



College English Textbooks for General Purposes: A Corpus-based Analysis of Lexical Coverage

Wenhua Hsu

(whh@isu.edu.tw)

I-Shou University, Taiwan

Abstract

This study aims to create a corpus of General English (GE) reading textbooks used in universities in Taiwan to form the basis of an analysis. The operational measures for comparison involved vocabulary size, vocabulary levels (distribution among the British National Corpus 1st–14th 1,000 high-frequency word families) and text coverage. Coxhead's (2000) Academic Word List (AWL) containing 570 word families was chosen as one of the base word lists. In addition, the Grades 1–9 Curriculum 2,000 basic English words required by Taiwan's Ministry of Education as well as the elementary and intermediate vocabulary covered in the General English Proficiency Test (GEPT) were lemmatized into word families, and then added to the base words, the BNC high frequency word lists and the AWL established in Nation's RANGE (n.d.) software. The GEPT is the accredited English proficiency test in Taiwan that college students are likely to encounter as an English graduation benchmark and the language requirement for the job market. The results show that a GE textbook can contribute to learning 49–415 interdisciplinary academic words. Beyond the 2,000-word level, a GE textbook can supply students with 162–2,001 new word families. It may be useful in preparing learners for an intermediate GEPT by covering 24.55% to 65% of the vocabulary involved in the test. It is hoped that the indices examined in this study would help English teachers to take into account vocabulary size and levels in curriculum design.

1 Introduction

English education in Taiwan starts officially in the third year of elementary school, though some private schools may begin English programs as early as in the first grade. During primary and secondary education (Grades 1–12), English courses aim to familiarize learners with basic English sentence structures and the most commonly used words. The content of English texts is broadly humanities-based and teaching is geared towards the general interest of students rather than to specific purposes.

The 2,000 basic English word list was published by Taiwan's Ministry of Education in 2003. It was developed based on the program design used by high schools in Taiwan, Korea, Japan and Shanghai as well as West's (1953) General Service List (GSL) of English words. Since then, the 2,000 basic English word list has served as a curricular standard for the English course design for elementary and high schools as well as cram schools. By and large, the 2,000 lexical items are presumed to be the minimum vocabulary of EFL high school graduates entering university.

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. In recent years, one widely implemented core educational policy in college English instruction was the adoption of English proficiency as a graduation requirement. This has provoked heated debate about the impact of English proficiency benchmarks on college English curricula. Some college English teachers speculate that test-oriented graduation benchmarks of

English proficiency will jeopardize any normal college English instruction and turn the classroom into a cram school. 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, business, medicine, science and technology to achieve an all-encompassing development of knowledge. Conversely, those who support English graduation criteria argue that testing is a necessary evil, especially when increasing the number of English credit hour requirement for students may not help in improving their English proficiency. Because of academic demands for English abilities and the language requirement for the job market, many universities/colleges¹ set graduation benchmarks of English language proficiency for their students.

Among various English proficiency tests used in institutions of higher education, 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 and even 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 commissioned by Taiwan's Ministry of Education and developed by the Language Training & Testing Center (LTTC) in 1999 (cf. <http://www.lttc.ntu.edu.tw>) especially for EFL learners at all levels of proficiency. GEPT has become a household name since it was first administered in 2002.

The reliability of the GEPT scores is widely accepted by the public. A number of studies related to the GEPT have been conducted by the LTTC, for example, on parallel-form reliability, test form and task comparability (Weir & Wu, 2002, 2006), and relating the GEPT reading comprehension tests to the Common European Framework of Reference for Languages (Wu & Wu, 2007). In Roever and Pan's (2008) introduction to the GEPT, reliability statistics show reliabilities mostly in the high .8 range, similar to the reliability figures of other large-scale test batteries. The test is therefore recognized by government agencies as a criterion for promotion. It is also used by the Ministry of Education and Academia Sinica in Taiwan as a means of evaluating the English abilities of applicants to 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 admissions, placement or graduation.

Currently, four levels of the GEPT are regularly administered: elementary, intermediate, high-intermediate and advanced. A fifth level, the superior, was administered only once and then suspended, pending further need. The GEPT elementary level is presumed to be appropriate for students who have studied English through junior high school (Grades 7–9). The GEPT intermediate level is seen as suitable for senior high school graduates (Grades 10–12) or university freshmen. The GEPT high-intermediate level is thought to be suitable for university graduates majoring in English. The GEPT advanced level is considered adequately difficult such that only someone with a graduate degree from a university in an English-speaking country 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. Examinees who pass the first stage are allowed to register for the second stage, with the speaking and writing sections of the test. In a range of English proficiency levels for graduation benchmarks stipulated by some universities/colleges for their students, the GEPT intermediate-level first stage (i.e. passing its listening and reading test) is widely accepted.

According to the Language Training & Testing Center, 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 (i.e. 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 (Wikipedia², n.d.). For English-major 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. (Wikipedia, n.d.). 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 (<http://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, 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 and materials used in the freshman and sophomore years may therefore play an important role in enhancing English abilities.

2 Literature Review

In light of the potential role of English for General Purposes 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 the skills of listening, speaking, reading and writing easier to perform. The limited vocabulary of EFL learners is a major source of difficulty in reading an English text.

West (1926) considered “one unknown word in every fifty words” to be the minimum threshold necessary for the adequate comprehension of a text (cited in Chujo, 2004, p. 231). That is, one needs to know sufficiently different words (types) to account for 98% of the running words (tokens) in a text. Native English-speaking children view a vocabulary load of two unknown words per hundred words (i.e. 98% lexical coverage³) as difficult reading (Carver, 1994). More specifically, 98% coverage is equivalent to roughly one unknown word for every five lines of text. Some researchers regard one unknown word in every twenty words (95% lexical coverage of a text) 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%), 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). If 95–98% coverage of a text 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.

Taking into consideration the above studies on lexical coverage (i.e. 95–98%), it can be concluded that as the density of unknown words increases, reading comprehension drops. Applying this to a test where students are unassisted while reading or listening to its content, it can be inferred that vocabulary size may be one of the predictors of test scores. As such, what is the optimum vocabulary goal at the tertiary level, if 95% lexical coverage is sustained? Namely, how large a vocabulary is needed for a graduating college EFL student?

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, 1997). In his recent study, Nation (2006) noted that if 98% coverage of a text 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. Earlier studies such as Carroll, Davies and Richman (1971) reported that the top 2,000 most frequent English words translate into a roughly 80% coverage for a longer text and the 5,000 most frequent words as high as 90%. Accordingly, to gain a lexical coverage of 95%, one needs to know some 12,000 words, which is much higher than Nation’s (2006) findings. A well-educated adult

native speaker of English has a vocabulary of around 17,000 words (Goulden, Nation, & Read, 1990).

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 1st–14th 1,000 high-frequency word lists (BNC HFWL), 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 an understanding of the top 6,000–6,500 most frequently-occurring words in the BNC HFWL are more likely to achieve 95% text comprehension in TOEFL than others with knowledge of fewer than 6,000 English words. Similarly, to get above average scores in TOEIC, one still needs to command a vocabulary of at least 4,500–5,000 words.

According to Nation (2001), words in non-fiction texts can be divided into four categories: (1) high-frequency or general service vocabulary, (2) academic vocabulary (also called sub-technical or semi-technical vocabulary), (3) technical vocabulary and (4) low-frequency vocabulary. High-frequency words refer to those basic general service English words which constitute the majority of all the running words in all types of writing. The most well-known general service vocabulary is West's (1953) General Service List of English Words (GSL). The GSL containing the most frequently-occurring 2,000 word families of English (3,372 word types) accounts for approximately 75% of the running words in non-fiction texts and around 90% of the running words in fiction (Hirsh & Nation, 1992; Nation & Hwang, 1995). Technical words are the ones used in a specialized field and are considerably different from subject to subject. About 5% of the words in an academic text are made up of technical vocabulary, with each subject containing roughly 1,000 word families (Nation, 2001). In an academic setting, ESP students do not see technical terms as a problem because these terms are usually the focus of discussion in class or in the specialist textbooks. Low-frequency words are rarely used terms. Academic vocabulary with medium-frequency of occurrence across texts of various disciplines (i.e. somewhere between the high-frequency words and technical words) has some rhetorical functions and communicative purposes. Acquiring these academic words (sub-technical vocabulary) seems to be essential when learners are preparing for EAP or ESP.

