Journal of English Teaching
e-ISSN: 2622-4224 | p-ISSN: 2087-9628
http://ejournal.uki.ac.id/index.php/jet

Potential Vocabulary Growth in EMI Programs: The Cases of Accounting and Civil Engineering in Taiwanese EFL Context

Wenhua Hsu
whh@isu.edu.tw
I-Shou University, Taiwan

Abstract
English-medium instruction (EMI) has become a nationwide trend in Taiwan's higher education institutions. Behind this rapid growth is the widespread belief that EMI provides English immersion, which facilitates incidental learning of the target language. However, not all EMI programs in EFL contexts provide the same immersion as those in the Anglosphere. English-medium university textbooks were therefore targeted as a research focus in that they offer non-English subject majors a sustainable channel for exposure to English in the EFL context. The researcher compiled two 4-million-token accounting and civil engineering textbook corpora and measured the vocabulary levels and amounts along the BNC/COCA word-frequency scale. Results show that accounting textbooks reached the 5th 1000-word-family level at 98% text coverage while civil engineering textbooks stretched to the 9th–10th 1000 level. Twelve repetitions were used as a benchmark for incidental learning. Only 1,479 word families beyond the first 3000 occurred 12+ times in the accounting corpus versus 3,397 word families in the civil engineering corpus. For EMI practitioners who are concerned with their students' vocabulary development, the results may serve as a reference for future investigations into other disciplines.

Keywords:
EMI, BNC/COCA, mid-frequency vocabulary, lexical text coverage

How to cite:
Hsu, W. (2021). Potential Vocabulary Growth in EMI Programs: The Cases of Accounting and Civil Engineering in Taiwanese EFL Context. *Journal of English teaching, 7*(3), 300-314. DOI: https://doi.org/10.33541/jet.v7i3.3118

INTRODUCTION

In recent years, English-medium instruction (EMI) has become a growing global phenomenon (Dearden, 2015). Macaro, Curle, Pun, An and Dearden (2018) defined EMI as "the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language (L1) of the majority of the population is not English" (p.9). According to Airey (2020), EMI courses have been expanding at a fast pace in many non-Anglophone countries since 2003 or earlier. In Taiwan where English is not an official language, one main reason for the rapid growth of EMI programs in higher education is the prospects of advance in global university rankings. Another reason is the widespread belief to varying extents that EMI provides immersion that facilitates the incidental learning of English (Yeh, 2014). It is generally held that domestic students can acquire disciplinary knowledge while simultaneously enhancing English abilities to prepare themselves for global competition. The belief of achieving this dual goal can date back to the success of French immersion education in Canadian schools in the 1960s, leading to mastery of both French and school subjects (Brinton, Snow & Wesche, 1989). In terms of incidental learning, previous studies have provided some evidence that constant exposure to English results in lexical gains over time (Brown, Waring & Donkaewbua, 2008; Pellicer-Sanchez & Schmitt, 2010; Vidal, 2011).

However, not all EMI programs provide the same English immersion as those in the core Anglosphere, where teachers and local students are predominantly native speakers of English and incoming non-native students benefit from immersion. In Taiwan's higher education, a significant number of faculty members in EMI programs are Taiwanese teachers, despite the fact that they have earned their PhD in one of the core Anglosphere and can speak English fluently. Quite a few international academics with a certain expertise are non-native speakers of English and a majority of international students come from Southeast and Northeast Asia (Ministry of Education, 2020), but they are key contributors to campus Englishization. Outside of the EMI classrooms, domestic students are exposed to their mother tongue (Taiwanese Mandarin) most of the time. Therefore, immersion in EFL EMI programs should not be viewed in the same light as that in majority-native-English-speaking countries.

Each EMI context has its own features and constructs its own set of ideologies that regulate the assessment of subject content and language learning (Baker & Huttner, 2019; Doiz & Lasagabaster, 2020; Lasagabaster, 2018). In consideration of fostering students' disciplinary literacy, EMI teachers should ideally have both specialist knowledge and linguistic knowledge (Schmidt-Unterberger, 2018). But in an investigation into 70 European universities with EMI programs, O'Dowd (2018) discovered that the concern of English improvement often plays second fiddle compared with that of disciplinary knowledge development. Jiang, Zhang and May (2019) also found out that under the constraint of class time, subject teachers generally regard adequate delivery of specialist content as their primary responsibility and do not consider English teaching within their remit.

