



To What Extent May EFL Undergraduates with EMI Develop English Vocabulary? The Case of Business Administration

Wenhua Hsu *

I-Shou University, Taiwan

Abstract

International colleges using the English-medium instruction (EMI) mode are gaining popularity among higher education institutions in Taiwan, where English is a foreign language (EFL). The researcher targeted English-medium university textbooks as a research focus, since they are first and foremost learning material of specialist knowledge and offer EFL non-English-majoring students a channel for exposure to English. A 7.2-million-token textbook corpus of business administration (BA) courses was compiled and the vocabulary level along the word-frequency scale of the British National Corpus and the Corpus of Contemporary American English was measured. Then the researcher sought to estimate how many new words EFL BA majors can encounter often enough for learning to occur. Results show that BA textbooks reached the 6th 1000-word-family level at 98% text coverage. Beyond the first 3000 word families, only 3,471 mid-frequency (the 4th to 9th 1000 level) word families and 547 low-frequency (from the 10th 1000 level onwards) word families occurred 12+ times. The occurrence of 12 times was assumed as a benchmark for incidental learning to occur. For international college practitioners who are concerned with their EFL students' vocabulary development, the results can serve as a reference for future investigations into other disciplines.

Keywords: lexical coverage, vocabulary levels, English-medium instruction, repetition

INTRODUCTION

English is a foreign language in Taiwan (an EFL context), and the official language is Mandarin. However, since 2017, Taiwan's Ministry of Education has started encouraging domestic universities to recruit international students to alleviate the impact of a plummeting birth rate on higher education (MOE, 2017). Also, in response to globalization, more and more universities in Taiwan have set up international colleges and offered English-medium instruction (EMI) curricula to raise their international visibility in world academic rankings. Therefore, EMI degree programs provided by international colleges and non-EMI degree programs by other colleges run parallel in many universities. International colleges enroll both international and domestic

students. International students mainly come from Indonesia, Vietnam, Thailand, the Philippines and Japan to take degree courses alongside Taiwanese students in English-taught programs (MOE, 2023).

Despite nearly twice the tuition fees for domestic students to study in international colleges, the popularity of EMI degree programs reflects the widespread belief that EMI provides immersion opportunities, which facilitate incidental learning of English (Yeh, 2014). It is generally held that students can simultaneously acquire disciplinary knowledge and enhance their English abilities by studying in EMI contexts. Past studies have provided some evidence that continuous exposure to English contributes to lexical gains over time (Brown, Waring & Donkaewbua, 2008; Pellicer-Sánchez & Schmitt, 2010; Vidal, 2011).

However, not all EMI programs provide the same English immersion as those in majority-native-English-speaking countries. In Taiwan, although universities with international colleges recruit international academics with a PhD in a specialized field, they are not necessarily native speakers of English. A significant portion of the faculty members in international colleges are Taiwanese teachers, even though they have earned a PhD from one of the core Anglosphere and can speak English fluently. Meanwhile, outside international colleges, Taiwanese students are usually exposed to their mother tongue. Such a context may contradict a typical immersion mode, in which teachers and local students are primarily native speakers of English. Given these discrepancies, immersion in the EMI degree programs provided by international colleges in Taiwan's universities should not be interpreted in the same light as immersion in English as a second language (ESL) contexts.

Moreover, in a survey of 70 European universities offering EMI programs, O'Dowd (2018) found that the concern of English improvement usually plays second fiddle compared with disciplinary knowledge development. Jiang, Zhang and May (2019) also unearthed the phenomenon that under the constraint of class time, subject teachers generally regard adequate delivery of specialist content as their responsibility and do not consider English teaching within their remit (Airey, 2012; Block & Moncada-Comas, 2019; Dearden & Macaro, 2016).

In non-EMI subject classrooms, specialist textbooks are, first and foremost, learning material for core knowledge. In EMI subject classrooms, as a source of input, English-medium textbooks offer students a channel for exposure to English. When students immerse themselves in textbooks, unknown English words may impede their comprehension of the subject content. However, if a new word repeatedly occurs, it can become a known word, since repetitions increase the salience of a word and enhance the retention of knowledge of that word over time. As such, the researcher targeted English-medium specialist textbooks as a research focus with a particular concern about potential vocabulary growth because vocabulary plays a fundamental role in developing language proficiency.

