English-Medium Instruction: How Many Mid-Frequency Words May EFL Accounting Undergraduates Learn?

Wenhua Hsu

whh@isu.edu.tw

I-Shou University, Taiwan

Abstract

In recent years, higher education programs with English as a medium of instruction (EMI) have been expanding at a swift pace in Taiwan, where Taiwanese Mandarin is the official language. Behind this rapid growth is the widespread belief that EMI provides immersion in English, which facilitates incidental learning of the target language. In this paper, English-medium textbooks were targeted as a research focus, since they are first and foremost learning material of core knowledge and offer students a channel for exposure to English. The researcher compiled a 4.1-million-token textbook corpus of accounting core courses and measured the vocabulary levels thereof along the word-frequency scale of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA). Then she sought to estimate how many mid-frequency word families accounting majors with EMI can encounter often enough to have an opportunity of learning them. Results showed that accounting textbooks reached the BNC/COCA 3rd–4th 1000-word-family level at 95% text coverage and stretched to the 5th–6th 1000 at 98% coverage. Of the 6000 mid-frequency word families (4th–9th 1000), only 1,274 word families occurred 12+ times. This frequency was assumed as a benchmark for incidental learning to occur. For EMI practitioners who are concerned with their students' English vocabulary development, the results can serve as a reference for future investigations into other disciplines.

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

Introduction

In recent years, English as a/the medium of instruction or English-medium instruction (EMI) has been gaining increasing popularity among higher education institutions in Taiwan, where Taiwanese Mandarin is the official language. Perceiving the upside of EMI as a means to enhance domestic students' English abilities for global competition, Taiwan's Ministry of Education (MOE) has been promoting EMI courses for university internationalization (Ministry of Education, 2017). The number of EMI degree programs and curricula has been regarded as one of the internationalization indices in university evaluation (Chen, 2011; Kao & Tsou, 2017; Li & Wu, 2018).

For curriculum internationalization, there is a growing trend in recruiting international academics with a specialty in a certain area, who are not necessarily native speakers of English, but they are key contributors to boosting speaking English on campus. In the current EMI programs, both international faculty and quite a few Taiwanese subject teachers with a Ph.D. degree from one of the core Anglosphere can speak English fluently. International students mainly come from Vietnam, Indonesia, Malaysia, Japan, Hong Kong and Mongolia to take degree courses alongside Taiwanese students (Ministry of Education, 2020).

Despite nearly twice the tuition fees for domestic students, another factor for the rapid growth of EMI programs is the widespread belief that EMI provides immersion opportunities, which facilitate incidental learning of English (Yeh, 2014). It is generally held that students can acquire disciplinary knowledge and enhance English abilities simultaneously. Past research has provided some evidence that continuous exposure to English results in lexical gains over time (Brown et al., 2008; Pellicer-Sanchez & Schmitt, 2010; Vidal, 2011).

However, not all EMI programs in English as a foreign language (EFL) contexts provide the same immersion as those in the core Anglosphere, where teachers and students are mostly native speakers of English. Previous studies have raised a number of concerns in relation to language learning in EMI programs. They are mainly about how much English proficiency students with EMI develop (Rogier, 2012; Yang, 2015) as well as whether native or non-native English-speaking subject teachers are critical to language learning (Inbar-Lourie & Donitsa-Schmidt, 2019; Kuteeva, 2020). 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 a survey of 70 European universities offering EMI programs, O'Dowd (2018) unearthed the phenomenon that the concern of language improvement usually plays second fiddle compared with disciplinary knowledge development. This may be because subject teachers do not see English teaching within their remit, resulting in a lack of deliberate language instruction as well as a lack of integration of content and language learning (Airey, 2012; Block & Moncada-Comas, 2019; Dearden & Macaro, 2016).

As aforementioned, in the current EMI context, non-native subject teachers may constitute a significant proportion of the faculty members. Without linguistic training, they may be more attentive to subject content in limited class time; meanwhile, outside of the EMI classrooms students have more exposure to their mother tongue than English. Such a context may contradict a typical immersion mode and this contradiction has led the researcher to target English-medium specialist textbooks as a research focus. In the classroom, specialist textbooks are first and foremost learning material of core knowledge and as a source of input, English-medium textbooks offer students a channel for exposure to English. Nevertheless, unknown words challenge EFL students while reading an English-medium specialist textbook.

In the process of learning a new language, vocabulary learning is an indispensable part, since a rich vocabulary plays a crucial role in performing language tasks (Qian, 2002; Uchihara & Clenton, 2020). The concern of this research was therefore with vocabulary expansion with a particular focus on mid-frequency words. Among a variety of academic disciplines, this study set out from accounting in that it represents one of the college majors and many business schools offer this discipline. It sought to answer two research questions:

1. What are the vocabulary levels of English-medium textbooks of accounting core courses?
2. Beyond the most frequent 3000 word families, how many mid-frequency words may EFL accounting undergraduates learn from their specialist textbooks?