Coxhead (2000) compiled a corpus of around 3.5 million running words from university textbooks and materials from four different academic areas (law, arts, commerce as well as science), and identified 570 academic word families, which were claimed to cover almost 10% of the total words in a general academic text. Her research suggested that for learners with academic goals, the academic word list contains the next set of vocabulary to learn after the top 2,000-word level. To put it concretely, after the top 2,000 word families on a frequency list, greater text coverage is gained by moving on to the 570 academic words (10% coverage) than by continuing to learn the next 1,000 words ("3–5%" coverage for the 3rd 1,000; Nation, 2006, p. 79). However, there is still a great discrepancy in the vocabulary capacities of an EFL learner and an English-native speaker with a vocabulary of 12,000–17,000 words, as mentioned.

Compared to 12,000–17,000 words, the requirement for a vocabulary size of 5,000–6,000 words in the current EFL context appears to be a more feasible goal for college teachers in assisting their students to meet vocabulary thresholds.

3 Method

The present study was undertaken to examine the vocabulary of General English textbooks used in colleges/universities in Taiwan. How can the goal of increasing vocabulary size to a particular target level be achieved in the classroom under real class conditions? What interdisciplinary academic vocabulary can freshman and sophomore English courses for general purposes provide through diverse and versatile content? What additional vocabulary is required for students under the assumption that 95% text comprehension is the threshold for passing an English proficiency test? If college GE textbooks fall short of the targets above, English instructors must then provide

supplementary materials to bridge the gap. By lexically comparing textbooks, this research sought to answer the following questions:

1. What percentage of the words in a General English reading textbook does Coxhead's (2000) Academic Word List cover? How many interdisciplinary academic words may one learn from a GE textbook?
2. If a senior high school graduate has a vocabulary size of the 2,000 basic English words required by Taiwan's Ministry of Education, how many new words may one learn from a GE textbook?
3. 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?
4. What is the vocabulary level of a GE textbook?

3.1 Textbook selection criteria

Since this study aimed to create a corpus of GE reading textbooks widely used in universities in Taiwan to form the basis of the analysis, the criteria for the inclusion of the books in the corpus were based on the popularity of GE textbooks according to sales data from eight major import bookstores. Among college GE textbooks ranging from low-intermediate to advanced levels, intermediate and high-intermediate level textbooks are commonly used. In total, thirty-six textbooks were chosen, five low-intermediate, thirteen intermediate, twelve upper-intermediate and six advanced (see Appendix A). Excluding exercises and supplementary readings, the main articles in each book chosen were scanned into thirty-six computer files, manually typed for some texts with illustrations, and proofread to ensure text completeness. 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. Proper nouns were separated from the counting of normal words for the following reasons:

1. The meanings of most proper nouns (well-known or not well-known) can be inferred from the context of a text and knowledge of them may be easily translated or gained through one's mother tongue.
2. Various types of text may contain different percentages of proper nouns. Proper nouns are not in the list of the most frequent 2,000 words. If proper nouns are included in the statistics of text coverage and vocabulary size, the results presented either in word types or word families may be distorted due to an unequal basis of comparison. To avoid such a bias, proper nouns were eliminated.

After removing proper nouns, the resulting corpus contained in total 617,927 tokens (running words), as Table 1 shows.

	GE Textbook	Book Level	Number of Texts	Tokens
1	Access Reading 3	low-intermediate	20	8,541
2	What a World 2	low-intermediate	18	11,337
3	What a World 3	intermediate	16	16,616
4	In Context-Steps to academic Reading	intermediate	38	24,304
5	Active Skills for Reading 3	intermediate	32	18,798
6	Active Skills for Reading 4	high-intermediate	32	26,962
7	Cause & Effect-reading 3	intermediate	25	16,975
8	Concepts & Comments-reading 4	high-intermediate	20	15,383
9	Hot Topics 2	intermediate	53	16,384
10	Hot Topics 3	high-intermediate	43	22,403
11	Interactions 1 for reading	low-intermediate	20	14,110
12	Interactions 2 for reading	intermediate	10	9,574
13	Mosaic Reading 1	high-intermediate	20	17,618
14	Mosaic Reading 2	advanced	34	27,512
15	NorthStar –high intermediate	high-intermediate	20	16,221
16	NorthStar focus on reading-advanced	advanced	20	17,323
17	Reading Challenge 2	low-intermediate	20	6,709
18	Reading Challenge 3	intermediate	20	7,927
19	Reading for Real-high-intermediate	high-intermediate	12	8,985
20	Reading for Real-advanced	advanced	12	10,014
21	Reading for Success 2	low-intermediate	32	11,505
22	Reading for Success 3	intermediate	32	18,566
23	Reading for the Real World 2	intermediate	24	15,587
24	Reading for the Real World 3	high-intermediate	24	16,114
25	Reading Matters 3	high-intermediate	51	44,199
26	Reading Matters 4	advanced	31	36,409
27	Select Readings-intermediate	intermediate	14	10,204
28	Select Readings-upper-intermediate	high-intermediate	12	15,277
29	Weaving It Together 3	intermediate	16	11,951
30	Weaving It Together 4	high-intermediate	18	16,805
31	World Class Reading 2	intermediate	14	8,968
32	World Class Reading 3	high-intermediate	14	15,411
33	Concepts for Today 4	intermediate	12	10,129
34	Topic for Today 5	high-intermediate	25	20,891
35	Tapestry Reading 4	advanced	36	29,474
36	College Reading 4	advanced	18	22,741
Excluding proper nouns, total tokens				617,927

Table 1 is arranged in ascending order according to the book level, or within the same series if there is one.

Table 1: Composition of a corpus of General English Reading textbooks

3.2 *The instrument*

Heatley, Nation and Coxhead (n.d.) created RANGE and FREQUENCY programs which incorporated the General Service List of English Words (GSL), Academic Word List (AWL) and British National Corpus High Frequency Word List (BNC HFWL) based on English words' occurring frequency, range and dispersion figures. The RANGE software can be used to compare a text against certain base word lists to see what words in the text are and are not in the lists, and to see what percentage of the vocabulary items in the text are covered by the lists, namely text coverage⁴. In particular, it can also be used to compare the vocabulary of many text files at a time to see how much of the same vocabulary they use (i.e. range) and the frequency of occurrence of the words in total and in each file.

To compare GE textbooks, the operational measures therefore involved vocabulary size, vocabulary levels (distribution among the BNC 1st–14th 1,000 high-frequency words) and text coverage.

3.3 *Choosing and creating base word lists*

Two categories of word lists were downloaded from the RANGE program (see below for the first two), and three word-family lists (the last three below) needed editing for this research. They were:

1. The BNC 1st–14th 1,000 high-frequency English word lists.
2. Coxhead's (2000) Academic Word List (AWL) containing 570 word families.
3. The 2,000 basic English words announced by Taiwan's Ministry of Education.
4. The GEPT elementary word list consisting of 2,263 words.
5. The GEPT intermediate word list with elementary vocabulary inclusive (totaling 4,947 words).

The GSL 2,000-word families available in RANGE as a base word list for this research did not seem to suffice in measuring vocabulary size and levels if a higher-education textbook included more advanced academic articles of a 3,000-word level or higher. As a result, 14,000 high-frequency word families, which were created from the British National Corpus and already built into the RANGE program, were adopted. The British National Corpus (BNC) with more than 100 million words is considered one of the largest corpora of present-day English usage in speech and in publications in the United Kingdom (Leech, Rayson, & Wilson, 2001). The 14,000 high-frequency word families were divided into fourteen base word lists, each containing exactly 1,000 word families. Apart from the BNC 1st–14th 1,000 word lists, some proper nouns and Roman numerals as well as spoken interjections and exclamations were also incorporated in the RANGE program, appearing as base word lists 15 and 16. Base list 15 (a proper noun list) and base list 16 (an interjection and exclamation list) were beyond the research focus and were hence not factored in.

