Which Word List Should I Teach? Using Word Lists to Support Textbook Vocabulary Instruction

Stuart G. Towns
sgtowns@gmail.com
School of Liberal Arts
King Mongkut's University of Technology Thonburi
Thailand

Received: 01/04/2020 Revised: 29/04/2020 Accepted: 30/04/2020

Abstract

Choosing which vocabulary to teach in an English classroom can be difficult, as many variables such as textbook content and student needs factor into the decision. To aid English instructors in this task, various word lists such as frequency bands from large corpora, research-based word lists, and proficiency levels can be consulted. However, it is not clear which (if any) of these sources are appropriate for any given language classroom, or what subset of the large lists should be taught. This paper attempts to shed light on this complex issue by conducting a case study analysis of textbook materials and student needs in an English for Academic Purposes (EAP) course at a university in Thailand. Reading passages from the course textbook and the vocabulary lists which accompany the texts were compared with the words which the students chose as being important for them to learn. The results of the case study showed that for this textbook and these students, the Common European Framework of Reference (CEFR) B2 level was an appropriate place to look for additional vocabulary. Although these results are specific to one course in one context, they showcase online tools that instructors can use to analyze their own classroom content and student needs, as well as describing generalizable insights that should be considered when choosing textbooks or creating class materials elsewhere.

Keywords: English for Academic Purposes, online tools, word lists, vocabulary instruction

Introduction

It is widely regarded that vocabulary knowledge plays a crucial role in students' academic success (Nagy & Townsend, 2012), yet the amount of vocabulary that students must know in an academic university context is truly daunting. In addition, there are no clear guidelines for which vocabulary should be taught to students by their instructors at any point in their learning journey. One easy solution to this dilemma would be for instructors to rely on the vocabulary that is supplied in an English course textbook. Most modern English textbooks attempt to provide all of the necessary content that a class would need throughout the entire course, providing a "handrail on a steep climb" (van Lier, 1996, p. 208) for both teachers and students. This "handrail" includes vocabulary instruction as well, providing reading passages for the students along with vocabulary lists of words that are found in the text, and vocabulary building exercises built around these words. Instructors could simply just focus on these recommended vocabulary words. This is likely to be the most common solution,

due to the importance of textbook content in many English course syllabi. But it is not clear if these sets of words are in fact the ones that the students need to or should learn at that time.

Another solution for instructors would be to focus on the most frequent words in the language first, since common words are the ones that learners are more likely to encounter. The frequency of words is directly related to language learner proficiency as well; studies have shown that as students become more proficient in English, they are able to use less frequent words (Engber, 1995; McNamara, Crossley, and McCarthy, 2010). Therefore, if an instructor knows their students' proficiency level, then word frequency bands at or near the student's level can help determine which words the student should learn. To accomplish this, instructors can use word lists that have been organized into 1,000-word bands based on the frequency of words found in large-scale corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).

However, it is not clear that frequency alone is the best metric for determining which words a student should learn. Another solution to the question of what vocabulary to teach would be to use word lists created based on the frequency of words that appear in authentic English texts, along with other pedagogically focused criteria (Coxhead, 2000). Over the past few decades, many word lists have been created with the purpose of highlighting which words learners should study. These word lists can be organized into three main groups based on Nation (2001) — common or core words, academic words, and technical words. Two of the most popular word lists which are still in wide use are the General Services List (GSL) created by West (1953) and the Academic Word List by Coxhead (2000). The GSL contains 2,000 of the most common word families and is considered to be the core vocabulary that second language learners should learn first. It was developed from a corpus of five million words. Words to include in the list were chosen based on frequency in the corpus and usefulness for student learning (West, 1953). The AWL came much later but was designed to complement the GSL. It was created from a 3.5-million-word corpus of academic texts that a first-year university student might encounter. It contains 570 word families that were not included in the GSL (Coxhead, 2011).

More recently, Browne, Culligan, and Phillips (2013) created a New General Service List (NGSL) and a New Academic Word List (NAWL). These two lists were compiled in an effort to improve the GSL and AWL but to follow the same model of a common core word list combined with an academic word list, just like the GSL and AWL. Both the NGSL and NAWL were created from much larger corpora than the GSL and AWL. The NGSL was based on a 273-million-word subset of the 1.6-billion-word Cambridge English Corpus, while the NAWL was based on a 288-million-word corpus which included an academic corpus, two oral academic corpora, and a textbook corpus (Browne, et al., 2013).