Amid a range of undergraduate programs offered by Taiwan's universities for international students and domestic students who want to receive EMI, the researcher set out from business administration (hereafter short for BA), since it is one of the most common majors offered by many universities and colleges. This study sought to answer the following two research questions (RQ):

RQ1. What vocabulary level may EFL BA undergraduates attain after finishing their degree programs in international colleges? Namely, what are the vocabulary levels of English-medium textbooks of BA core courses?

RQ2. Beyond the most frequent 3000-word families, how many words from English-medium specialist textbooks may EFL BA undergraduates encounter often enough for learning to occur?

LITERATURE REVIEW

In early studies, Goulden, Nation, and Read (1990) estimated that an educated native speaker of English has approximately a vocabulary of 20,000-word families. Despite years of study, the vocabulary goal of learning 20,000 words is still far beyond the reach of EFL learners like ours. Fortunately, Nation (2006) reported that knowledge of the first 9000-word families would suffice to provide 98% text coverage (i.e., two unknown words per hundred words) and allow the average person to read a wide range of authentic texts without too many unknown words being a hurdle to comprehension. Compared with 20,000-word families, the vocabulary goal of learning the first 9000-word families is more feasible.

To estimate the vocabulary level and calculate the number of words contained in a text, Nation (2020) has endeavored to compile large-scale word-family lists for many years. Thus far, he has compiled 28,000+ word families and ranked the first 25,000 into twenty-five 1000-word-family lists based on their dispersion and frequency in the 100-million-word British National Corpus (from the 1980s to the early 1990s) and one-billion-plus-word Corpus of Contemporary American English (from 1990 onwards). The combined corpus is hereafter referred to as BNC/COCA. The rationale behind the frequency ranking is that frequently-occurring words are more likely to be learned than infrequently-occurring words (Nation, 2006).

Along the BNC/COCA word-frequency scale (1st—25th 1000), Schmitt and Schmitt (2014) divided the 25 vocabulary levels into three bands—high-frequency vocabulary (the 1st—3rd 1000 word families), mid-frequency (4th—9th 1000) and low-frequency vocabulary (beyond the 9th 1000). They further stressed the importance of mid-frequency vocabulary for proficient use based on Nation's (2006) estimate that knowledge of the first 9000-word families is essential to fluent reading of all sorts of texts. Concerning the validity of the BNC/COCA twenty-five 1000-word-family lists, Dang and Webb (2016) as well as Dang, Webb and Coxhead (2020) conducted a series of tests on different word lists and concluded that the BNC/COCA word-family lists

performed better on a variety of texts than the other wordlists. In this study, the researcher also used the ranked BNC/COCA word-family lists to measure the vocabulary levels and amounts in English-medium textbooks of BA courses.

Vocabulary levels and vocabulary thresholds needed for adequate comprehension are two sides of the same coin. Measuring the vocabulary level of a text can be approached from lexical coverage of text, which Nation (2006) defined as “the percentage of running words in the text known by the reader” (p. 61). For example, when learners know 98% of the total words of a text (98% text coverage), they would encounter two unknown words per hundred words being read. Although lexical coverage of text does not equate to comprehension degrees, Schmitt, Jiang and Grabe (2011) detected that the two have a positive linear relationship, indicating that when the percentage of known words in a text increases, the likelihood of better comprehension also increases. For adequate comprehension, Laufer and Ravenhorst-Kalovski (2010) and Nation (2006) proposed 98% coverage as a vocabulary threshold. In this study, the researcher also adopted the putative 98% text coverage as a cutoff for measuring the vocabulary levels of English-medium textbooks of BA core courses.