Coxhead's (2000) 570 academic word-family list has also been included in the RANGE program. This list was adopted for the present study to measure how frequently the academic words across disciplinary domains occur in a GE textbook and to compare the extent to which a GE textbook can prepare a learner for reading professional texts containing such sub-technical vocabulary.

As mentioned above, in 2003, Taiwan's Ministry of Education set a vocabulary goal of 2,000 basic English words (referred to hereafter as TBEWL 2000) for primary and secondary schools (Grades 1–12). It was therefore presumed in this research that freshmen who have passed the college entrance exam had a vocabulary size of at least 2,000 words. In order to have the same word-family building basis for comparison as the AWL and BNC 1st–14th 1,000 word lists in the RANGE program, the TBEWL 2000 headwords were lemmatized into word families following Nation's (2001) principles. The principles used in the RANGE software to make word families were based on Bauer and Nation's (1993) six-level scale, which includes all the affixes from levels 2 to 6.

- | | |
|---------|--|
| Level 2 | Regularly inflected words are part of the same family. The inflectional categories are - plural; third person singular present tense; past tense; past principles; -ing; comparative; superlative; possessive. |
| Level 3 | -able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, all with restricted uses. |
| Level 4 | -al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-, all with restricted uses. |
| Level 5 | -age (leakage), -al (arrival), -ally (idiotically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom; officialdom), -eer (black marketeer), -en (wooden), -en (widen), -ence (emergence), -ent (absorbent), -ery (bakery, trickery), -ese (Japanese; officialese), -esque (picturesque), ette (usherette, roomette), -hood (childhood), -i (Israeli), -ian (phonetician; Johnsonian).... |
| Level 6 | -able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-. |

(Nation, 2001, p. 268)

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 meanings 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). For instance, the headword *ache* and its complete family members are *ache*, *aches*, *ached*, *aching* and *achy*. From the example *ache*, a word family consists of a headword, its inflected forms and its closely related derived forms, even if the part of speech is not the same (Nation, 2001).

The notion is useful when we are concerned with vocabulary size in a reading textbook. The 2,000 most frequent words on the BNC HFWL include 2,000 base forms, their inflected forms and derivatives, thereby making a total of 11,941 different words (types).

Another two base word lists replacing the BNC high-frequency word lists pre-built into the RANGE were the elementary and intermediate vocabulary involved in the GEPT (2,263 words at the elementary and 4,947 words at the intermediate level, including the previous 2,263 words). They were also made into word families following the same procedure as above. Subsequently, to calculate text coverage and to assess the vocabulary levels of the thirty-six GE textbooks, the RANGE computing program was run each time against one of the following five base word lists, using the same counting units, word types and word families: (1) AWL 570, (2) BNC HFWL 1st–14th 1000, (3) TBEWL 2000, (4) GEPT elementary word list, and (5) GEPT intermediate word list.

4 Results and discussion

4.1 The number and coverage of interdisciplinary academic words in a GE textbook

According to Coxhead (2000), the 570 academic word families account for approximately 10% of the tokens in academic texts but only 1.4% of the total words in a fiction collection of the same size. The most frequent word families in the AWL are, for example, *approach*, *constitute*, *identify*, *indicate*, *interpret*, *specific*, *vary* and so on. Some of the least frequent vocabulary among the 570 academic words across disciplines is *adjacent*, *conceive*, *collapse*, *incline*, *persist*, *whereby* and *notwithstanding* etc. The middle column in Table 2, which shows the AWL occurrence in tokens and in percentage, addresses the first research question, “What percentage of the words in a General English reading textbook does Coxhead’s (2000) Academic Word List cover?” For instance, the textbook *Reading Matter 4* contained 36,409 running words after deleting proper nouns, with 785 different types of academic words appearing in the texts 1,980 times. The 1,980 occurrences arising out of 415 academic word families made up 5.44% of the total words in the textbook. In this corpus, *Reading Matters 4* included the most interdisciplinary academic words, having 415 word families out of the AWL 570.

Taking a further look at Table 2, one can see that none of the GE textbooks in the corpus seemed to meet the figure in the literature, i.e. 10% coverage counted in tokens with regard to the commonly useful academic words that students may come across in reading professional articles. The top five textbooks of higher AWL coverage were *College Reading 4* (7.3% in tokens; 14.41% in word types), *Reading for the Real World 3* (6.54% in tokens; 15.25% in types), *Reading for the Real World 2* (5.93% in tokens; 13.55% in types), *Hot Topics 3* (5.71% in tokens; 12.48% in types) and *Reading Matter 4* (5.44% in tokens; 12.82% in types). Four of these five books were labeled by the publishers as suitable for high-intermediate and advanced EFL learners (see Table 1 for book levels). *What a World 2* for low-intermediate learners had the lowest AWL coverage, with 1.3% coverage counted in tokens and 3.76% in types. It is not surprising to see that the higher AWL coverage and the higher-level GE textbooks bore some relationship to each other, since advanced textbooks are meant to teach advanced vocabulary and to expand students’ vocabulary breadth. In a similar vein, it is not disappointing to find that basic GE textbook contained little academic vocabulary since lower-level books serve other purposes in language learning.

General English Textbooks	Number of words	AWL occurrence in tokens/ %	AWL occurrence in types/ %	AWL occurrence in families
Reading Matters 4	36,409	1980/ 5.44%	785/ 12.82%	415
Reading Matters 3	44,199	2110/ 4.77%	776/ 11.09%	408
Tapestry Reading 4	29,474	1590/ 5.39%	671/ 11.83%	394
Active Skills for Reading 4	26,962	1249/ 4.63%	615/ 12.23%	363
Mosaic Reading 2	27,512	1247/ 4.53%	605/ 10.41%	361
College Reading 4	22,741	1659/ 7.30%	607/ 14.41%	351
Topic for Today 5	20,891	1116/ 5.34%	509/ 11.77%	327
Reading for the Real World 3	16,114	1054/ 6.54%	519/ 15.25%	313
Hot Topics 3	22,403	1279/ 5.71%	507/ 12.48%	312
Reading for the Real World 2	15,587	924/ 5.93%	477/ 13.55%	310
Reading for Success 3	18,566	877/ 4.72%	431/ 11.22%	292
Mosaic Reading 1	17,618	660/ 3.75%	398/ 10.45%	275
Active Skills for Reading 3	18,798	1000/ 5.32%	444/ 12.47%	270
Weaving It Together 4	16,805	723/ 4.30%	403/ 10.46%	267
Select Readings-hi-intermediate	15,277	609/ 3.99%	371/ 11.00%	266
World Class Reading 3	15,411	815/ 5.29%	387/ 10.93%	260
NorthStar reading-advanced	17,323	581/ 3.35%	360/ 9.75%	254
Concepts for Today 4	10,129	530/ 5.23%	339/ 12.81%	237
NorthStar –high intermediate	16,221	508/ 3.13%	292/ 8.73%	211
Select Readings-intermediate	10,204	472/ 4.63%	280/ 12.29%	211
Concepts & Comments- reading 4	15,383	607/ 3.95%	303/ 10.85%	202
Hot Topics 2	16,384	701/ 4.28%	311/ 10.63%	198
In Context	24,304	943/ 3.88%	291/ 8.84%	183
Reading for Success 2	11,505	374/ 3.25%	232/ 8.49%	174
Reading for Real-advanced	10,014	305/ 3.05%	220/ 8.40%	166
What a World 3	16,616	416/ 2.50%	208/ 7.17%	158
Reading Challenge 3	7,927	342/ 4.31%	215/ 10.98%	155
Interactions 1 for reading	14,110	632/ 4.48%	221/ 8.45%	151
Interactions 2 for reading	9,574	474/ 4.95%	219/ 10.51%	149
Reading for Real hi-intermediate	8,985	235/ 2.62%	183/ 8.01%	149
World Class Reading 2	8,968	291/ 3.24%	193/ 8.47%	148
Weaving It Together 3	11,951	276/ 3.29%	167/ 7.54%	130
Cause & Effect-reading 3	16,975	372/ 2.19%	163/ 6.49%	116
Reading Challenge 2	6,709	223/ 3.32%	118/ 7.20%	91
Access Reading 3	8,541	298/ 3.49%	102/ 6.82%	80
What a World 2	11,337	147/ 1.30%	72/ 3.76%	49

Thirty-six textbooks are tabulated in descending order according to the number of occurrence of academic word families.