Similarly, in the current EMI setting, where there may not be much attention paid to language per se by subject teachers, English-medium specialist textbooks turn to be an

important source of input. Namely, they offer non-English majors a channel for exposure to English. In view of a lack of content and language integrated learning (Airey, 2012; Block & Moncada-Comas, 2019; Dearden & Macaro, 2016), the researcher targeted university textbooks as a research focus because they are first and foremost learning material of specialist knowledge. When students immerse themselves in textbooks, unknown English words may impede their comprehension of subject content. However, if a new word repeatedly occurs, it has the potential to 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 further narrowed the focus on vocabulary with a particular concern about the extent to which EFL students with EMI may develop their English vocabulary and if there is a difference between different subjects in potential vocabulary growth.

Amongst a range of undergraduate disciplines, the researcher set out from accounting and civil engineering, since they represent two different fields of study and many universities provide the degree programs of these two subjects. This study sought to address two research questions (RQ):

RQ1: After finishing EMI undergraduate programs, what vocabulary levels may EFL accounting and civil engineering students attain?

RQ2: Beyond the first 3000 word families, how many words from English-medium specialist textbooks may accounting and civil engineering undergraduates encounter often enough for learning to occur?

LITERATURE REVIEW

Early research like Goulden, Nation and Read (1990) conjectured that well-educated native speakers of English have a vocabulary of around 20,000 word families. This rough estimate was based upon a rule of thumb that native speakers add circa 1000 new words a year to their lexical repertoire. However, despite years of study, the vocabulary goal of learning 20,000 word families is still far beyond the reach of EFL learners like ours. Fortunately, knowledge of the first 8000—9000 word families would suffice to provide 98% text coverage (with the encounter of two unknown words per hundred words) and allow the average person to read a wide range of unsimplified texts without too many unknown words being a substantial hurdle, according to Nation (2006). Compared with 20,000 word families, this vocabulary goal appears more feasible.

To compute vocabulary amounts contained in texts, Nation (2020) has endeavored to compile a large scale of word-family lists for over many years. So 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 and one-billion-plus-word Corpus of Contemporary American English (hereafter referred to as BNC/COCA). The rationale behind the frequency ranking is that frequently-occurring words stand a favorable opportunity of being 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 (the 1st—3rd 1000 levels), mid-frequency (4th—9th 1000) and low-frequency vocabulary (beyond the 9th 1000). They further illustrated the essentiality of mid-frequency vocabulary for proficient use in support of Nation's (2014) 9000-word-family target for wide reading. Concerning the validity of the BNC/COCA twenty-five 1000-word-family lists, Dang and Webb (2016) 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 others. In this study, the researcher also utilized the ranked BNC/COCA word-family lists to measure the vocabulary levels and amounts contained in university textbooks.

Measuring the vocabulary level of a text can be approached from text coverage, which Nation (2006) defined as "the percentage of running words in the text known by the reader" (p. 61). For instance, when learners know 98% of the total words of a text (i.e. 98% text coverage), it signals that they would meet two unknown words per hundred words being read. Although text coverage does not equate to comprehension degrees, Schmitt, Jiang and Grabe (2011) detected that the two have a positive linear relationship, which shows that as the percentage of known words in a text increases, the likelihood of better comprehension also increases. For adequate comprehension of a text, Laufer and Ravenhorst-Kalovski (2010) as well as 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 English vocabulary levels of accounting and civil engineering textbooks. When the vocabulary level of a text is measured at 98% text coverage, the outcome also reflects the vocabulary threshold needed for adequate comprehension of that text.

As aforementioned, pertinent to immersion 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, Cobb & Meara, 1998). However, lexical researchers have diverged on the minimal repetitions that are required for incidental learning. Early research reported that 6 repetitions would be enough to incidentally learn a new word (Rott, 1999). However, in Vidal's (2011) experiments on the effects of reading and listening on incidental vocabulary acquisition, the greatest number of vocabulary gains occurred in the reading groups under 2 to 3 repetitions. On the contrary, Waring and Takaki (2003) maintained that a minimum of 20 repetitions would be necessary for incidental learning. In a similar vein, Pellicer-Sanchez and Schmitt (2010) upheld that 10 to 15 repetitions would be required to gain rich knowledge of a word. Referring to prior studies, Nation (2014) took a middle ground and used 12 repetitions as a threshold to measure reading amounts for word learning. Following Nation (2014), the researcher also chose 12 times as the cutoff to measure the number of word families beyond the first 3000 with 12+ occurrences in accounting and civil engineering textbooks.