Literature Review

Vocabulary Levels and Lexical Text Coverage

Regarding English vocabulary amount, Goulden, Nation and Read (1990) reported that there were well over 54,000 word families in the Webster's Third International Dictionary, excluding foreign words, proper nouns and abbreviations. They also figured out that native English-speaking university graduates have a vocabulary of circa 20,000 word families based on the presumption that people learn about 1000 new words per year.

Drawing upon the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), Nation (2017) compiled 25,000 word families and classified them into 25 frequency bands with each containing 1000 word families. The BNC/COCA word-frequency scale is a useful benchmark in measuring the vocabulary level of a text, because the 25,000 word families are ranked based on the frequency of occurrence in enormous, genre-balanced corpora. The reasoning behind vocabulary levels is that high-frequency words are more likely to be met and learned than low-frequency words (Nation, 2006).

Schmitt and Schmitt (2014) advocated that the first 3000 word families be a yardstick for high-frequency vocabulary and labelled the first 4000 to 9000 word families as mid-frequency vocabulary and the words after the first 9000 as low-frequency vocabulary. They further stressed the importance of learning mid-frequency words according to Nation's (2006) gauge that knowledge of the first 9000 word families plus proper nouns would provide 98% lexical coverage of a variety of texts (i.e., knowing 98% of the total words of a text). Relative to 20,000 word families (a well-educated native speaker's lexicon), mastery of the first 9000 word families would allow the average person to read unsimplified texts well with the encounter of two unknown words per 100 words.

This has brought a beacon of hope to EFL learners. The most frequent general words occur across all kinds of texts. If EFL learners already know high-frequency words such as the first 3000 word families, they would know a large proportion of the words in a text (89% to 95%) as per Nation (2006) and Schmitt and Schmitt (2014). Subsequent to the first 3000 word families, covering the lexical

shortage would largely depend on mastering words from the mid-frequency band and they should be given more attention, as Schmitt and Schmitt (2014) have urged.

Nation (2006) defined lexical text coverage as "the percentage of running words in the text known by the reader" (p. 61). It is a putative measure in association with reading comprehension. Applying statistical methods to their data, Laufer and Ravenhorst-Kalovski (2010) discovered that the amount of vocabulary knowledge accounted for more than half of the variance (64%) in the comprehension scores. Schmitt et al. (2011) also identified a positive linear relationship between lexical coverage and text comprehension, with the figures showing that the likelihood of good comprehension increases with more known words.

Through regression analysis to predict comprehension degrees, Hu and Nation (2000) reported that with a vocabulary achieving merely 80% text coverage, no participant could read fluently, but at 90% vocabulary command, a minority of participants were able to read well. When increasing to 95% lexical mastery, a majority of them attained good comprehension. At 100% coverage, most of the participants could read rather easily. Hirsh and Nation (1992), Hu and Nation (2000) as well as Nation (2006) maintained that 98% lexical coverage (leaving no more than 2% words unknown) is necessary, while Laufer (1989) recommended 95% coverage, allowing learners to read with some assistance. Taken together, 95% and 98% are generally thought to be lexical coverage targets for assisted reading and unassisted reading respectively.

Measuring the vocabulary level of a text hinges on the predetermined lexical text coverage. For instance, setting a vocabulary threshold at 98% coverage, English newspapers and novels measure up to the first 8000-9000 word-family levels, while at 95% coverage, they register the first 4000 word-family level (Nation, 2006). In other words, knowing 95% and 98% of the total words of English newspapers and novels as vocabulary goals entails a good command of the first 4000 word families as well as the first 8000-9000 respectively. This research also adopted these two putative percentage points as cut-offs for measuring the vocabulary levels of English-medium specialist textbooks.

Repetition and Incidental Learning

Amid various factors that affect learning, crucial to incidental learning is repetition, since the learning of a new word is seldom accomplished by merely one encounter (Horst et al., 1998). When the number of repeated occurrences of an unknown word increases, the likelihood of learning that word increases. Nevertheless, there is no definite number of repetitions that warrants the acquisition of a new word, since context clues, the word itself and the learner's English proficiency may impact its learning (Webb, 2008; Zahar et al., 2001).