Table 2: The AWL coverage of GE textbooks

Given the assumption that the 570 interdisciplinary academic words are all new to college freshmen, Table 3 demonstrates that the academic lexical items that can be learnt by using one of the thirty-six GE textbooks range from as little as 49 to as much as 415 word families. This result produces an overall picture of the amount of academic vocabulary a student will be equipped with after taking a GE course for one year.

Below is a list of the most and the least frequently-occurring academic headwords in the corpus of GE textbooks (see Table 3).

Academic headword	Range (across the number of textbooks)	Headword occurrences	Word family occurrences
create	36	126	438
final	36	44	220
area	35	188	314
job	34	244	371
major	34	135	176
similar	34	124	150
adjacent	1	1	1
allocate	1	1	1
coincide	1	1	2
discrete	1	1	1
denote	1	1	2
forthcoming	1	1	1
hierarchy	1	1	1
infer	1	1	1
notwithstanding	1	1	1
overlap	1	1	1
paradigm	1	1	2
parameter	1	1	1
simulate	1	1	1
scope	1	1	1
regime	1	1	1

Table 3: The most and the least frequent academic headwords in thirty-six GE textbooks

In the present corpus, two academic words *create* and *final* as well as their family members, occurred in all the 36 GE textbooks. The word *area* as well as the three words, *job*, *major* and *similar*, appeared in 35 and 34 textbooks respectively. Words such as *adjacent*, *discrete* and *regime*, though among the interdisciplinary academic words, only showed up once in one GE textbook.

As can be seen from Table 3, the word, *create*, by itself appeared in all 36 textbooks, with a total of 126 occurrences. When its family members (*created*, *creates* and *creating*, etc.) were included, they showed up as many as 438 times. Words like *create*, *job* and *similar* do not show an academic flavor in a strong sense. These words appear to be common language components that slip invisibly in and out of everyday conversation, content domain talk and specialized reading materials. It is highly possible that EFL students may have encountered and learnt these words elsewhere before. On the other hand, least frequent words such as *parameter*, *allocate* and *hierarchy* reveal a sense of content domain, for example, in the field of business and in academic research papers.

Coxhead (2000) suggested that learners with a good command of the top 2,000 most frequent words need to consider their intended use of English. If they intend to use English for a wide range of purposes, they need to continue learning high-frequency vocabulary. If, however, learners intend to further their academic studies, their vocabulary learning should go in a different direction. After the 2,000-word level, academic vocabulary gives a better return in relation to learning efforts for those learners pursuing academic interests. Taken together, the Academic Word List containing 570 word families could be used to set vocabulary learning goals for GE courses as a launch pad for further English for academic purposes programs. For instance, if 50% of the headwords in the AWL are expected to be taught within an academic year, then a GE textbook should ideally include at least 285 predominantly academic words. If properly selected or managed, a GE textbook can contribute to the goal of academic vocabulary learning.

4.2 The number of new words contained in a GE textbook beyond the 2,000-word level

As far as the 2,000-word level is concerned, the word lists, BNC HFWL 2000, TBEWL 2000 and GEPT elementary vocabulary should be compared first. The three word lists are summarized in Table 4 for easy comparison.

Word List	Number of lexical items contained	Number of word types extended	Number of word families made	Word family overlap with BNC HFWL 2000/ resemblance (in %)	TBEWL2000 and GEPT elementary overlap/ resemblance (in %)
BNC HFWL2000	2,000	11,941	2,000	X	
1 st 1,000	1,000	6,348	1,000		
2 nd 1,000	1,000	5,593	1,000		
TBEWL 2000	2,000	10,379	1,963	1,282/ 64.7%	1,925/ 97.64%
GEPT elementary	2,263	10,554	1,980	1,387/ 69.7%	

Table 4: Comparison among the Word Lists, BNC HFWL 2000, TBEWL 2000 and GEPT elementary

BNC HFWL 2000 and TBEWL 2000 contain 2,000 lexical items each, while the GEPT elementary word list includes 2,263 headwords. After extending to word types, the BNC HFWL 2000 has 2,000 word families with 11,941 different words (types) involved. The numbers of the word families made from TBEWL 2000 and GEPT elementary word list are very close (1,963 versus 1,980) and the numbers of word types expanded from these two word lists are also approximately equal (10,379 vs. 10,554). The overlap of the headwords between the TBEWL 2000 and the GEPT elementary word list is 1,925 words, while resemblance between the TBEWL 2000 and the BNC HFWL 2000 is 64.7%, with 1,282 word families occurring in both [$1282 * 2 \div (1963 + 2000) = 64.7\%$]. Likewise, in comparing headwords with headwords, the coverage of GEPT elementary vocabulary in the BNC HFWL 2000 is 69.7%, with an overlap of 1,387 word families [$1387 * 2 \div (1980 + 2000) = 69.7\%$].

In contrast, when TBEWL 2000 and GEPT elementary vocabulary are compared, the overlap is as high as 97.64%, with both lists containing the same 1,925 word families [$(1925 * 2) / (1963 + 1980) = 97.64\%$]. Those in the TBEWL 2000 that are not in the BNC HFWL 2000 are words closely related to daily life in Taiwan and involving some Taiwan-specific culture. These words are, for instance, *chopsticks*, *dumpling*, *mango* in food; *typhoon*, *air conditioner* in climate and temperature; *dragon*, *panda*, *cockroach* in animals and insects and so on, showing a local color. A significantly high overlap between TBEWL 2000 and the GEPT elementary vocabulary can be partly ascribed to the fact that both lists of vocabulary were particularly developed for EFL learners in Taiwan and are somewhat test-oriented.

Despite the fact that the three word lists have similar numbers of lexical items and word families generated from within (i.e. about 2,000 word families), their comparison raised another concern, "Which base word list performs best across the thirty-six GE textbooks in terms of text coverage?" Alternatively, assuming that a newly matriculated college student has the vocabulary of TBEWL 2000, BNC HFWL 2000 or GEPT elementary vocabulary, with which 2,000-word level can one gain a higher percentage of reading comprehension in an English text for general purposes?

Table 5 demonstrates that in comparison with the other two 2000-word level lists, the BNC HFWL 2000 seemed to offer higher text coverage and hence a good command of it enables better comprehension of a GE text. BNC HFWL 2000 provides a consistently higher percentage of text coverage across all the GE textbooks than the other two lists. Though TBEWL 2000 and GEPT elementary words bore a close resemblance to each other (97.64%), in general terms, TBEWL 2000 was inferior to GEPT elementary vocabulary in text coverage. To explore Research Question 2, "If a senior high school graduate has a vocabulary size of the 2,000 basic English words required by Taiwan's Ministry of Education, how many new words may one learn from a GE textbook?", it was decided that TBEWL 2000 would serve as the computing basis for analysis, since it

represents the average level of a college freshman's vocabulary capacity in accordance to the requirement of Taiwan's Ministry of Education.