As mentioned, the current context (i.e., international colleges in EFL universities) is in an EMI mode. Pertinent to immersion in EMI is incidental learning. Among many factors that are critical to incidental learning is repetition, since a single encounter with a new word does not offer learners sufficient experiences in a variety of contexts to support the robust learning of that word (Horst, 2013; Horst, Parsons & Bryan, 2011). However, lexical researchers have diverged on the minimal repetitions that are required for incidental learning. Based on previous studies on the number of encounters needed for word learning, ranging from 5 to 16 repetitions, Nation (2014) reasoned that for uptake to happen, a word’s occurrence in a variety of contexts totaling 12 times would just be enough to develop knowledge of that word. Following Nation (2014), the researcher adopted 12 occurrences as a cutoff to address RQ2 concerning how many words beyond the first 3000-word families EFL BA undergraduates may encounter often enough for learning to occur.

For incidental learning, the amount of input must be large enough to meet new words enough times. Nation (2014) chose classic novels and Hsu (2019, 2020) selected VOA news articles and TED Talks English transcripts for extensive reading. Nation (2014) computed that learners would need to read a minimum of 3 million words of novels to encounter most of the first 9000-word families 12 times or more (hereafter 12+). Focusing upon mid-frequency vocabulary learning (specifically, encountering 800+ word families from each of the 4th to 9th 1000 levels 12+ times), Hsu (2019, 2020) estimated that learners would need to read 6 million words of VOA news or 4.8 million words of TED Talk English transcripts at the minimum.

Similarly, when EFL BA undergraduates finish their EMI degree programs in international colleges, they will have read many English-medium specialist textbooks, comparable to voluminous reading. In view of this, the researcher was concerned about

the English vocabulary development of EFL BA majors during four years of college study.

RESEARCH METHOD

Compiling a textbook corpus of business administration (BA)

Referring to the BA undergraduate programs in international colleges of prestigious universities in Taiwan, the researcher identified twelve core courses that most BA departments require their students to take regardless of their selected area of specialization in BE, such as marketing, human resources, finance and investment. The BA core courses include three foundation courses (principles of accounting, fundamentals of management and micro- & macro-economics) within the College of Business and Management and four fundamental professional courses (marketing, human resources, production & operation, international business) as well as five specialized BA courses (finance & investment, money & banking, risk management & insurance, management information system, business policy & strategy).

The researcher recorded a list of BA textbook titles based on the catalogs of internationally reputable textbook publishers for BA majors and the reference lists that BA professors provide for their students. The same textbooks were noted down on the priority list for later screening. After this textbook listing, the selection of textbooks with the 5th or higher edition was made to ensure their popularity. Moreover, considering sufficient content for use for one semester at least, the textbook for inclusion in the candidate list must have 600+ pages. This arbitrary decision was made after a series of textbook comparisons regarding the number of editions and pages.

When the textbook listing was finalized, three BA teachers offered their help to confirm reputable textbooks for each compulsory course by ticking the appropriate box on a survey list. Next to each textbook were two questions (concerning whether they recognize this textbook and whether they would consider it for classroom use) with three options (yes, no and uncertain) for a check. The textbooks without a consensus of two teachers ticking yes for either of the two questions were removed. After a sequence of selections, the number of candidate textbooks across twelve core courses was reduced to 85.

From the shortlist, the researcher randomly picked four among six to seven candidate textbooks for each core course, totaling 48 textbooks (see Table 1). It was assumed that BA undergraduates in international colleges under the EMI mode would read these textbooks at some stage within four years of university study. All the sampled textbooks in PDF format were downloaded from digital databases (Oxford Scholarship ebook Online, My iLibrary ebook and Netlibrary), which were subscribed to by the university libraries in Taiwan for educational and research purposes. They were saved as plain texts in UTF-8, with photos, graphs and tables automatically removed. After deleting the

front matter, back matter and references, the BA textbook corpus had approximately 7.2 million tokens (7203,241 running words).