Early research corroborated that six repetitions of a word would suffice for incidental learning (Rott, 1999). However, an optimistic finding can be seen in Vidal's (2011) survey that the biggest number of new words being learned occurred in two to three repetitions. In contrast, Waring and Takaki (2003) contended that to have a 50% probability of retaining a newly-learned word in three months'

time, 8+ repetitions would be needed. They further pointed out that 20+ repeated occurrences should be a prerequisite for incidental learning. Similarly, Webb (2007) upheld that 10+ repetitions would be necessary for learning the full meaning of a word. Pellicer-Sanchez and Schmitt (2010) also endorsed 10 to 15 repetitions to gain a rich knowledge of a word.

There seems to be little consensus in respect to the minimal repetitions that are required for incidental learning of a new word. Referring to previous studies, Nation (2014) took a middle-ground approach and adopted 12 repetitions as a calculation basis to compute the amounts of input necessary for learning most of the 2nd to 9th 1000 word families. This study also chose 12 times as a threshold for incidental learning to occur and thereby measured the number of mid-frequency word families appearing in accounting textbooks.

Needed Input for Most Mid-Frequency Words Learning

To assess the possibility of incidental learning through voluminous reading, Nation (2014) randomly selected 25 novels from Project Gutenberg and measured how much novel input is needed to encounter most of the words at a certain 1000-word-family level at least 12 times for learning to occur. According to Nation (2014), learners with knowledge of the first 3000 word families would need to read a minimum of 0.5 million words of novels in order to meet most of the word families from the 4th 1000 at least 12 times for a chance to acquire them. Learners with mastery of the first 4000 word families would need to read about one million words to encounter most of the words from the 5th 1000. Likewise, to meet most of the words from the 9th 1000 over 12 times, learners may read three million words of novels.

Following Nation (2014), Hsu (2019, 2020) used VOA news and TED Talk transcripts as input in lieu of novels to estimate minimal amounts to encounter 800+ word families from each of the 4th to 9th 1000 over 12 times. Results showed that VOA news articles generally reached the 6th 1000-word-family level at 98% text coverage, while TED talks of different topics mostly arrived at the 5th 1000 level except science- and technology-related talks extending to the 6th 1000. Hsu (2020) concluded that to develop mid-frequency vocabulary up to the 9th 1000 level, learners would need to read six million words of VOA news or 4.8 million words of TED Talks transcripts as opposed to three million words of novels.

Similarly, when EFL accounting majors with EMI finish their compulsory coursework, they will have read a fairly large amount of English-medium specialist textbooks. In view of voluminous textbook reading, the researcher was concerned about EMI accounting undergraduates' mid-frequency vocabulary growth within four years of college study.

Methodology

Building a Textbook Corpus of Accounting Core Courses

Referring to the curricula offered by the departments of accounting in prestigious universities, the researcher identified nine courses that most accounting departments require their students to take regardless of their selected area of specialization in accountancy. They include accounting fundamentals, one of the foundation courses (such as economics, statistics, and management) within the Faculty of Business and Management, as well as compulsory courses—intermediate accounting, advanced accounting, cost accounting, managerial accounting, financial accounting, auditing, taxation and accounting information systems.

Based on the catalogues of internationally reputable textbook publishers for accounting majors and certified public accountant exam-takers as well as the reading lists that accounting professors provide for their students, the researcher noted down a list of accounting textbook titles. The same books were put in the priority list for later screening. Subsequent to this book listing was the selection of textbooks with the 11th or higher edition to ensure their popularity. Moreover, in consideration of sufficient content, the textbook for inclusion in the candidate list must have 800+ pages in length. This arbitrary decision was made after a series of comparison among textbooks in terms of the number of editions and pages.

When the textbook listing was finalized, five accounting teachers were requested for assistance to confirm prestigious textbooks for each course. A survey list was given to them for ticking a box for each candidate textbook. Beside each textbook were two questions (concerning if they know this book and if they would consider using it in class) with three options (yes, no and uncertain) for them to tick. The textbooks without a consensus (without at least two of the five teachers ticking yes for either of the two questions) were removed. After a sequence of selections, the number of candidate textbooks across nine core courses was reduced to 51.

From the shortlist, the researcher chose one textbook for each course, totaling nine core textbooks (see Table 1). It was assumed that EMI accounting undergraduates would read them when taking compulsory courses. All the sampled accounting textbooks in PDF format were downloaded from e-book databases, which were subscribed by university libraries in Taiwan for educational and research purposes. They were then converted into plain texts in UTF-8. The textbook corpus had approximately 4.1 million tokens (running words), after the removal of front matter, references and back matter.