In order to encounter new words enough times for learning to occur, the amount of input must be large enough. Nation (2014) selected novels and Author (2019, 2020) chose

VOA news articles and TED Talks English transcripts for voluminous reading. To meet most of the first 9000 word families 12+ times, learners would need to read 3 million words of novels at the minimum, as per Nation (2014). With a particular focus on mid-frequency vocabulary learning (specifically, the encounter with 800+ word families from each of the 4th to 9th 1000 levels 12+ times), Author (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 least.

Similarly, when EFL students finish their EMI degree programs, they will have read a rather large amount of English-medium specialist textbooks, which can be comparable to voluminous reading. In view of this, the researcher was concerned about accounting and civil engineering majors' English vocabulary development in four years of college study.

METHOD

Compiling a Textbook Corpus of Accounting and Civil Engineering Core Courses

Referring to the accounting and civil engineering undergraduate programs offered by elite universities, the researcher identified some core courses that most accounting and civil engineering departments require their students to take regardless of selected areas of specialization in the field. For accountancy, the core curricula include accounting fundamentals, accounting information systems, intermediate accounting, managerial accounting, cost accounting, financial accounting, advanced accounting, taxation and auditing. For the undertaking of construction and maintenance of public and private works, civil engineering students need to take these required courses: calculus, general physics, engineering mathematics, foundation engineering, surveying, engineering graphics, soil mechanics, fluid mechanics, mechanics of materials, mechanics of solids, geomatics as well as construction, transportation, structural, environmental, architectural, hydraulic, geotechnical and seismic engineering.

Based on the catalogues of internationally reputable textbook publishers for accounting and civil engineering majors as well as the reading lists that accounting and civil engineering professors provide for their students, the researcher noted down a list of textbook titles. The identical textbooks were put in the priority list for subsequent screening. To ensure popularity and wide use, the accounting textbooks with the 11th+ edition and the civil engineering textbooks with the 4th+ edition were selected. In consideration of at least one semester use, the textbooks for further inclusion in the candidate list must have 700-800 pages at the minimum. The selection criteria were admittedly arbitrary but they were decided after a series of comparison among the textbooks. After that, five accounting teachers and six civil engineering teachers were consulted to confirm influential textbooks for each course. The textbooks without the agreement of at least two teachers by checking yes on one of the two questions (regarding if they know this textbook and if they would use it in class) were removed from the candidate book list.

From the shortlist, the researcher randomly selected one textbook for each course, totalling 9 accounting textbooks and 19 civil engineering textbooks. It was assumed that accounting and civil engineering majors with EMI would read these specialist textbooks for their required courses. All the sampled textbooks in PDF were downloaded from paid electronic databases for research purposes and were then saved as plain texts in UTF-8. After the deletion of references as well as front matter and back matter, the accounting textbook corpus had around 4 million running words while the civil engineering corpus had 6 million tokens. To make a comparison on an equal footing in terms of the reading amount, the civil engineering textbook corpus was further cut down by one-third for each book. Table 1 shows the composition of each corpus.

Table 1
Composition of the Accounting and the Civil Engineering Textbook Corpora

Accounting core courses	Tokens	Accounting core courses	Tokens
Accounting fundamentals	513,626	Intermediate accounting	708,476
Advanced accounting	414,011	Managerial accounting	344,707
Cost accounting	427,339	Financial accounting	325,852
Auditing	454,987	Taxation	446,772
Accounting info systems	364,243	Total:	4,000,013