Table 1. Composition of the BA textbook corpus

| Core courses | Tokens | Core courses | Tokens |
|------------------------|---------|-----------------------------|---------|
| Accounting & auditing | 600,161 | Micro- & macro-economics | 604,624 |
| Finance & investment | 600,985 | Management | 613,514 |
| Human resources | 596,985 | Marketing | 586,384 |
| Money & banking | 597,643 | Management info system | 610,771 |
| International business | 614,063 | Risk management & insurance | 609,048 |
| Production & operation | 598,491 | Business policy & strategy | 570,572 |
| Total: 7203,241 tokens | | | |

Instrument and processing

The researcher first measured the vocabulary levels of BA textbooks along the BNC/COCA word-frequency scale and then calculated the number of mid-frequency word families (the 4th–9th 1000) with 12+ occurrences. The vocabulary threshold of a text can be gauged by counting the number of the ranked BNC/COCA 1000-word-family lists needed from the first 1000 until the text coverage accumulates to 98%. Meanwhile, the vocabulary level of that text can be extrapolated based on which 1000-word-family level is the last one added when the cumulative coverage reaches 98%.

The AntWordProfiler Version 2.1.0 (Anthony, 2023) was implemented to analyze the lexical profiling of BA textbooks. In addition to the ranked 25 BNC/COCA 1000-word-family lists, another three ever-growing lists of proper nouns, abbreviations and transparent compounds complied by Nation (2017) were also installed in the program. The words that were placed at the 'Words NOT Found In Base Lists' (hereafter called off-list) by the AntWordProfiler were further examined.

If off-list words were personal or geographical names, they were supplemented to the existing proper noun list. For hyphenated transparent compounds in the off-list (e.g., paid-in, spin-off, work-in-process), TextMate 2.0 was applied to replace these words with a space. The hyphens were summarily removed so the AntWordProfiler would not mistake these compounds for off-list words. Similarly, closed compounds in the off-list (e.g., breakeven, baseline, cashflow) were added to the existing compound list to avoid double counting if their component words are already in the BNC/COCA word lists.

Proper nouns, transparent compounds and abbreviations do not make reading arduous. Excluding their text coverage would inflate the vocabulary level and overestimate the vocabulary threshold required for adequate comprehension (Nation, 2006; Nurmukhamedov & Webb, 2019) since the reading load they put on is manageable.

Recognizing a proper noun may not be difficult due to transliteration or the capitalization of its first letter. One can effortlessly infer the meaning of a transparent

compound from its constituent words if the learner is already familiar with its individual words. Inevitably, acronyms and abbreviations occur in the BA register, but they do not pose a hurdle to reading. They are usually explained in context, with the complete form in parentheses or glossed in the appendix. The acronyms and abbreviations that appeared in the off-list were added to the abbreviation list.

Following previous research taking the text coverage of these three types of words into account (Nation, 2006; Rodgers & Webb, 2011; Webb & Rodgers, 2009), the researcher also counted them in to avoid overestimation until 98% coverage was achieved.

Word family as a counting unit for lexical coverage and repetition

For reading purposes, English word recognition is essential for comprehension. Word lemmas as a counting unit may be less preferable to word families in terms of an overestimate of the recognition vocabulary amount needed for good comprehension, according to Webb (2021). A word lemma contains a base form and inflectional forms belonging to the same part of speech, while a word family also includes derivational affixes (Nation & Meara, 2010). For instance, fuel and fuels are nouns, and fuel, fuels, fueled and fueling in verb forms are different lemmas. They would be categorized into two lemmas or counted as one word family, leading to different vocabulary sizes. To avoid overestimation, 'word family' was used as a counting unit for lexical coverage.

Likewise, word families were used for repetition count (e.g., the headword depreciate and its family members depreciated, depreciates, depreciating, depreciation, depreciations belonging to the same word family). In the current context, college-matriculated students already have some knowledge of English word-building rules, which aids in the learning of a new word (Nagy et al., 1989). When a new word and its family members appear in various contexts, learners with inflectional and derivational knowledge would recognize them, guess the meaning from context and strengthen knowledge from multiple encounters, which contribute to learning that word family. For this reason, the occurrences of a word family's stem as well as its inflected and derived forms were summed up so that the combined frequency would show the number of exposure to that word family.