In addition to common core and academic words, Nation (2001) also included a category of technical words, or words that are used in specific domains or industries. Many new word lists have been built around technical words in fields such as engineering (Mudraya, 2006; Ward, 2009), medicine (Wang, Liang, & Ge, 2008) and nursing (Yang, 2015), agriculture (Martínez, Beck, & Panza, 2009), business (Hsu, 2011), chemistry (Valipouri & Nassaji, 2013), environmental science (Liu & Han, 2015), and computing (Chen & Lei, 2019). Researchers in Thailand have also published word lists for Thai university studies in various fields such as business (Tangpijaikul, 2014), engineering (Watson-Todd, 2017), and multi-disciplinary EAP classes (Pojanapunya, 2019).

Another potential source of word lists comes from the Common European Framework of Reference (CEFR), which is a language-neutral guide to levels of proficiency in a language being learned. It has three groups, Basic User, Independent User, and Proficient User, split into six levels A1/A2, B1/B2, and C1/C2 respectively. The main components of the CEFR levels are communication skills, but there are also vocabulary lists available that that show which words a language learner should know at each level (What is the CEFR, 2015). Many modern English textbooks are targeted at certain CEFR levels, however, the CEFR levels of the students in a class may vary widely. In addition, each CEFR level contains thousands of words, so even students who are on the same level may have very different vocabulary knowledge.

It was reported in 2014 that Thailand's Education Ministry desired to incorporate CEFR standards into English language instruction in Thailand, both for the creation of materials as well as assessments (Intathep, 2014). This was followed in 2016 by the creation of guidelines by the Commission of Higher Education stating that university-created proficiency tests should be based on the CEFR (Commission of Higher Education, as translated and cited in Baker & Jarunthawatchai, 2017). Due to these governmental guidelines, teachers have started using CEFR in their teaching practice. A study investigating teacher perspectives on the impact of CEFR on language learning in Thailand reported that a majority of teacher participants (thirty-three Thai English as a Foreign Language (EFL) university lecturers) had used CEFR levels to help them create examinations (Kanchai, 2019). However, while many Thai teachers see the benefits in using CEFR to create proficiency tests, another study showed that Thai teachers are not using CEFR standards in creating teaching materials (Franz & Teo, 2017). Using CEFR word lists as a way to create vocabulary learning materials for students might be a way to implement Thai governmental policy for the classroom.

Even though each of these word lists prioritize words for students to learn, the lists themselves are still very large, containing hundreds or thousands of words. Due to the limitations of time in the language classroom, teachers must be further selective in choosing which words they should teach their students. One solution to this issue could be to choose "opaque words" from the lists, that is, words which have academic meanings that are different from the most common meanings of the word (Watson Todd, 2017). In addition, while word lists might be useful for an instructor in designing vocabulary instruction and analyzing the vocabulary levels of academic texts (Gardner & Davies, 2014), it is not clear which of the many word lists are the most appropriate for a particular classroom of students, or which subset of words from the chosen list should be prioritized. In the context of Thailand, research into the use of word lists focus on the evaluation of curriculum and textbooks (Poonpon, Honsa, & Cowan, 2001; Wan-a-rom, 2012), national English exams (Sukying, 2010), graded readers (Wan-a-rom, 2008), and the proficiency of students (Wiriyakaruyn, 2018). But there does not seem to be any research into how teachers in Thailand might use these word lists to select the most appropriate words.

There does not seem to be a lot of guidance for instructors between the easy but limited "handrail" of relying on the listed vocabulary words in textbooks, and the overwhelming number of words on various word lists that have been created over the years. For instructors who are required to use textbooks, it is not clear how to navigate this space in order to provide the most efficient and effective learning environment for their students. Therefore, this study is an attempt to use the framework of word lists to help instructors supplement the readings in their existing textbooks with tailored vocabulary word lists for their students. The study is

presented as a case study into the materials and student needs of a specific first-year university English for Academic Purposes course in Bangkok, Thailand. It uses online word list analysis tools to compare the vocabulary found in textbook reading passages, the words that the textbook authors have chosen for the students, and words chosen as problematic by the students themselves, with the purpose being to provide guidance for the instructor in selecting vocabulary to focus on in the course.