Civil engineering core courses	Tokens	Civil engineering core courses	Tokens
Engineering mathematics	93,861	Calculus	294,685
General physics	284,343	Foundation engineering	83,863
Engineering graphics	85,291	Surveying	248,096
geomatics	340,214	Mechanics of solids	157,670
Mechanics of materials	354,120	Soil mechanics	204,141
Fluid mechanics	192,824	Transportation engineering	213,084
Architectural engineering	387,471	Structural engineering	160,621
Hydraulic engineering	135,748	Construction engineering	256,223
Environmental engineering	193,139	Seismic engineering	114,004
Geotechnical engineering	200,839	Total:	4,000,237

Procedures

To analyze the lexical profiling of accounting and civil engineering textbooks, the vocabulary analysis program AntWordProfiler (Anthony, 2014) was installed with the ranked BNC/COCA twenty-five 1000-word-family lists as well as the three ever-growing lists of proper nouns, acronyms and transparent compounds complied by Nation (2017). Along the BNC/COCA word-frequency scale, the researcher first measured the vocabulary levels of accounting and civil engineering textbooks by calculating how many of the ranked twenty-five 1000-word-family lists from the 1st 1000 were needed until the text coverage of each 1000 word families plus proper nouns, acronyms and transparent compounds accumulated to 98%. Meanwhile the vocabulary thresholds of accounting and civil engineering textbooks for adequate comprehension were extrapolated, when the cumulative coverage reached 98%.

In the middle of implementing the AntWordProfiler, the words that were classified as the 'Words NOT Found In Base Lists' (hereafter named as off-list) were further examined. If off-list words were personal or geographical names, they were added to the existing proper noun list. Moreover, a fairly large number of acronyms appear in the accounting and civil engineering textbooks. An acronym is usually glossed in appendix with the full form or the explanation thereof can be found in context in its first appearance. The acronyms in the off-list were supplemented to the acronym list. The hyphen of a hyphenated transparent compound was removed so that it would not be treated as an off-list word by the program. A closed transparent compound (without a hyphen or a space) may be mistaken as an off-list word as well. The closed transparent compounds in the off-list were added to the existing transparent compound list.

Apart from the BNC/COCA word family lists, some researchers (e.g. Nation 2006; Nurmukhamedov & Webb, 2019; Webb & Rodgers, 2009) took proper nouns, acronyms and transparent compounds into account when adding up the text coverage of each word list. Excluding the text coverage of these three categories would inflate the vocabulary level of a text as well as the vocabulary demand of that text. In view of this, the researcher also added their text coverage to that of the base word lists until the cumulative coverage achieved 98%.

Likewise, to prevent the vocabulary amount of a text from being overestimated, word families were decided as a counting unit. Repeated exposure to a new word may increase the likelihood of incidentally learning that word. In the present setting where college entrance exams cover the English subject test, university matriculated students already have some ideas of English word-building rules, which contribute to the learning of an unknown word (Nagy et al., 1989). When a word and its family members occur in a range of contexts, learners with knowledge of inflectional and derivational morphemes would recognize the word family, guess its meaning from context and consolidate knowledge from multiple encounters. The occurrences of a word family' stem plus its inflected and derived forms were counted together so that the combined frequency would indicate the number of repeated exposure to that word family.

FINDINGS AND DISCUSSION

Vocabulary Levels of Accounting and Civil Engineering Textbooks

As mentioned earlier, English-medium university textbooks are the primary source of English input for EFL students, so RQ1 'After finishing EMI undergraduate programs, what vocabulary levels may EFL accounting and civil engineering students attain?' may be addressed from the vocabulary levels of their specialist textbooks. Table 2 provides a snapshot of the vocabulary levels of accounting and civil engineering textbooks at 98% text coverage and the vocabulary distribution among the BNC/COCA word lists.

The accounting and civil engineering textbook corpora contained approximately equal amount of words (both having about 4 million words, see Table 1). As shown in Table 2, the first 3000 word families, which Schmitt and Schmitt (2014) urged as a yardstick for high-frequency vocabulary, accounted for 91.88% and 83.93% of the total words of the accounting corpus and civil engineering corpus respectively. This manifests

the relative importance of knowledge of the first 3000 word families because their text coverage was much larger than that of the total combined remaining 1000-word-family lists by a large margin. In Taiwan, mastery of the most frequent 3000 word families has been generally considered as the graduation benchmark for senior high school students (Grade 12), since the English test in the General Scholastic Ability Test involves a 4,500-word vocabulary (College Entrance Examination Center, 2021).