Table 1

Composition of the Accounting Textbook Corpus

Accounting core courses	Textbook titles (Authors)	Edition (Year)	Publisher	Tokens
Accounting fundamentals	Accounting principles (Weygandt, Kimmel & Kieso)	13th (2018)	Wiley	530,735
Intermediate accounting	Intermediate accounting (Kieso, Weygandt & Warfield)	17th (2019)	Wiley	725,585
Advanced accounting	Advanced accounting (Hoyle, Schaefer, Doupnik)	14th (2021)	McGraw-Hill	431,120
Managerial accounting	Managerial accounting: Creating value in a dynamic business environment (Hilton & Platt)	12th (2020)	McGraw-Hill	361,816
Cost accounting	Horngren's cost accounting: A managerial emphasis (Datar & Rajan)	17th (2021)	Pearson	444,448
Financial accounting	Financial accounting (Weygandt, Kimmel & Kieso)	11th (2020)	Wiley	342,961
Auditing	Auditing and assurance services: An integrated approach (Arens, Elder, Beasley & Hogan)	17th (2020)	Pearson	472,096
Taxation	Essentials of taxation: Individuals and business entities (Nellen, Cuccia, Persellin, Young & Maloney)	24th (2021)	Cengage Learning	463,881
Accounting information systems	Accounting information systems (Romney, Steinbart, Summers & Wood)	15th (2021)	Pearson	381,352
			Total:	4,153,994

As shown in Table 1, the nine core textbooks contained different numbers of tokens. They were left intact, since lexical text coverage was to be measured in percentage, which was not subject to an equal number of running words. Different textbooks with different amounts of text also reflect the real situation.

Procedure and Data Processing

The researcher first measured the vocabulary levels of accounting textbooks along the BNC/COCA word-frequency scale and then calculated the number of mid-frequency word families with 12+ occurrences. As reviewed in the literature, previous research on lexical coverage has

consistently used 95% (a lower criterion) or 98% (an upper criterion) as a benchmark. 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 of each 1000 word families accumulates to 95% or 98%. Meanwhile, the vocabulary level of that text can be extrapolated based upon which 1000-word-family level is the last one being added when the cumulative coverage reaches 95% or 98%.

To analyze the lexical profiling of accounting textbooks, the vocabulary analysis program AntWordProfiler (Anthony, 2014) was utilized by installing the BNC/COCA 25,000 word families as well as the three lists of proper nouns, compounds and acronyms complied by Nation (2017). After running the program, the words that were placed in the 'Words NOT Found In Base Lists' (hereafter termed as off-list) were further examined. If an off-list word was a geographical or personal name, it was added to the proper noun list.

The hyphen of a hyphenated compound was substituted by a space so that it would not be mistaken by the AntWordProfiler for an off-list word (e.g., *year-to-date*, *zero-coupon*, *write-off*, *paid-in*, *spin-off*, *work-in-process*, *after-tax*). If the constituent words of a hyphenated compound have already been included in the twenty-five 1000-word-family lists, the deletion of the hyphen can avoid their being put in the off-list. Closed compounds refer to two-word combinations without a space and may be put in the off-list as well (e.g., *breakeven*, *baseline*, *breakdown*, *cashflow*). The closed compounds in the off-list were added to the compound list.

Inevitably, acronyms occur in the accounting register. They are usually explained in context, with the full form given in parentheses or glossed in appendix (e.g., *IAS* for *International Accounting Standards*, *ROI* for *return on investment*, *EFTS* for *electronic funds transfer system*, *FIFO* for *first-in-first-out*). The acronyms that appeared in the off-list were added to the acronym list.

When calculating the lexical coverage of a text, Nation (2006) took proper nouns, compounds and acronyms into consideration, apart from the base word lists. One can recognize a proper name from its spelling with little effort. It is not difficult to infer the meaning of a closed compound from its component words if it is semantically compositional. It is also not laborious to look an acronym up for its full form in appendix or in context. As such, excluding the text coverage of proper nouns, compounds and acronyms would overestimate the vocabulary level of a text. Therefore, their text coverage was added to that of the base word lists until the cumulative percentage achieved 95% or 98%.

Word Families Used for Repetition Count

To avoid inflating the vocabulary amount of a text, 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 the learning of a new word (Nagy et al., 1989). When a new word and its family members appear in a variety of contexts, learners with inflectional and derivational knowledge would recognize them, guess

the meaning from context and strengthen knowledge from multiple encounters, which contribute to the learning of 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 Accounting Core Textbooks

Table 2 provides a snapshot of the vocabulary levels of accounting textbooks at 95% and 98% text coverage and the vocabulary distribution among the BNC/COCA 1st–25th 1000-word-family levels.

