analyses from different sources of data. It would also be interesting to examine how books within one series may provide a pathway for language development.

Last but not least, this research was only a preliminary attempt to survey the GE textbooks used in an EFL tertiary setting. Through a corpus-based lexical study, the researcher hopes to draw more attention of instructors and learners to GE courses when choosing a textbook. The aim of this study has been to raise that awareness.

### References

- Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Carver, R.P. (1994). Percentage of unknown words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior*, 26, 413–437.
- Chujo, K. (2004). Measuring vocabulary levels of English textbooks and tests using a BNC lemmatized high frequency word list. In J. Nakamura (Ed.), *English corpora under Japanese eyes* (pp. 231–249). Amsterdam: Rodopi.
- College Entrance Examination Center.(2017). General Scholastic Ability Test (GSAT). Retrieved from <http://www.ceec.edu.tw/>
- Heatley, A., Nation, I.S.P., & Coxhead, A. (2004). RANGE [Computer software]. Retrieved from <https://www.victoria.ac.nz/lals/about/staff/paul-nation>



- Hirsh, D., & Nation, I.S.P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.
- Hu, H.M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From human thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20–34). Cambridge: Cambridge University Press.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Laufer, B., & Sim, D.D. (1985). Measuring and explaining the threshold needed for English academic purposes texts. *Foreign Language Annals*, 18, 405–413.
- Nagy, W., Anderson, R., Schommer, M., Scott, J.A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24, 262–281.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I.S.P. (2012). The BNC/COCA word family lists 25,000. Retrieved from <https://www.victoria.ac.nz/lal/about/staff/paul-nation>
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52, 513–536.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary, description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- West, M. (1953). *A general service list of English words*. London: Longman, Green, & Co.