Table 2
Vocabulary Levels of Accounting and Civil Engineering Textbooks along the BNC/COCA Scale

BNC/COCA word lists	Coverage of accounting textbooks in tokens (%)	Cumulative coverage of accounting textbooks	Coverage of civil engineering textbooks in tokens (%)	Cumulative coverage of civil engineering textbooks
Proper nouns	2.25%	2.25%	2.80%	2.80%
Acronyms	0.32%	2.57%	1.49%	4.29%
Compounds	0.46%	3.03%	1.72%	6.01%
1st 1000	62.95%	65.98%	58.72%	64.73%
2nd 1000	18.57%	84.55%	14.68%	79.41%
3rd 1000	10.36%	94.91%	10.53%	89.94%
4th 1000	2.14%	97.05%	3.21%	93.15%
5th 1000	1.01%	**98.06%**	2.12%	95.27%
6th 1000	0.58%	98.64%	1.01%	96.28%
7th 1000	0.33%	98.97%	0.57%	96.85%
8th 1000	0.17%	99.14%	0.48%	97.33%
9th 1000	0.07%	99.21%	0.33%	97.66%
10th 1000	0.14%	99.35%	0.57%	**98.23%**
11th 1000	0.05%	99.40%	0.21%	98.44%
12th 1000	0.16%	99.56%	0.38%	98.82%
13th 1000	0.04%	99.60%	0.13%	98.95%
14th 1000	0.02%	99.62%	0.08%	99.03%
15th 1000	0.01%	99.63%	0.11%	99.14%
16th 1000	0.03%	99.66%	0.10%	99.24%
17th 1000	0.05%	99.71%	0.15%	99.39%
18th 1000	0.01%	99.72%	0.04%	99.43%
19th 1000	0.01%	99.73%	0.06%	99.49%
20th 1000	0.00%	99.73%	0.05%	99.54%
21st 1000	0.01%	99.74%	0.03%	99.57%
22nd 1000	0.00%	99.74%	0.02%	99.59%
23rd 1000	0.01%	99.75%	0.02%	99.61%
24th 1000	0.01%	99.76%	0.01%	99.62%
25th 1000	0.02%	99.78%	0.02%	99.64%
Off-list	0.22%	100%	0.36%	100%
Total				

Note: The bolded figures indicate the vocabulary level at which the cumulative text coverage has already reached 98%.

Despite the significance of the first 3000 word families, EMI civil engineering freshmen with only this vocabulary size may have some difficulty in reading their specialist textbooks as opposed to accounting freshmen with EMI. The cumulative 94.91% and 89.94% coverage in accounting and civil engineering textbooks in turn provided by the first 3000 word families plus proper nouns, acronyms and transparent compounds means unfamiliarity with 5.09% and 10.06% of the total words in accounting and civil engineering textbooks respectively. In other words, accounting freshmen with knowledge of the most frequent 3000 word families would encounter one unknown word in roughly every 20 words, while civil engineering freshmen would meet one unknown word in every 10 words or less. In terms of interruption frequency to the flow of reading (e.g. dictionary or glossary look-up or guessing meaning from context), the slightly below 95% coverage in accounting textbooks (i.e. 94.91%) that knowledge of the first 3,000 word families plus proper nouns and transparent compounds would provide may still fall within the parameters where good comprehension occurs (Nation, 2006). In contrast, encountering one unknown word in every 10 words (in less than one line of text) may make civil engineering freshmen become discouraged.

By the 5th 1000 level, slightly over 98% coverage (98.06%) in accounting textbooks was attained. Specifically, knowledge of the most frequent 5000 word families plus proper nouns and so forth would suffice to provide 98% text coverage. In comparison with Nation's (2014) report that a minimum of 9000-word-family vocabulary is needed to read a diversity of unsimplified texts, accounting textbooks are a lot easier to read, since they use a condensed vocabulary, converging within the range of the first 5000 word families. To put it in perspective, if EMI accounting students have a 5000-word-family vocabulary, they would be able to read their specialist textbooks at quite ease in terms of not facing a daunting amount of dictionary look-up (viz. one unknown words in every 50 words). However, this also implies that studying English-medium accounting textbooks alone may not result in EFL accounting majors meeting a wide variety of English words and therefore may not lead to great vocabulary expansion.