Table 2

Vocabulary Levels of Accounting Textbooks Along the BNC/COCA Scale

Word lists	Coverage of accounting textbooks as a whole	Coverage of accounting fundamentals	Coverage of intermediate accounting	Coverage of advanced accounting	Coverage of cost accounting
Proper nouns, compounds and acronyms	3.04%	3.41%	2.82%	3.00%	2.91%
1st 1000	62.93%	61.99%	63.31%	62.15%	64.11%
2nd 1000	18.49%	19.66%	18.86%	18.81%	17.9%
3rd 1000	10.32% (Cum. 94.78%)	9.62% (Cum. 94.68%)	9.88% (Cum. 94.87%)	10.54% (Cum. 94.5%)	10.66% (Cum. **95.58%**)
4th 1000	2.16% (Cum. **96.94%**)	1.86% (Cum. **96.54%**)	2.08% (Cum. **96.95%**)	2.94% (Cum. **97.44%**)	2.07% (Cum. 97.65%)
5th 1000	1.01% (Cum. 97.95%)	1.05% (Cum 97.59%)	1.05% (Cum. **98%**)	0.68% (Cum. **98.12%**)	0.94% (Cum. **98.59%**)
6th 1000	0.68% (Cum. **98.63%**)	0.91% (Cum. **98.5%**)	0.55% (Cum. 98.55%)	0.79% (Cum. 98.91%)	0.38% (Cum. 98.97%)
7th–25th 1000	1.37%	1.50%	1.45%	1.09%	1.03%
Total tokens	4,153,994	530,735	725,585	431,120	444,448

Table 2 (continued)

Word lists	Coverage of managerial accounting	Coverage of auditing	Coverage of taxation	Coverage of financial accounting	Coverage of accounting info systems
Proper nouns, compounds and acronyms	2.63%	1.81%	4.22%	3.58%	2.66%
1st 1000	64.38%	63.29%	64.7%	61.38%	64.3%
2nd 1000	17.94%	16.57%	17.37%	19.79%	16.1%
3rd 1000	10.39% (Cum. **95.34%**)	13.59% (Cum. **95.26%**)	8.35% (Cum. 94.64%)	9.64% (Cum. 94.39%)	11.15% (Cum. 94.21%)
4th 1000	1.86% (Cum. 97.2%)	2.09% (Cum. 97.35%)	2.69% (Cum. **97.33%**)	1.96% (Cum. **96.35%**)	2.19% (Cum. **96.4%**)
5th 1000	0.97% (Cum. **98.17%**)	0.99% (Cum. **98.34%**)	0.75% (Cum. **98.08%**)	1.18% (Cum. 97.53%)	1.23% (Cum. 97.63%)
6th 1000	0.6% (Cum. 98.77%)	0.52% (Cum. 98.86%)	0.68% (Cum. 98.76%)	0.87% (Cum. **98.45%**)	0.7% (Cum. **98.33%**)
7th–25th 1000	1.23%	1.14%	1.24%	1.55%	1.67%
Total tokens	361,816	472,096	463,881	342,961	381,352

Notes. 1. Coverage = % in tokens. 2. Cum. = cumulative. 3. Bolded figures indicate the level at which the cumulative coverage of that textbook has already reached 95% (the 3rd–4th 1000 level) and 98% (the 5th–6th 1000).

The accounting textbook corpus contained 4,153,994 running words (see Table 2). The 1st 1000 word families made up 62.93% of the total words of the corpus, the 2nd 1000 word families 18.49%, the 3rd 1000 word families 10.32% and so forth. As shown, the cumulative coverage of the first 3000 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 accounting textbooks involving specialist knowledge of different content areas used a condensed vocabulary, converging at the first 3000 word families.

By the 4th 1000-word-family level, 96.94% text coverage (> 95%) was attained. It was thereby extrapolated that knowledge of the most frequent 3000–4000 word families plus proper nouns, compounds and acronyms would provide 95% coverage. To put it in perspective, if EMI accounting majors have a vocabulary capacity of the first 3000–4000 word families, they would be able to read specialist textbooks rather smoothly in terms of the frequency of interruptions (e.g., consulting a dictionary or guessing unknown words). Moreover, different from unassisted extensive reading, textbooks are usually read along with teacher's instruction, which further lowers the vocabulary load.

Overall, there was not much variation in vocabulary levels among accounting textbooks across nine core courses either at 95% or 98% coverage. Cost accounting and managerial accounting as well as auditing required the least vocabulary to reach 95% coverage (see Table 2, > 95% coverage at the 3rd 1000 level). Mastery of the first 5000 word families would reach 98% coverage in these three subjects rather easily (> 98%). For the other six courses, the slightly below 95% coverage, which knowledge of the first 3,000 word families would provide, still fell within the range where good comprehension may occur (Laufer, 1989). Setting the threshold at 98% coverage, accounting fundamentals, financial accounting and accounting information systems were a little bit more vocabulary-demanding than the other subjects (specifically, entailing more than the first 5000 word families but fewer than the first 6000 to reach 98% coverage). If accounting undergraduates have a good command of the first 6000 word families, they may feel at ease without frequent dictionary lookups while reading a specialist textbook.