Level	GE Textbooks	TBEWL2000 % coverage (in tokens)	GEPT elemen- tary % coverage	BNC HFWL2000 % coverage
low-intermediate	Access Reading 3	91.21 %	91.98 %	92.80 %
intermediate	Active Skills for Reading 3	85.70 %	87.57 %	90.33 %
high-intermediate	Active Skills for Reading 4	83.48 %	85.45 %	88.29 %
intermediate	Cause & Effect 3	90.18 %	91.62 %	92.18 %
advanced	College Reading 4	79.39 %	82.05 %	87.29 %
high-intermediate	Concepts & Comments 4	84.89 %	86.90 %	89.57 %
intermediate	Concepts for Today 4	84.41 %	86.25 %	89.58 %
intermediate	Hot Topics 2	87.95 %	89.39 %	92.04 %
high-intermediate	Hot Topics 3	83.48 %	84.96 %	88.86 %
intermediate	Interactions 2 reading	86.37 %	88.52 %	90.61 %
low-intermediate	Interactions 1 reading	88.05 %	89.73 %	91.41 %
high-intermediate	Mosaic Reading 1	84.82 %	86.58 %	88.59 %
advanced	Mosaic Reading 2	80.10 %	82.54 %	85.22 %
advanced	NorthStar-advanced	86.72 %	87.98 %	90.10 %
high-intermediate	NorthStar- high intermediate	86.89 %	87.87 %	90.12 %
low-intermediate	Reading Challenge 2	88.67 %	90.40 %	91.62 %
intermediate	Reading Challenge 3	84.82 %	87.22 %	89.20 %
advanced	Reading for Real-advanced	85.58 %	87.46 %	89.41 %
high-intermediate	Reading for Real-hi- intermediate	87.67 %	88.98 %	90.20 %
low-intermediate	Reading for Success 2	85.44 %	87.25 %	87.84 %
intermediate	Reading for Success 3	83.16 %	85.18 %	87.29 %
intermediate	Reading for the Real World 2	81.16 %	84.15 %	87.61 %
high-intermediate	Reading for the Real World 3	79.28 %	81.76 %	86.39 %
high-intermediate	Reading Matters 3	81.92 %	83.98 %	86.93 %
advanced	Reading Matters 4	82.21 %	84.45 %	88.21 %
intermediate	Select Readings-intermediate	89.10 %	90.91 %	93.30 %
high-intermediate	Select Readings-hi- intermediate	85.54 %	87.33 %	89.83 %
intermediate	In Context	89.80 %	91.54 %	92.28 %
advanced	Tapestry Reading 4	82.12 %	84.25 %	87.16 %
high-intermediate	Topic for Today 5	81.67 %	83.45 %	86.78 %
high-intermediate	Weaving It Together 4	82.48 %	84.17 %	87.22 %
intermediate	Weaving It Together 3	88.54 %	89.92 %	90.14 %
low-intermediate	What a World 2	88.65 %	89.89 %	89.24 %
intermediate	What a World 3	86.22 %	88.26 %	88.43 %
intermediate	World Class Reading 2	87.59 %	89.06 %	89.92 %
high-intermediate	World Class Reading 3	81.07 %	83.28 %	85.72 %

Table 5 is arranged in ascending order according to the book level, or within the same series if there is one.

Table 5: Text coverage by the three word lists, TBEWL 2000, GEPT elementary and BNC HFWL 2000

As can be seen from Table 6, the columns, *the TBEWL 2000 base word list* and *not in the TBEWL2000*, present how much of the vocabulary in a GE textbook will be familiar to the learner and how many words the learner may not know. For example, 7,790 of the running words in *Access Reading 3* were in the TBEWL2000 list, while 751 tokens were not. These 751 occurrences involving 324 different words (types) could possibly be college freshmen's new words. By and large, if a college student reads all the texts of a GE course book within an academic year, he/she would come across 324-4,003 new words beyond the 2,000-word level, 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 for word type to word family range from 1.54 to 2.18. This broad range of 324 to 4,003 new words (corresponding to roughly 162 to 2,001 word families, see Note⁵) implies that the vocabulary learning goals subject to the selection of a GE textbook are quite wide-ranging. By learning vocabulary at the rate of 162 new headwords (and hence their family members) per academic year over the college course of four years, it does not seem likely that students can achieve the 5,000-word threshold level as proposed by Laufer (1997).

GE Textbook	TBEWL 2000		Not in the TBEWL2000		Total	
	tokens/ %	types/ %	tokens/ %	types/ %	tokens	types
Access Reading 3	7790/ 91.21	1172/ 78.34	751/ 8.97	324 / 21.66	8541	1496
Active Skills for Reading 3	16110/ 85.70	1999/ 56.15	2688/ 14.30	1561/ 43.85	18798	3560
Active Skills for Reading 4	22509/ 83.48	2505/ 49.80	4453/ 16.52	2525/ 50.20	26962	5030
Cause & Effect 3	15308/ 90.18	1769/ 70.45	1667/ 9.82	742/ 29.55	16975	2511
College Reading 4	18053/ 79.39	2100/ 49.87	4688/ 20.61	2111/ 50.13	22741	4211
Concepts & Comments 4	13059/ 84.89	1671/ 59.83	2324/ 15.11	1122/ 40.17	15383	2793
Concepts for Today 4	8550/ 84.41	1532/ 57.90	1579/ 15.59	1114/ 42.10	10129	2646
Hot Topics 2	14410/ 87.95	1829/ 62.49	1974/ 12.05	1098/ 37.51	16384	2927
Hot Topics 3	18702/ 83.48	2158/ 53.11	3701/ 16.52	1905/ 46.89	22403	4063
Interactions 2	8269/ 86.37	1359/ 65.24	1305/ 13.63	724/ 34.76	9574	2083
Interactions 1	12424/ 88.05	1726/ 66.03	1686/ 11.95	888/ 33.97	14110	2614
Mosaic Reading 1	14943/ 84.82	2025/ 53.16	2675/ 15.18	1784/ 46.84	17618	3809
Mosaic Reading 2	22036/ 80.10	2586/ 44.52	5476/ 19.90	3223/ 55.48	27512	5809
NorthStar-advanced	15022/ 86.72	2061/ 55.84	2301/ 13.28	1630/ 44.16	17323	3691
NorthStar-high intermediate	14094/ 86.89	1935/ 57.86	2127/ 13.11	1409/ 42.14	16221	3344
Reading Challenge 2	5949/ 88.67	1226/ 74.85	760/ 11.33	412/ 25.15	6709	1638
Reading Challenge 3	6724/ 84.82	1277/ 65.22	1203/ 15.18	681/ 34.78	7927	1958
Reading for Real-advanced	8570/ 85.58	1540/ 58.82	1444/ 14.42	1078/ 41.18	10014	2618
Reading for Real-high intermediate	7877/ 87.67	1432/ 62.70	1108/ 12.33	852/ 37.30	8985	2284
Reading for Success 2	9830/ 85.44	1599/ 58.49	1675/ 14.56	1135/ 41.51	11505	2734
Reading for Success 3	15439/ 83.16	1956/ 50.94	3127/ 16.84	1884/ 49.06	18566	3840
Reading for the Real World 2	12650/ 81.16	1923/ 54.63	2937/ 18.84	1597/ 45.37	15587	3520
Reading for the Real World 3	12775/ 79.28	1765/ 51.85	3339/ 20.72	1639/ 48.15	16114	3404
Reading Matters 3	36206/ 81.92	2994/ 42.79	7993/ 18.08	4003 / 57.21	44199	6997
Reading Matters 4	29932/ 82.21	2791/ 45.60	6477/ 17.79	3330/ 54.40	36409	6121
Select Readings-intermediate	9092/ 89.10	1487/ 65.28	1112/ 10.90	791/ 34.72	10204	2278
Select Readings-high intermediate	13068/ 85.54	1848/ 54.77	2209/ 14.46	1526/ 45.23	15277	3374
In context	21826/ 89.80	2104/ 63.93	2478/ 10.20	1187/ 36.07	24304	3291
Tapestry Reading 4	24205/ 82.12	2636/ 46.48	5269/ 17.88	3035/ 53.52	29474	5671
Topics for Today 5	17061/ 81.67	2060/ 47.65	3830/ 18.33	2263/ 52.35	20891	4323
Weaving It Together 4	13860/ 82.48	1975/ 51.25	2945/ 17.52	1879/ 48.75	16805	3854
Weaving It Together 3	10582/ 88.54	2133/ 67.67	1369/ 11.46	1019/ 32.33	11951	3152
What a World 2	10050/ 88.65	1371/ 71.67	1287/ 11.35	542/ 28.33	11337	1913
What a World 3	14326/ 86.22	1769/ 60.98	2290/ 13.78	1132/ 39.02	16616	2901
World Class Readings 2	7855/ 87.59	1492/ 65.50	1113/ 12.41	786/ 34.50	8968	2278
World Class Readings 3	12493/ 81.07	1789/ 50.52	2918/ 18.93	1752/ 49.48	15411	3541