By the 10th 1000 level, 98.23% coverage in civil engineering was met. Compared with accounting textbooks, civil engineering textbooks had the wider dispersion of words, spreading to the 9th —10th 1000 level at 98% coverage. This implies that civil engineering textbooks would rival novels, of which the vocabulary level reaches the 8th to 9th 1000 at 98% text coverage (Nation, 2006). By analogy, constant immersion in English-medium civil engineering textbooks through voluminous reading may bring about a beacon of considerable vocabulary expansion.

The apparent discrepancy between the 5th 1000 level and the 9th —10th 1000 level indicates that majoring in civil engineering is more vocabulary-demanding than specializing in accounting. For fluent, easy reading, a threshold of 9000—10,000 word families is suggested for EMI civil engineering students. In contrast, EFL students planning to study EMI accounting undergraduate programs may need to know a minimum of the first 5000 word families in order to perform reading tasks well. The striking difference in vocabulary levels also hints that different EMI subject programs may give

rise to different amounts of vocabulary learning if matriculating freshmen start with a vocabulary of the first 3000 word families only.

In answer to RQ1, Table 2 shows that accounting textbooks reached the 5th 1000-word-family level while civil engineering textbooks stretched to the 9th–10th 1000 level at 98% text coverage. The latter gives us a beacon of hope concerning what vocabulary level civil engineering students may attain after finishing their degree program with EMI. If EMI civil engineering students ideally increase their vocabulary to 9000—10,000 word families in line with the vocabulary level of their specialist textbooks, with this vocabulary size they would be able to read a diversity of authentic texts. However, the results for RQ2 paint a different picture.

Potential Vocabulary Growth in EMI Accounting and Civil Engineering Programs

This study aimed at vocabulary development beyond the most frequent 3000 word families, since Taiwanese university freshmen ought to have had knowledge of the English base words after passing the nationwide college entrance exam with the English subject included. As reviewed in the literature, the learning of 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 3 at the cutoff of 12 repetitions provides an overall picture of the amount of potential vocabulary growth for EMI accounting and civil engineering students as a result of constant exposure to English-medium specialist textbooks.

Table 3

Number of Word Families beyond the First 3000 Word Families Occurring 12+ Times

BNC/COCA word-frequency scale	4 million words of accounting textbooks across 9 core courses	4 million words of civil engineering textbooks across 19 core courses	3 million words of novels (Nation 2014, p.6)	6 million words of VOA news (Author 2019, p.417)	4.8 million words of TED Talks transcripts (Author 2020, p.197)
Vocabulary level at 98% text coverage	5th 1000	9th—10th 1000	8th—9th 1000	6th 1000	5th—6th 1000
4th 1000	464	624	945	996	800+
5th 1000	292	486	929	985	800+
6th 1000	213	363	904	962	800+
7th 1000	144	292	857	898	800+
8th 1000	88	242	817	851	800+
9th 1000	63	199	805	805	800
Mid-frequency vocabulary (4th –9th 1000) subtotal	1,264	2,206	5,257	5,497	4,800+
Low-frequency vocabulary (10th—25th 1000) subtotal	215	1,191	N.A.	N.A.	N.A.
Total	1,479	3,397	N. A.	N. A.	N.A.

DOI: https://doi.org/10.33541/jet.v7i3.3118

Table 3 shows that the 4-million-token accounting textbook corpus across nine core courses contained 1,264 word families from the BNC/COCA 4th to 9th 1000 levels and 215 word families from the 10th to 25th 1000, totaling 1,479 word families beyond the first 3000 appearing 12+ times. When finishing nineteen compulsory courses (circa 4 million words), civil engineering undergraduates would meet 2,206 mid-frequency word families and 1,191 low-frequency word families, totaling 3,397 word families with 12+ repetitions. As shown in the potential increase of 3,397 word families from reading English-medium civil engineering textbooks greater than that of 1,479 word families from reading accounting textbooks by 2.3 times, there is a difference between different major subjects in the amount of incidental vocabulary growth.