Compared with a vocabulary size of the first 9000 word families needed to read a diversity of texts (Nation, 2006), accounting textbooks are a lot easier because knowledge of the first 5000–6000 word families would suffice to provide 98% coverage. However, this signals that reading English-medium accounting textbooks may not result in students encountering a great variety of words and therefore may not bring about massive vocabulary expansion.

In answer to research question 1 'What are the vocabulary levels of English-medium textbooks of accounting core courses?', Table 2 demonstrates that accounting textbooks generally reached the 3rd–4th 1000-word-family level at 95% text coverage and stretched to the 5th–6th 1000 at 98% coverage. EFL students planning to study EMI accounting undergraduate programs may need to know a minimum of the first 3000–4000 word families and optimally the first 5000–6000 word families in order to perform reading tasks well in their field.

Mid-Frequency Word Families in Accounting Textbooks

The 4.1-million-token accounting textbook corpus across nine core courses contained a total of 7,458 word families from the BNC/COCA 1st to 25th 1000 with 2,881 word families from the 1st to 3rd 1000 (see Column 2 in Table 3 for the total and the 1st-3rd 1000 subtotal). The first 3000 word families were not considered, since they have often been regarded as the minimal vocabulary that senior high school graduates should master and were therefore assumed to be known. Table 3 answers research question 2 'Beyond the most frequent 3000 word families, how many mid-frequency words may EFL accounting undergraduates learn from their specialist textbooks?', demonstrating that a total of 3,300 (55%) out of the 6,000 mid-frequency word families (4th–9th 1000) would be encountered when accounting undergraduates complete their core courses. The 3,300 mid-frequency word families included words occurring fewer than 12 times and quite a few one-timers (words appearing only once) such as *agony, aesthetic, assassin, clutch, cynical, despair*, to name but a few. Only 1,274 (21%) out of

the 6,000 mid-frequency word families appeared 12+ times (see Column 2 in Table 3 for the 4th–9th 1000 subtotal in parentheses).

Table 3

Number of Word Families from the 1st to 25th 1000 in Each Textbook and the Textbooks as a Whole

Word level	Core textbooks as a whole	Accounting fundamentals	Intermediate accounting	Advanced accounting	Cost accounting
1st–3rd 1000	2881 (2484)	2425 (1293)	2485 (1465)	2155 (1245)	2424 (1421)
4th 1000	833 (466)	463 (89)	506 (136)	362 (100)	501 (120)
5th 1000	694 (293)	312 (44)	350 (71)	237 (36)	316 (55)
6th 1000	598 (214)	245 (30)	283 (42)	182 (29)	233 (30)
7th 1000	454 (145)	146 (16)	178 (26)	133 (18)	169 (17)
8th 1000	413 (90)	129 (17)	135 (10)	79 (13)	140 (7)
9th 1000	308 (66)	82 (5)	102 (9)	62 (5)	76 (8)
4th–9th 1000 subtotal	3300 (1274)	1377 (201)	1554 (294)	1055 (201)	1435 (237)
10th–25th 1000 subtotal	1277 (218)	333 (32)	335 (36)	223 (22)	295 (23)
Total	7458 (3976)	4135 (1526)	4374 (1795)	3433 (1468)	4154 (1681)

Word level	Managerial accounting	Auditing	Taxation	Financial accounting	Accounting information systems
1st–3rd 1000	2425 (1391)	2241 (1322)	2311 (1337)	2183 (1101)	2386 (1412)
4th 1000	452 (84)	397 (111)	439 (110)	362 (74)	462 (116)
5th 1000	310 (47)	245 (44)	295 (48)	226 (29)	333 (62)
6th 1000	218 (30)	183 (26)	231 (34)	158 (21)	228 (32)
7th 1000	142 (15)	133 (17)	137 (16)	96 (12)	154 (29)
8th 1000	104 (12)	88 (11)	115 (16)	78 (11)	128 (12)
9th 1000	73 (5)	54 (7)	77 (7)	57 (4)	81 (12)
4th–9th 1000 subtotal	1299 (193)	1100 (216)	1294 (231)	977 (151)	1386 (263)
10th–25th 1000 subtotal	281 (27)	232 (30)	325 (42)	183 (22)	268 (38)
Total	4005 (1611)	3573 (1568)	3930 (1610)	3343 (1274)	4040 (1713)

Note. The figures in parentheses indicate the number of word families occurring 12+ times.