Table 6: Text coverage by TBEWL 2000 across GE textbooks

Conversely, learning 2,001 new words over the course of one year leads the researcher to speculate on the possibility of this occurring in the current EFL context, especially where there is only 2-3 hours of English instruction per week. If Nation's (2001) estimate that native speakers read

about 10-12 books per year to acquire 1,000 words is correct, then setting the vocabulary goal 2001 new words for one GE textbook, to be learnt within one year, would be akin to 'building castles in the air'. In view of this, the vocabulary size of a textbook cannot be ignored and should be taken into account from the start, at the onset of syllabus design.

4.3 Percentage of the GEPT intermediate vocabulary covered in a GE textbook

Research Question 3, "To what extent does a GE textbook cover the GEPT intermediate vocabulary?" can be further presented in another way, "How well does a GE textbook prepare for the intermediate GEPT?" Table 7 below gives us some indication of the task facing an EFL learner when preparing the vocabulary necessary for the intermediate GEPT. Listed at the top of the table, of the 3,788 word families in the GEPT intermediate list⁶, 2,463 occurred in the textbook *Reading Matters 3*. This GE textbook, which was claimed by the publisher to be suitable for students at the high-intermediate level (see Table 1), would possibly prepare learners for 65.02% (2,463 divided by 3,788) of the vocabulary they would encounter in an intermediate GEPT. At the bottom of Table 7, *Access Reading 3* for low-intermediate learners covers 24.55% of the vocabulary of the intermediate GEPT (930 divided by 3,788). This implies that when a student is preparing for the intermediate GEPT, studying just one GE textbook at the high-intermediate level or even at the advanced level may not suffice.

As far as the difficulty of a GE textbook – as measured by the coverage of the GEPT intermediate vocabulary – is concerned, *Access Reading 3* is the least difficult among the 36 books (97.31% text coverage, see Table 7, bottom left). When one reaches the vocabulary level of the intermediate GEPT, i.e. 3,788 word families, the textbook *Access Reading 3* would become less challenging to him/her. The vocabulary knowledge of 3,788 word families would enable one to understand 97.31% of the total words in *Access Reading 3*. A student would be able to gain adequate understanding of the texts in *Access Reading 3* in relative ease compared to the other textbooks. On the contrary, *Mosaic Reading 2* is the most difficult, with only 90.59% of the words in its texts known to the learner (equivalent to almost one new word in each single line of text read), even if one is already equipped with the vocabulary of the intermediate GEPT.

Textbook	% Text Coverage (in tokens)	Occurrences of the intermediate GEPT word families	% preparation for the GEPT intermediate vocabulary
Reading Matters 3	91.85%	2,463	65.02%
Reading Matters 4	92.80%	2,277	60.11%
Mosaic Reading 2	90.59%	2,264	59.77%
Tapestry Reading 4	92.38%	2,220	58.61%
Active Skills for reading 4	93.05%	2,119	55.94%
Topics for Today 5	91.82%	1,871	49.39%
College Reading 4	92.35%	1,858	49.05%
Weaving It Together 4	92.85%	1,852	48.89%
Mosaic Reading 1	93.07%	1,847	48.76%
NorthStar-advanced	94.35%	1,819	48.02%
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	92.3%	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%
Reading for Success 2	93.28%	1,454	38.38%
Reading for Real-advanced	94.27%	1,449	38.25%
Interactions 1	96.56%	1,427	37.67%
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%
What a World 2	93.55%	1,083	28.59%
Reading Challenge 2	96.04%	1,015	26.8%
Access Reading 3	97.31%	930	24.55%

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

Table 7: Text coverage and word families covered by the GEPT intermediate vocabulary across GE textbooks

4.4 The vocabulary level of a GE textbook

Vocabulary levels were defined as the number of words counted from the top of BNC HFWL accounting for 95% of the running words in that textbook if we accept the assumption of 95% text coverage as the minimum for successfully guessing meanings from context and gaining reasonable comprehension. As mentioned previously, there are fourteen 1,000-word bands created from the British National Corpus in the RANGE program. Thus, the BNC HFWL was used to identify the division among the diverse vocabulary levels contained within the GE textbooks. The text coverage of each 1,000-word band in the target textbook was calculated by counting the number of 1,000-word bands needed until the total coverage reached 95%. Running *Mosaic Reading 2* and

Select Readings-intermediate on the base words, the BNC HFWL 1st–14th 1,000 through the RANGE program, one can see that the vocabulary level of *Mosaic Reading 2* was 13,000 words with its cumulated text coverage reaching 95.04%, and *Select Readings-intermediate* was 3,000 words with its cumulated text coverage at 95.32%.

Table 8 reveals that the vocabulary distribution of *Mosaic Reading 2* among the different bands of the ranked BNC high-frequency word lists is more diverse than that of *Select Readings-intermediate*. One can see a striking difference in the vocabulary levels of these two books, one for intermediate learners and the other for advanced learners. The 13,000-word level hints that *Mosaic Reading 2* had a much larger vocabulary and using this textbook would result in learners working on a wider variety of vocabulary, some of which they would encounter only once or no more than a few times throughout the book. In contrast, *Select Readings-intermediate* used a smaller vocabulary, converging at the 2,500–3,000 word level along the scale of the BNC HWFL. Because the BNC 1st–14th 1,000 English words are ranked in accordance with their frequency of occurrence, with the 1st 1,000 words being the most frequent and correspondingly the 14th 1,000 words the least frequent, a book with a higher vocabulary level can be interpreted as having more English words appearing in the latter 1000-word bands. Table 9 is a complete list of vocabulary levels across the 36 GE textbooks and their vocabulary distribution among the BNC HFWL 1st–14th 1000, with the cumulated text coverage reaching 95%.

BNC HFWL	Select Readings-intermediate	Mosaic Reading 2
	tokens/ text coverage	tokens/ text coverage
1 st 1000	8,765/ 85.90%	21,070/ 76.58%
2 nd 1000	755/ 7.40%	2,376/ 8.64%
3 rd 1000	206/ 2.02%	803/ 2.92%
	By this level, the cumulated coverage 95.32%	
4 th 1000	162/ 1.59%	66/ 2.40 %
5 th 1000	43/ 0.42%	391/ 1.42%
6 th 1000	64/ 0.63%	217/ 0.79%
7 th 1000	27/ 0.26%	145/ 0.53%
8 th 1000	23/ 0.23%	180/ 0.65%
9 th 1000	15/ 0.15%	94/ 0.34%
10 th 1000	23/ 0.23%	64/ 0.23%
11 th 1000	10/ 0.10%	54/ 0.20%
12 th 1000	4/ 0.04%	41/ 0.15%
13 th 1000	4/ 0.04%	51/ 0.19%
	By this level, the cumulated coverage 95.04%	
14 th 1000	10/ 0.10%	36/ 0.13 %

Table 8: Vocabulary levels of *Select Readings-intermediate* and *Mosaic Reading 2*

There are four apparent exceptions in Table 9 with regard to the book level claimed by the publishers and the vocabulary level measured by the BNC HFWL 1st–14th 1,000. Two low-intermediate GE textbooks, *Reading for Success 2* and *What a World 2*, had a vocabulary level surpassing 7,000 words and a wider dispersion along the scale of BNC high-frequency words. Counter to expectations, these two books at the low-intermediate level actually had more middle- and low-frequency English words. By contrast, two advanced books, *NorthStar-advanced* and *Reading for Real-advanced*, had a denser distribution of words, scattering in the range of the 4th and 5th 1000-word bands. The above two situations show that it is possible to select an advanced-level GE textbook with a lower vocabulary level. The discrepancies in vocabulary levels among GE textbooks imply that textbook authors may not apply the same standard in their selection of words while writing their teaching materials for the publishers. It can also be ascribed to the fact that articles in textbooks for different reading purposes may involve different levels of difficulty.