RESULTS AND DISCUSSION

Vocabulary levels of BA core textbooks

Since English-medium specialist textbooks are the major source of input for EFL students in the current context, RQ 1 'What vocabulary level may EFL BA undergraduates attain after finishing their degree programs in international colleges?' can be addressed from the vocabulary levels of English-medium textbooks of BA core courses. Table 2 provides a snapshot of (1) the overall vocabulary levels of BA textbooks at 98% text coverage and (2) the vocabulary distribution along the BNC/COCA word-frequency scale.

Table 2. Vocabulary distribution among the BNC/COCA word lists for the BA textbooks

| BNC Word Lists | Tokens | % coverage in tokens | Cumulative % coverage including proper nouns etc. | Word Families |
|--|-----------|----------------------|---|---------------|
| Proper nouns etc. | 169,276 | 2.35% | 2.35% | |
| 1 st 1,000 | 5,525,606 | 76.71% | 79.06% | 1,000 |
| 2 nd 1,000 | 757,781 | 10.52% | 89.58% | 995 |
| 3 rd 1,000 | 266,520 | 3.70% | 93.28% | 937 |
| 4 th 1,000 | 186,564 | 2.59% | 95.87% | 891 |
| 5 th , 1,000 | 103,006 | 1.43% | 97.30% | 826 |
| 6th 1,000 | 52,584 | 0.73% | 98.03% | 804 |
| 7 th 1,000 | 36,016 | 0.50% | 98.53% | 691 |
| 8 th 1,000 | 25,932 | 0.36% | 99.89% | 549 |
| 9 th 1,000 | 17,288 | 0.24% | 99.13% | 447 |
| 10 th –25 th 1,000 | 62,668 | 0.87% | 100% | 1095 |
| Total | 7,203,241 | 100% | | 8,235 |

Note: Proper nouns etc. include proper nouns, abbreviations and transparent compounds. The bolded figures indicate the level at which the cumulative text coverage has already reached 98% (at the 6th 1000).

The BA Textbook Corpus contained 7,203,241 tokens (running words) and involved 8,235 word families. The first BNC/COCA 1,000 word families accounted for 76.71% of the total tokens, the second 1,000 word families 10.52%, the third 1,000 word families 3.7% and so on. Taking the coverage of proper nouns, abbreviations and transparent compounds into account, by the sixth 1,000-word-family level, 98% text coverage was attained. It was thereby extrapolated that knowledge of the most frequent 6000 word families plus proper nouns, abbreviations and transparent compounds would provide 98% text coverage.

As shown, the cumulative coverage of the first 2000 word families was greater than that of the remaining 1000-word-family lists by a large margin. The text coverage of the 4th 1000 reduced to below 3% and fell to less than 1% at the 6th 1000. From that level onwards, each additional 1000 word families provided a very small increase in text coverage. This reveals that the BA textbooks involving specialist knowledge of different content areas used a condensed vocabulary, converging at the first 6000 word families. This also suggests that if EFL BA undergraduates have a good command of the most frequent 6000 word families, they would be able to read English-medium BA textbooks rather smoothly in terms of interruptions (e.g., consulting a dictionary or guessing unknown words). Moreover, unlike unassisted extensive reading, textbooks are usually read along with the teacher's instruction, further lowering the vocabulary load.

Table 2 answers the first research question, showing that knowledge of the most frequent 6000 word families was generally the vocabulary threshold required for adequate comprehension of a BA textbook at 98% text coverage.

Table 3. Vocabulary levels of BA textbooks across twelve subject areas at 98% text coverage

| Textbooks of BA course | Vocab. Level | Textbooks of BA course | Vocab. Level |
|------------------------|--------------|-------------------------------|--------------|
| accounting & auditing | 6000 | micro- & macro-economics | 6000 |
| finance & investment | 6000 | management | 6000 |
| human resources | 6000 | marketing | 6000 |
| money & banking | 6000 | management information system | 9,000 |
| international business | 6000 | risk management & insurance | 6000 |
| production & operation | 5000 | business policy & strategy | 7,000 |

Note: The vocabulary level of the textbooks for each BA core course was obtained by counting the number of words from the BNC/COCA first 1000 word families plus proper nouns and so on until the accumulated text coverage reached 98%.