Past studies on incidental learning reported that new word learning rarely happens after one or two meetings (Horst et al., 1998). As aforementioned, 12 encounters were adopted in this research as a threshold for learning to occur. Please note that the figures in parentheses in Table 3 indicate the number of word families at a particular 1000 word-family level occurring 12 times or more.

As can be seen in Table 3, when studying Accounting Fundamentals, EMI accounting students would meet 1,377 mid-frequency word families (4th–9th 1000) and 333 low-frequency word families (10th–25th 1000) including one-timers. Similarly, the other textbooks would provide opportunities for meeting mid-frequency words with intermediate accounting providing the most inclusion (1,554 word families) and financial accounting providing the least inclusion (977 word families). If considering 12+ times as an indicator of occurring often enough for learning to happen, the number of mid-frequency words in each textbook would greatly reduce to no more than 300 word families, ranging from 151 to 294 word families.

It is worth noting that a mid-frequency word appearing fewer than 12 times in a textbook may be met a few more times later when reading another textbook, which makes its total frequency exceeding 12 times. When more and more accounting textbooks are read, the number of mid-frequency word families accumulating to 12+ times would gradually increase, as 1,274 mid-frequency word families have shown in the textbooks as a whole versus 151 to 294 mid-frequency word families in a single textbook. Despite 1,274 word families, it is still far below the goal of learning most of the words from the 4th to 9th 1000 levels (6000 word families). At the 4th 1000 level, fewer than a half of 1000 word families (466 word families) occurred 12+ times, and up through the 9th 1000, only 66 word families appeared 12 times or more.

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 three million words of novels would help learners to meet most of the first 9,000 word families often enough for incidental learning to happen. However, in the present situation, even though EMI accounting students finish reading circa 4.1 million words of textbooks, they would still not meet most of the 1000 word families at least 12 times from the 4th 1000 onwards (see Column 2 in Table 3 for 466 word families at the 4th 1000 level). At the 5th 1000, there was a reduction to 694 word families from 833 at the 4th 1000 with only 293 word families appearing 12+ times, which were a lot fewer than 800 word families as per Nation's (2014) criterion in his study on novels as input. With the advance towards the upper BNC/COCA word-frequency scale, the distribution of mid-frequency and low-frequency words become more and more sporadic, with fewer and fewer words occurring 12+ times throughout the textbooks. In line with Table 2, Table 3 reconfirms that accounting textbooks used a very small vocabulary, mainly within the first 3000 word families.

This may be because, in most novels, a large number of different words are used. It may also be the diversity of themes involved in novels that result in the richness of vocabulary. In contrast to novels, the vocabulary recycling of accounting specialist textbooks is very strong, which may reduce

students' vocabulary load when they read another accounting textbook. But accounting majors may make little progress in new word learning after the first 3000 word families.

Though beyond the present research questions, a look at Table 4 shows the nature of mid-frequency words appearing in accounting textbooks. Table 4 enumerates 15 most frequently-occurring word families beyond the first 3000 in the accounting textbooks. The 15 headwords (*inventory*, *depreciate*, *equity*, etc.) as well as their family members occurred in all of the nine accounting core textbooks.

Table 4

Top15 Mid-Frequency Word Families in the Accounting Textbook Corpus

Headword	Frequency level	Frequency of occurrence	Range (across the number of textbooks)	Text coverage
inventory	5th 1000	13,940	9	0.34%
depreciate	7th 1000	6,618	9	0.16%
equity	4th 1000	6,436	9	0.15%
overhead	4th 1000	5,883	9	0.14%
dividend	4th 1000	5,499	9	0.13%
deduct	4th 1000	3,852	9	0.09%
debit	8th 1000	3,680	9	0.09%
entity	4th 1000	3,118	9	0.08%
merchandise	6th 1000	2,923	9	0.07%
consolidate	4th 1000	2,669	9	0.06%
incur	5th 1000	2,578	9	0.06%
accrue	6th 1000	2,431	9	0.06%
ledger	7th 1000	2,318	9	0.06%
receipt	4th 1000	2,251	9	0.05%
fraud	4th 1000	2,151	9	0.05%
			Total coverage	1.59%

From Table 4, it is not surprising to see that words like *inventory, depreciate, equity, overhead* and *ledger* reveal a sense of accounting content domain in connection with bookkeeping. Meanwhile, words like *dividend, debit, merchandise* and *accrue* show a financial flavour in relation to profit. However, the words *entity, consolidate, incur* and *fraud* do not show a business flavour in a strong sense and may appear in different subject matter. Apart from in the field of accounting, the lay-technical words *deduct* and *receipt* may be commonly used in daily conversation. In other words, among the 15 mid-frequency word families, 60% (9/15) of the words are closely related to accounting and finance.