Some low-frequency words appear more frequently in a certain genre or subject matter and can function as technical words in certain fields. This echoes Nation's (2001) remark that "one person's technical vocabulary is another person's low-frequency word" (p. 20). By the same token, some vocabulary presumed difficult by some authors may be considered easy by others.

GE Textbook	Book Level	Vocabulary Level
Mosaic Reading 2	advanced	13,000
Reading Matters 3	high-intermediate	9,000–9,500
World Class Reading 3	high-intermediate	9,000
Reading for Success 2	low-intermediate	7,000–7,500
Topic for Today 5	high-intermediate	7,000–7,500
What a World 2	low-intermediate	7,000–7,500
Tapestry Reading 4	advanced	7,000
Reading for the Real World 3	high-intermediate	6,000–6,500
Active Skills for Reading 4	high-intermediate	6,000
College Reading 4	advanced	6,000
Mosaic Reading 1	high-intermediate	6,000
Reading for Success 3	intermediate	6,000
Weaving It Together 4	intermediate	6,000
Reading for the Real World 2	intermediate	5,500–6,000
Reading Matters 4	advanced	5,500–6,000
Hot Topics 3	high-intermediate	5,500–6,000
What a Word 3	intermediate	5,000–5,500
Concepts for Today 4	intermediate	4,500–5,000
NorthStar-focus on reading-advanced	advanced	4,500–5,000
NorthStar-high intermediate	high-intermediate	4,500–5,000
Active Skills for reading 3	intermediate	4,000–4,500
Reading Challenge 3	intermediate	4,000–4,500
Reading for Real-advanced	advanced	4,000–4,500
Reading for Real-high-intermediate	high-intermediate	4,000–4,500
Select Readings-high-intermediate	high-intermediate	4,000–4,500
Concepts & Comments 4	high-intermediate	4,000
Interactions 2 for reading	intermediate	4,000
Hot Topics 2	intermediate	3,500–4,000
Weaving It Together 3	intermediate	3,500–4,000
World Class Reading 2	intermediate	3,500–4,000
Interactions 1 for reading	low-intermediate	3,500–4,000
Access Reading 3	low-intermediate	3,000–3,500
Cause & Effect 3	intermediate	3,000–3,500
Reading Challenge 2	low-intermediate	3,000–3,500
In Context-Steps to Academic Reading	intermediate	3,000
Select Readings-intermediate	intermediate	2,500–3,000

The vocabulary level of each GE textbook was obtained by counting the number of words from the top of BNC HFWL until the cumulated lexical coverage reached 95%.

Table 9: Vocabulary levels of the GE Textbooks measured by the BNC HFWL 1st–14th 1000 at the accumulated text coverage reaching 95%

Despite these few exceptions, Table 9 provides a profile of the GE textbooks' vocabulary levels can be summarized as below.

1. In the present textbook market in Taiwan, a low-intermediate GE textbook generally contains vocabulary reaching the level of the BNC 3rd–4th 1,000 high-frequency words, while an intermediate book would normally reach the level of 4,000–4,500 high-frequency words.

2. Assuming an increase of 500–1,000 words for each level of progression, upper-intermediate and low-advanced textbooks should likewise contain a vocabulary breadth of 5,000–5,500 words, and 6,000 words or above.

5 Conclusion

5.1 Findings and pedagogical implications

The present corpus-based study of GE textbooks discussed the role a GE textbook may play in vocabulary learning in a range of ways. The principal concern was four-fold:

Coxhead's (2000) 570 academic word families account for 1.3%–6.54% of the total words in a GE textbook. Approximately 49 to 415 interdisciplinary academic words can be learnt from a GE reading textbook. If academic vocabulary learning is the goal, teachers may need to choose a GE textbook containing more readings for academic purposes or supplement them with different sources in this regard. With more exposure to academic texts, learners will consolidate the vocabulary knowledge acquired from GE courses.

Beyond the 2,000-word level, a GE textbook in the corpus can supply students with 162 to 2,001 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 a student's vocabulary size has reached a certain level. If a learner's vocabulary is small, he/she may not enjoy reading due to the enormous vocabulary load. However, if a learner's 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. Considering students' heterogeneous English proficiency levels resulting from different learning backgrounds, it is recommended that a placement test including a test of the vocabulary level should be carried out for college freshmen upon admission in order to subsequently divide English classes into several levels. Gauging students' vocabulary breadth by means of an entry placement test and placing them in the proper class level would enable a college teacher to select an English textbook for general purposes with an appropriate level of reading difficulty.

A GE textbook may be useful in preparing learners for the intermediate GEPT, with coverage of 24.55% to 65% of the vocabulary involved in the test. If a student intends to gain a high score on the intermediate GEPT, so as to achieve the English graduation benchmarks, he/she would need to make a determined and continuous effort to expand his/her vocabulary during the college years. To compensate for insufficient number of class hours, teachers can give reading assignments through the implementation of an extensive extracurricular reading scheme 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 further deepen their lexical knowledge.

In relation to the topics and articles they contain, GE textbooks offer texts of varying vocabulary levels, ranging from the BNC HFWL top 3,000 most frequently-occurring words to the BNC HFWL 13,000. In the present research, the book levels claimed by the publishers did not seem to coincide with their vocabulary levels in a few instances. This finding may be helpful in raising teachers' awareness of the importance of considering the vocabulary level when choosing an English textbook. They should treat the book level claimed by the publishers with caution. Apart from this awareness, it is also crucial to set vocabulary goals at a reasonable size and level for the course of study.

5.2 Limitations and recommendations

This research was only a preliminary attempt to survey the GE textbooks used in an EFL tertiary setting. Its focus on the vocabulary component by no means implies that vocabulary size,

levels and lexical coverage are exclusive factors in selecting a GE book for class use. Through a lexical corpus study, the researcher hopes to draw the attention of instructors and learners to GE courses. A good command of vocabulary is essential for success at higher levels of education. The current research data may serve as a basis for in-depth research into the acquisition of vocabulary. Other parameters such as syntax and content area knowledge may be worth investigating but are beyond the focus of the current study. For the purpose of data triangulation, further research might achieve further findings through a qualitative analysis of students' perception toward language textbooks in relation to vocabulary load and reading difficulty. It would also be interesting to examine how a textbook can be used and how books within one series may provide a pathway for language development.

Last but not least, vocabulary may not be the only component of an English course, but it is a component that learners notice and that can occupy a lot of their learning time. It is a component that deserves more attention from course designers and the aim of this study has been to generate that attention.

Notes

¹ In Taiwan, a university contains more than twelve academic departments, in contrast with a college, which has less than three. A university normally recruits senior high

² The mappings of GEPT, TOEFL, TOEIC and IELTS scores to the CEFR can be obtained from Wikipedia online at <http://en.wikipedia.org/wiki/CEFR>.

³ Coverage here is defined 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).

⁴ Coverage here refers to the percentage of the running words in a text or corpus that are also covered by a particular word list.

⁵ Based on Table 2, the proportion of word types to word families ranges from 1.54 to 2.18. The denominator 2 was adopted for easy estimation when transforming word types to word families. (E.g. $324 \div 2 = 162$; $4003 \div 2 = 2001$)

⁶ The GEPT intermediate vocabulary contains 4,947 lexical items, including the 2,263 elementary level words. When made into word families, the intermediate GEPT involves 3,788 word families.