Table 3 shows a slight variation in vocabulary levels or vocabulary demands for textbooks across 12 BA core courses, except for production & operation, management information system (MIS) and business policy & strategy. BA textbooks converged at the 6000 word-family level at 98% coverage, with two subjects extending to the 7000 and 9000 word families (see Table 3). Production & operation textbooks needed the least vocabulary (the most frequent 5000 word families) to reach 98% text coverage, while MIS textbooks required the most (9000 word families) to reach 98% coverage. This figure may be overestimated. The MIS textbooks contained many computer-related lexis, ranked in the latter bands of the 25 ranked BNC/COCA 1,000-word family lists. For example, reboot listing in the fourteenth 1000-word family list and malware in the nineteenth 1000 list may be familiar to E-generation learners with limited vocabulary since these words have become household terms. As a result, the vocabulary demand for MIS may be smaller than the first 9000 word families when considering the vocabulary threshold.

In view of the convergent vocabulary level for BA textbooks (within the range of the most frequent 6000 word families), it was conjectured that business technical and sub-technical vocabulary may be intensely scattered from the 1st to 6th 1000 word families (e.g., deficit, mortgage, asset in the 3rd 1000 and dividend in the 4th 1000). These words with a business flavor in connection with money appear to be common language components that slip invisibly in and out of everyday conversation and business domain talk. This may partly explain why some business terms and jargon are common in ordinary English and the more generally accessible nature of business English in contrast with medical, legal, and science and technology English.

Potential vocabulary growth for EFL BA undergraduates receiving EMI

The 7.2-million-token BA Textbook Corpus across 12 core courses contained 8,235 word families from the BNC/COCA 1st to 25th 1000, including 2,932 word families from the 1st to 3rd 1000 (refer back to Table 2). The first 3000 word families were not

considered, since they are often regarded as the base vocabulary that senior high school graduates should have mastered before the nationwide college entrance exam with the English subject included. Therefore, they were assumed to be known. As reviewed in the literature, learning a new word seldom occurs after only one or two encounters (Horst, Cobb & Meara, 1998). If 12 occurrences as an indicator of appearing often enough for learning to happen are sensible in accordance with Nation (2014), Table 4 at the cutoff of 12 repetitions provides an overall picture of the amount of potential vocabulary growth for EFL BA undergraduates in international colleges as a result of constant exposure to English-medium specialist textbooks. Namely, Table 4 answers RQ 2 'Beyond the most frequent 3000 word families, how many words from English-medium specialist textbooks may EFL BA undergraduates encounter often enough for learning to occur?', demonstrating that 4,208 (70%) out of the 6,000 mid-frequency word families (4th—9th 1000) and 1,095 out of the low-frequency word families (10th—25th 1000) would be encountered when BA undergraduates finish their core courses. The 4,208 mid-frequency word families and 1,095 low-frequency word families included words occurring fewer than 12 times.

The figures without parentheses in the right column of Table 4 are bona fide the numbers of word families occurring 12 times or more. A total of 4,018 word families beyond the first 3000 word families appeared 12 times or more.

Table 4. Number of words beyond the first 3000 word families occurring 12+ times in BA textbooks

| BNC/COCA word-frequency scale | Number of words occurring 12+ times in BA textbooks |
|--|---|
| 4 th 1000 | 822/ (891) |
| 5 th 1000 | 800/ (826) |
| 6 th 1000 | 788/ (804) |
| 7 th 1000 | 431/ (691) |
| 8 th 1000 | 333/ (549) |
| 9 th 1000 | 297/ (447) |
| Mid-frequency vocabulary (4 th —9 th 1000) | 3,471/ (4,208) |
| subtotal | |
| Low-frequency vocabulary (10 th —25 th 1000) | 547/ (1,095) |
| Total | 4,018/ (5303) |

Note: The figures in parentheses indicate the number of word families at a particular 1000 word-family level appearing in BA textbooks, including the number of words occurring fewer than 12 times (see Table 2 for the right column).