The total coverage of the word families beyond the first 3000 in the accounting textbooks as a whole was 5.22% (see back Table 2 for the coverage summation of the 4th-25th 1000). Compared with 5.22% text coverage, the 15 most frequent mid-frequency word families actually accounted for 1.59% (see Table 4), which once again substantiates a small vocabulary used in accounting textbooks.

Conclusion and Implications

This lexical research was a preliminary study on an EMI accounting bachelor's degree program in an EFL setting. It had a dual purpose: to measure 1) the vocabulary levels of accounting textbooks and 2) the amount of mid-frequency vocabulary contained in accounting textbooks. Generally, accounting textbooks involving different specialist knowledge reached the 3rd–4th 1000-word-family level at 95% text coverage and extended to the 5th–6th 1000 at 98% coverage. Even though EMI accounting majors complete their required courses, continual reading of English-medium accounting 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 4.1-million-token accounting textbook corpus, only 1,274 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 EMI accounting majors' English vocabulary size would level off at the first 3000–4000 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 EMI programs may be to extensively read English newspapers, novels and all sorts of English articles.

Although this research contributes to the literature of EMI research, it has been worked narrowly within the field of accounting. 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 inspirations for future qualitative analyses of EFL learners' lexical needs and perceptions of English-medium specialist textbooks regarding reading difficulty.

Acknowledgements

The researcher wishes to express her profound gratitude to Emeritus Professor Paul Nation for the free BNC/COCA word lists, which were indispensable to this study. This research was supported by a Grant (MOST 110-2410-H-214-003) from the Ministry of Science and Technology in Taiwan.

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. (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. https://doi.org/10.1111/ijal.12246

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

Chen, J.-J. (2011). *A study of the policy performance and influence of "The Top University Development Plan" in Taiwan* [Unpublished master's thesis]. National Taiwan Normal University, Taiwan.

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/sllt.2016.6.3.5

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. https://doi.org/10.1080/13670050.2020.1727409

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

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696. https://doi.org/10125/67046

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/10125/66953

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. http://dx.doi.org/10.15639/teflinjournal.v31i2/181-203

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430. https://doi.org/10125/66973

Inbar-Lourie, O., & Donitsa-Schmidt, S. (2019). EMI Lecturers in international universities: Is a native/non-native English-speaking background relevant? *International Journal of Bilingual Education and Bilingualism, 23*(3), 301–313. https://doi.org/10.1080/13670050.2019.1652558

Kao, Y-T., & Tsou, W. (2017). EMI course assessment: A survey study of the issues. In W. Tsou, & S-M. Kao (Eds.), *English as a medium of instruction in higher education: Implementation and classroom practices in Taiwan* (pp. 183–205). Springer.

Kuteeva, M. (2020). Revisiting the 'E' in EMI: Students' perceptions of standard English, lingua franca and translingual practices. *International Journal of Bilingual Education and Bilingualism, 23*(3), 287-300. https://doi.org/10.1080/13670050.2019.1637395

Lasagabaster, D. (2018). Fostering team teaching: Mapping out a research agenda for English-Medium Instruction at university level. *Language Teaching*, *51*(3), 400–416. https://doi.org/10.1017/S0261444818000113

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). Multilingual Matters.

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. https://doi.org/10125/66648

Li, N., & Wu, J. (2018). Exploring assessment for learning practices in the EMI classroom in the context of Taiwanese higher education. *Language Education & Assessment, 1*(1), 28–44. https://doi.org/10.29140/lea.v1n1.46

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. (2020). 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. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*s, *63*(1), 59–82. http://dx.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. https://doi.org/10125/66881

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

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-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. https://doi.org/10125/66652

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513–536. http://dx.doi.org/10.1111/1467-9922.00193

Rogier, D. (2012). *The effects of English-medium instruction on language proficiency of students enrolled in higher education in the UAE* [Unpublished doctoral thesis]. University of Exeter.

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. https://doi.org/10.1017/S0272263199004039

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

Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research, 24*(4), 540–556. https://doi.org/10.1177/1362168818799371

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

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. https//doi.org/10125/66776

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics,28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245. https://doi.org/10125/66826

Yang, W. (2015). Content and language integrated learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism, 18*(4), 361–382. https://doi.org/10.1080/13670050.2014.904840

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

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541–572. http://dx.doi.org/10.3138/cmlr.57.4.541

About the Author

Wenhua Hsu is currently teaching at I-Shou University in Taiwan. She holds a PhD in the field of ESP from the University of Essex, UK and MBA from Kansas State University, USA. Her interest includes frequent academic/sub-technical and lay-technical vocabulary as well as lexical bundles in specialized fields.