Acknowledgements

1. The researcher wishes to thank Professor Paul Nation, whose website containing high-frequency word lists and RANGE program was indispensable to this study.
2. The researcher is grateful to the anonymous reviewers of this paper for their valuable comments.
3. The researcher wishes to express her profound gratitude to the following publishers for their permission to reproduce copyright material for the current research.

Corpus use of content: From *Select Readings Intermediate Student's Book* and *Select Readings Upper-Intermediate Student's Book* copyright © 2004 Reproduced by permission of Oxford University Press; from *What a World 2*, *What a World 3* copyright © 2005, *NorthStar: Focus on Reading and Writing-high intermediate* and *NorthStar: Focus on Reading and Writing-advanced* copyright © 2004 Used by permission of Pearson Education, Inc.; from *Reading for Real High-Intermediate* and *Reading for Real Advanced* copyright © 2001 Reproduced by permission of Lynx Publishing; from *Reading for Success 2* and *Reading for Success 3* copyright © 2004 Reproduced by permission of Taiwan ELT; from *Interactions 1 & 2*, *Mosaic 1 & 2*: Reading copyright © 2007, *World Class Readings 2 & 3: A Reading Skills Text* copyright © 2004 Reproduced by permission of McGraw-Hill; from *Reading for the Real World 2*, *Reading for the Real World 3* copyright © 2004, *Reading Challenge 2*, *Reading Challenge 3* copyright © 2005 Reproduced by permission of Compass.

4. The researcher also wishes to thank the National Science Council (NSC) in Taiwan for its research grant for the time-consuming corpus building and manual proofreading.
5. Without all mentioned, the current research would never have been completed.

References

- Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Carroll, J.B., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. New York: American Heritage.

- 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.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Goulden, R., Nation, I.S.P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363.
- Heatley, A., Nation, I.S.P., & Coxhead, A. (n.d.). RANGE and FREQUENCY programs. Retrieved January 15th, 2008, from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>
- 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.
- Language Training and Testing Center (n.d.). General English Proficiency Test. Retrieved March 3rd, 2008, from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_main.htm
- 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., & Sim, D.D. (1985). Measuring and explaining the threshold needed for English academic purposes texts. *Foreign Language Annals*, 18, 405–413.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Harlow: Pearson Educational.
- 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., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35–41.
- 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.
- Roeber, C., & Pan, Y. (2008). Test review: General English Proficiency Test. *Language Testing*, 25(3), 403–408.
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary, description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Weir, C.J., & Wu, J. (2002). Parallel-form reliability: A case study of the GEPT intermediate spoken performance test. In *The Proceedings of the Fifth Academic Forum on English Language Testing in Asia* (pp. 29–58). Tokyo: The Society for Testing English Proficiency.
- Weir, C.J., & Wu, J. (2006). Establishing test form and individual task comparability: A case study of a semi-direct speaking test. *Language Testing*, 23(2), 167–197.
- West, M. (1926). *Learning to read a foreign language*. London: Longman, Green & Co.
- West, M. (1953). *A general service list of English words*. London: Longman, Green, & Co.
- Wikipedia. (n.d.). Common European framework of reference for languages: use in language testing. Retrieved October 23, 2008, from <http://en.wikipedia.org/wiki/CEFR>
- Wu, J., & Wu, R. (2007). *Using CEFR in Taiwan: The perspective of a local examination board*. Paper presented at the Fourth European Association for Language Testing and Assessment (EALTA) conference, Sitges, Spain.

Appendix A

Thirty-six General English Textbooks:

- Ackert, P., & Lee, L. (2005). *Reading and vocabulary development 3: Cause and effect*. Boston, MA: Thomson Heinle.
- Ackert, P., & Lee, L. (2005). *Reading and vocabulary development 4: Concepts and comments*. Boston, MA: Thomson Heinle.
- Anderson, N.J. (2003). *Active skills for reading 3*. Boston, MA: Thomson Heinle.
- Anderson, N.J. (2003). *Active skills for reading 4*. Boston, MA: Thomson Heinle.
- Benz, C., & Schuemann, C. M. (2006). *College Reading 4*. New York: Houghton Mifflin.
- Bernard, J. & Lee, L. (2004). *Select readings-upper intermediate*. Oxford: Oxford University Press.
- Broukal, M. (2005). *What a world 2: Amazing stories from around the globe*. New York: Pearson Education.
- Broukal, M. (2005). *What a world 3: Amazing stories from around the globe*. New York: Pearson Education.
- Broukal, M. (2004). *Weaving it together 3: Connecting reading and writing*. Boston, MA: Thomson Heinle.
- Broukal, M. (2004). *Weaving it together 4: Connecting reading and writing*. Boston, MA: Thomson Heinle.
- Collins, T. (2005). *Access reading 3: Reading in the real world*. Boston, MA: Thomson Heinle.
- Constantino, M., & Dabbs, T. (2001). *Reading for real-high intermediate*. Vancouver, CA: Lynx.
- Constantino, M., & Dabbs, T. (2001). *Reading for real-advanced*. Vancouver, CA: Lynx.
- English, A.K., & English, L.M. (2004). *NorthStar: Focus on reading and writing-high intermediate*. New York: Addison Wesley Longman.
- Graber, B., & Babcock, P. (2004). *Reading for the real world 3*. Sachse, TX: Compass.
- Hartmann, P., & Kirn, E. (2007). *Interactions 2: Reading*. New York: McGraw-Hill.
- Kirn, E., & Hartmann, P. (2007). *Interactions 1: Reading*. New York: McGraw-Hill.
- Lee, L., & Gundersen, E. (2004). *Select readings-intermediate*. Oxford: Oxford University Press.
- Malarcher, C., & Janzen, A. (2005). *Reading challenge 2*. Sachse, TX: Compass.
- Malarcher, C., & Janzen, A. (2005). *Reading challenge 3*. Sachse, TX: Compass.
- Miller, J.L., & Cohen, R.F. (2004). *NorthStar: Focus on reading and writing-advanced*. New York: Addison Wesley Longman.
- Pavlik, C. (2006). *Hot topics 2*. Boston, MA: Thomson Heinle.
- Pavlik, C. (2006). *Hot topics 3*. Boston, MA: Thomson Heinle.
- Rashap, A., Thomas, V. & Edwards, W. (2004). *Reading for success 2*. Baltimore, MD: American ELT.
- Rashap, A., Tong, V. & Edwards, W. (2004). *Reading for success 3*. Baltimore, MD: American ELT.
- Rogers, B. (2004). *World class readings 2: A reading skills text*. New York: McGraw-Hill.
- Rogers, B. (2004). *World class readings 3: A reading skills text*. New York: McGraw-Hill.
- Smith, L.C., & Mare, N.N. (2004). *Concepts for today 4*. Boston, MA: Thomson Heinle.
- Smith, L.C., & Mare, N.N. (2004). *Topics for today 5*. Boston, MA: Thomson Heinle.
- Sokolik, M.E. (2000). *Tapestry: Reading 4*. Boston: Heinle & Heinle.
- Wegmann, B., & Knezevic M. (2007). *Mosaic 1: Reading*. New York: McGraw-Hill.
- Wegmann, B., & Knezevic M. (2007). *Mosaic 2: Reading*. New York: McGraw-Hill.
- Wholey, M.L., & Henein, N. (2007). *Reading Matters 3: An integrated approach to reading*. New York: Houghton Mifflin.
- Wholey, M.L., & Henein, N. (2002). *Reading Matters 4: An integrated approach to reading*. New York: Houghton Mifflin.
- Zukowski/Faust, J., Johnston, S.S., & Templin, E.E. (2002). *Steps to academic reading 4: In context*. Boston, MA: Thomson Heinle.
- Zwier, L.J., & Stafford-Yilmaz, L. (2004). *Reading for the real world 2*. Sachse, TX: Compass.