As mentioned earlier, in support of extensive reading, Nation (2014) used novels as input to estimate the minimal amount to read to gain 12+ encounters with 800+ word families from each of the 4th to 9th 1000 levels. Continually reading up to 3 million words of novels would help learners to meet most of the first 9,000 word families often enough for incidental learning. Hsu (2019, 2020) reported her findings that although

VOA news only reaches the 6th 1000-word-family level at 98% coverage and TED Talk English transcripts attain the 5th to 6th 1000 levels, by voluminously reading VOA news up to 6 million words or reading TED Talk English transcripts up to 4.8 million words, EFL learners would get sufficient input to encounter most of the first 9000 word families enough times.

Compared with Nation (2014) and Hsu (2019, 2020), even though EFL BA undergraduates finish reading circa 7.2 million words of English-medium specialist textbooks, they would still not meet most (>800) of the mid-frequency vocabulary at least 12 times (see Table 4 for < 800 word families from the 6th 1000 onwards). With the advance towards the 9th 1000-word-family level, the number of mid-frequency vocabulary occurring 12+ times becomes smaller and smaller, assuming that 12 repetitions are the necessary threshold for incidental learning to occur.

This may be because, in most novels, news articles and public talks, a large number of different words are used. It may also be the diversity of topic areas involved in novels, news, and talks that results in a rich vocabulary. In contrast, with limited vocabulary use, the vocabulary recycling of BA textbooks is strong, which may reduce students' vocabulary load when they go on to another compulsory textbook. However, BA majors may make little progress in mid-frequency vocabulary learning after the 6th 1000 level.

Despite the potential learning of 3,471 mid-frequency word families from BA textbooks (see Table 4), it is still far below the target of learning most of the 4th to 9th 1000 word families, namely at least 4,800 out of 6000 mid-frequency word families as per Nation (2014) and Hsu (2019, 2020). If knowledge of the first 9000 word families (providing 98% coverage of a variety of texts) is the vocabulary goal, the lexical provision by BA textbooks is apparently not enough. In line with Tables 2 and 3, Table 4 reconfirms that BA textbooks used vocabulary mainly within the first 6000 word families.

CONCLUSION AND IMPLICATIONS

This lexical research was a preliminary study on an international BA undergraduate degree program in an EFL setting. It had a dual purpose: to measure (1) the vocabulary levels of English-medium BA textbooks and (2) the potential vocabulary growth beyond the first 3000 word families from reading BA textbooks. Generally, BA textbooks involving different specialist knowledge reached the 6th 1000-word-family level at 98% text coverage. Even though EFL BA majors complete their core courses, continual reading of English-medium BA textbooks will still not help them to encounter most of the 6000 mid-frequency word families often enough for learning to occur. Data shows that in the 7.2-million-word BA textbook corpus, only 3,471 mid-frequency word families occurred 12+ times. Namely, an academic program delivered in full English does not necessarily warrant the highest inclusion of mid-frequency words.

In view of convergent vocabulary, it is highly likely that EFL BA majors' English vocabulary size would level off at the first 6000 word families if they do not read

English texts outside of their specialist domain, which may often be the case in EFL settings. The value of this study has been to raise this awareness. For vocabulary growth, one advice to EFL undergraduates in international colleges may be to extensively read English newspapers, novels and all sorts of English articles.

In the current EMI subject classrooms, EFL students may pay little attention to the rule-governed aspect of language form and syntax, the repetition of vocabulary in a longer text and sustained exposure to English-medium specialist textbooks will help in strengthening the retention of the vocabulary. For this reason, EMI instruction is still highly advocated.

Although this research contributes to the literature of EMI research, it has been worked narrowly within the field of business administration. The findings may serve as a basis of comparison for investigations into other academic disciplines in the EMI mode. It is hoped that this research may provide some inspiration for future qualitative analyses of EFL learners' lexical needs and perceptions of English-medium specialist textbooks regarding reading difficulty.