Table 3 shows that the 4-million-token accounting textbook corpus across nine core courses contained 1,264 word families from the BNC/COCA 4th to 9th 1000 levels and 215 word families from the 10th to 25th 1000, totaling 1,479 word families beyond the first 3000 appearing 12+ times. When finishing nineteen compulsory courses (circa 4 million words), civil engineering undergraduates would meet 2,206 mid-frequency word families and 1,191 low-frequency word families, totaling 3,397 word families with 12+ repetitions. As shown in the potential increase of 3,397 word families from reading English-medium civil engineering textbooks greater than that of 1,479 word families from reading accounting textbooks by 2.3 times, there is a difference between different major subjects in the amount of incidental vocabulary growth.

If knowledge of the first 9,000 word families, which would provide 98% coverage of a variety of unsimplified texts according to Nation (2014), is the vocabulary goal, we can just focus our attention to mid-frequency vocabulary (4th to 9th 1000) for the time being. As shown in Table 3, only 1,264 (21.07%) and 2,206 (36.77%) out of the 6000 mid-frequency word families occurred 12+ times in the respective 4-million-token accounting and civil engineering textbook corpora. The outcomes are far below the goal of learning most of the mid-frequency words. In contrast with specialist textbooks, reading 3 million words of novels, 6 million words of VOA news or 4.8 million words of TED Talk English transcripts would enable EFL learners to encounter 5,257 (87.6%), 5,497 (91.6%) and 4800+ (80%+) mid-frequency word families respectively (Table 3).

Apparently, even though EMI students finish reading circa 4 million words of English-medium specialist textbooks such as accounting and civil engineering, they would still not encounter most of the 1000 word families (>800) enough times from the 4th 1000 onwards. Table 3 shows that with the advance towards the upper BNC/COCA word-frequency scale, the number of mid-frequency vocabulary (4th to 9th 1000) and low-frequency vocabulary (10th to 25th 1000) for potential incidental learning (appearing 12+ times) becomes smaller and smaller. Compared with novels, news and speeches, the vocabulary recycling of specialist textbooks is much stronger, which may reduce EFL students' vocabulary load when they study from one compulsory course to another. In particular, only a small mid-frequency and low-frequency vocabulary is frequently used in accounting textbooks (see Table 3). In comparison with novels (providing 5,257 mid-frequency word families with sufficient occurrences), news (5,497) and talks (4800+), accounting majors would make much smaller progress in mid-frequency word learning (merely 1,264 word families). Although civil engineering textbooks reach the 9th—10th 1000 level, they would not result in greater vocabulary growth as we have expected of from voluminous reading, as far as a potential growth of 2,206 out of 6000 mid-frequency word families is concerned (only 36.77%). This may be because a large number of

different words are used in novels, news articles and speeches due to the diversity of topic areas, which brings about the richness of vocabulary.

Civil engineering textbooks contain many frequent sub-technical and highly-technical words, ranking in the upper bands of the BNC/COCA word-frequency scale (so called mid-frequency and low-frequency words). Initially, matriculating civil engineering students with EMI may be discouraged by many rarely-seen-before technical words (e.g. flange, aquifer and torsion, to name but a few). However, over time, civil engineering students would get acquainted with these frequent new word with the aid of in-class lectures and textbook illustrations. This suggests that EMI civil engineering freshmen may have an imminent need for a list of frequent civil engineering specialized words for learning. The word list is worth developing but beyond the present research focus. Compared with civil engineering novices, accounting freshmen with EMI are luckier. Many words beyond the first 3000 word families that are used frequently in the accounting register (e.g. invoice, merchandize, depreciate, deduct, incur, defer, debit and receipt, to name but a few) do not show a technical flavor in a strong sense. This is because these frequent words that slip in and out of daily conversation and business domain talk turn out to be common words and therefore they may not be so unfamiliar to accounting novices.

Table 3 answers RQ2 'Beyond the first 3000 word families, how many words from English-medium specialist textbooks may accounting and civil engineering undergraduates encounter often enough for learning to occur?', demonstrating that a total of 1,479 and 3,397 word families beyond the base words would be encountered sufficient times for potential vocabulary growth when accounting and civil engineering undergraduates complete their required core courses.