REFERENCES

Airey, J. (2012). "I don't teach language." The linguistic attitudes of physics lecturers in Sweden. *AILA Review*, 25(1), 64–79. <https://doi.org/10.1075/aila.25.05air>

Anthony, L. (2023). *AntWordProfiler* (Version 2.1.0) [Computer software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software>

Block, D., & Moncada-Comas, B. (2019). English-medium instruction in higher education and the ELT gaze: STEM lecturers' self-positioning as NOT English language teachers. *International Journal of Bilingual Education and Bilingualism*. Advance online publication. <https://doi.org/10.1080/13670050.2019.1689917>

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading while listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136–163. <https://doi.org/10125/66816>

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

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, 26(4), 617–641. <https://doi.org/10.1177/1362168820911189>

Dearden, J., & Macaro, E. (2016). Higher education teachers' attitudes towards English medium instruction: A three-country comparison. *Studies in Second Language Learning and Teaching*, 6(3), 455–486. <https://doi.org/10.14746/slbt.2016.6.3.5>

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363. <https://doi.org/10.1093/applin/11.4.341>

Horst, J. (2013). Context and repetition in word learning. *Frontiers in Psychology*, 4(149), 1–11. <https://doi.org/10.3389/fpsyg.2013.00149>

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-23. <https://doi.org/10.125/66953>

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, 2, 1-11. <https://doi.org/10.3389/fpsyg.2011.00017>

Hsu, W. (2019). Voice of America (VOA) news as voluminous reading material for mid-frequency vocabulary learning. *RELC Journal*, 50(3), 408-421. <https://doi.org/10.1177/0033688218764460>

Hsu, W. (2020). Can TED Talk transcripts serve as extensive reading material for mid-frequency vocabulary learning? *TEFLIN Journal*, 31(2), 181-203. <https://doi.org/10.15639/teflinjournal.v31i2/181-203>

Jiang, L., Zhang, L.J. & May, S. (2019). Implementing English medium instruction (EMI) in China: Teachers' practices and perceptions, and students' learning motivation and needs. *International Journal of Bilingual Education and Bilingualism*, 22(2), 107-119. <https://doi.org/10.1080/13670050.2016.1231166>

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. <http://hdl.handle.net/10125/66648>

Ministry of Education. (2017). Directions governing the MOE subsidy for colleges and universities to promote and refine English-taught degree programs. [Data file]. Retrieved from <http://edu.law.moe.gov.tw/LawContent.aspx?id=GL000575#lawmenu>

Ministry of Education. (2023). Foreign student number in 2020. Retrieved from <https://stats.moe.gov.tw/>

Nagy, W., Anderson, R., Schommer, M., Scott, J. A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24(3), 262-282. <https://doi.org/10.2307/747770>

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>

Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), 1-16. <http://hdl.handle.net/10125/66881>

Nation, I. S. P. (2017). *The BNC/COCA Level 6 word family lists* (Version 1.0.0) [Data set]. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

Nation, I.S.P. (2020). The BNC/COCA word family lists. Retrieved from https://www.wgtn.ac.nz/_data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf

Nation, I. S. P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd Edition) (pp. 34-52). Edward Arnold.

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, 52(2), 188-200. <https://doi.org/10.1017/S0261444819000028>

O'Dowd, R. (2018). The training and accreditation of teachers for English medium instruction: An overview of practice in European Universities. *International Journal of Bilingual Education and Bilingualism*, 21(5), 553–563. <https://doi.org/10.1080/13670050.2018.1491945>

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22(1), 31–55. <https://doi.org/10125/66652>

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689–717. <https://doi.org/10.5054/tq.2011.268062>

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

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258. <http://dx.doi.org/10.1111/j.1467-9922.2010.00593.x>

Webb, S. (2021). Word families and lemmas, not a real dilemma: Investigating lexical units. *Studies in Second Language Acquisition*, 43(5), 973–984. <https://doi.org/10.1017/S0272263121000760>

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>

Yeh, C.-C. (2014). Taiwanese students' experiences and attitudes towards English-medium courses in tertiary education. *RELC Journal*, 45(3), 305–319. <https://doi.org/10.1177/0033688214555358>