CONCLUSION

This research was a preliminary comparative study on two EMI undergraduate programs in an EFL setting. Its goal is twofold: to measure (1) the vocabulary levels of English-medium accounting and civil engineering textbooks and (2) the amount of vocabulary beyond the first 3000 word families contained in accounting and civil engineering textbooks. Generally, accounting textbooks reached the 5th 1000-word-family level at 98% text coverage while civil engineering textbooks stretched to the 9th—10th 1000 level.

Results show that in the respective 4-million-token accounting and civil engineering textbook corpora, only 1,479 word families from the 4th to 25th 1000 levels occurred in the accounting corpus 12 times or more with the inclusion of 1,264 mid-frequency word families (4th to 9th 1000) versus 3,397 word families beyond the first 3000 appearing in the civil engineering corpus 12+ times, including 2,206 word families from the mid-frequency bands.

Even if EMI students like accounting and civil engineering majors complete their compulsory courses, continual reading of their English-medium specialist textbooks will still not enable them to meet most of the 6000 mid-frequency word families often enough for learning to occur. This signals that an academic program taught in full English does not necessarily ensure maximum encounters with mid-frequency words. In view of a convergent vocabulary frequently used in specialist textbooks, it is highly likely that non-English subject majors' vocabulary size would level off at the first 3000 word families plus approximately one-fifth to one-third of the mid-frequency vocabulary (1264/6000 to

2206/6000), 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 research has been to raise this awareness. When EMI has become a nationwide trend at tertiary education in an EFL setting like Taiwan, the present results contribute to an understanding of what expectations we may reasonably have of the vocabulary development in an EMI program. With respect to the 9000-word-family goal for reading a wide range of authentic texts, one advice for EFL EMI students is to constantly read English newspapers, novels and all sorts of English articles.

Though this study contributes to the literature of EMI research, yet it has been worked within a narrow focus on the fields of accounting and civil engineering. The results 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 inspirations for future qualitative analyses of EFL learners' lexical needs and perceptions of English-medium specialist textbooks regarding vocabulary load and reading difficulty.

ACKNOWLEDGEMENTS

This research received the grant from Taiwan's Ministry of Science and Technology, project number MOST 110-2410-H-214-007.

REFERENCES

Airey, J. (2012). "I don't teach language." The linguistic attitudes of physics lecturers in Sweden. *AILA Review, 25*(1), 64–79.

Airey, J. (2020). The content lecturer and English-medium instruction (EMI): epilogue to the special issue on EMI in higher education. *International Journal of Bilingual Education and Bilingualism, 23*(3), 340-346.

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software

Baker, W., & Huttner, J. (2019). 'We are not the language police': Comparing multilingual EMI programmes in Europe and Asia. *International Journal of Applied Linguistics, 29*(1), 78–94.

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.

Brinton, D. M., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. Boston: Heinle & Heinle.

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.

College Entrance Examination Center (2021). The GSAT English. Retrieved from: https://www.ceec.edu.tw/en/xmdoc/cont?xsmsid=0J180519944235388511

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics, 167*(2), 132–158.

Dearden, J. (2015). English as a medium of instruction: A growing global phenomenon. British Council Report. Retrieved from

https://www.britishcouncil.org/education/ihe/knowledge-centre/english-language-higher-education/report-english-medium-instruction

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.

Doiz, A., & Lasagabaster, D. (2020). Dealing with language issues in English-medium instruction at university: A comprehensive approach. *International Journal of Bilingual Education and Bilingualism, 23*(3), 257-262.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341–363.

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.

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.

Lasagabaster, D. (2018). Fostering team teaching: Mapping out a research agenda for English-Medium Instruction at university level. *Language Teaching, 51*(3), 400–416.

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.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching, 51*(1), 36-76.

Ministry of Education. (2020). Foreign student number in 2020. [Data file]. 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.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Reviews, 63*(1), 59–82.

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.

Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

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

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching, 52*(2), 188-200.

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.

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*(1), 31–55.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition, 21*(4), 589–619.

Schmidt-Unterberger, B. (2018). The English-medium paradigm: a conceptualisation of English-medium teaching in higher education. *International Journal of Bilingual Education and Bilingualism, 21*(5), 527–539.

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.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258.

Waring, R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language,15*(2), 130–163.

Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427.

Yeh, C.-C. (2014). Taiwanese students' experiences and attitudes towards English-medium courses in tertiary education. *RELC Journal, 45*(3), 305–319.