Therefore, the research question for this study is: How might instructors use word list tools to help augment the textbook by selecting additional vocabulary for their students to learn?

Methodology

This study was conducted with students from a freshman-level university English course in Thailand. The course is entitled "Academic Reading and Writing" and uses the *Pathways: Reading Writing, and Critical Thinking 3* textbook from National Geographic Learning. This textbook is intended for students whose English proficiency is at a CEFR B2 level and is organized in a similar manner to most textbooks on the market. There are ten units in the textbook, with each unit including two reading passages of approximately 700-800 words, with ten vocabulary words highlighted in each of the readings. The remainder of each unit contains exercises to practice vocabulary, reading comprehension, and writing.

During this study, some classroom time was spent on reading practice for the students. On these days, the instructor first asked the students to complete the vocabulary exercises from the textbook which focused on the ten highlighted words in the readings. Then, the students read the passage in preparation for working on the reading comprehension exercises. The teacher encouraged students to mark unfamiliar words in the reading passages and to look up the word definitions to help them understand the text. At the end of the semester, the words which students marked in their textbooks as being problematic were collected. Six students had marked at least 20 words across the four texts, with the remainder of the students in the class not using this method of marking words consistently. Therefore, the words chosen by these six students comprise the data for this study.

Table 1 provides further information about the data in this study. For the first of the four texts, only the first half of the text was assigned in the class and therefore only consisted of 492 words. However, all ten vocabulary words highlighted by the textbook author were in the first half of the text. The remaining three reading passages were all read in full by the class, and had text lengths of 761, 875, and 787 words respectively. Each of the four texts contained ten highlighted words. However, in Text 3, two of the words were actually phrasal verbs and therefore not included in this study because the automated analysis tools work with single words only.

All of the words that the students chose were collected and sorted by the number of students who chose each word. Table 2 contains how often words were chosen by students. Of the 159 total words chosen by students, 74 words were chosen by just one student, 51 words were chosen by two students, 19 words were chosen by three students, ten words were chosen by four students, and five words were chosen by five students. No words were chosen by all six students. As can be seen in Table 2, 34 words were chosen by at least three of the six students.

Table 1: Data for this study: Text lengths and number of chosen words

Text #	# total words in the text	# Words chosen by author	% of words chosen by author	# words chosen by students	% of words chosen by students
1	492	10	2.0%	34	6.9%
2	761	10	1.3%	53	7.0%
3	875	8	0.9%	35	4.0%
4	787	10	1.3%	37	4.7%
TOTAL	2915	38	1.3%	159	5.5%

Table 2: Number of words chosen for this study

Number of students	Number of words chosen	Total Words grouped by analyzed or not analyzed
1	74	125 words not analyzed in this study
2	51	
3	19	34 words analyzed in this study
4	10	
5	5	
6	0	
Total	159	

This study will focus on these 34 words chosen by students and will ignore the 125 infrequently chosen words. There are several reasons for this decision. First, the list of 34 student chosen words is comparable in size to the 38 total words in the vocabulary lists in the textbook, making comparisons clearer. Secondly, there is a limit to how many words could be introduced to the students in one lesson. If the instructor elected to add these most common chosen words to the explicit vocabulary instruction for the class, then the vocabulary requirements for each reading passage would be almost doubled on average. This seems like an acceptable level considering classroom time constraints.

This study will compare the words marked by students as being problematic (hereafter referred to as "student-chosen words") and the words given as vocabulary to learn in the textbook (hereafter referred to as "author-chosen words"). The list of 38 author-chosen and 34 student-chosen words can be seen in Table 3. This table also includes how many students chose each word.

Table 3: Author-chosen words and student-chosen words

Words chosen by the textbook authors	analysis, archaeologist, attempt, basically, combination, committing, consumption, detectives, determined, disciplining, enhance, establish, examined, extended, families, gender, generally, heal, identities, increasingly, industrial, intense, majority, mentioned, moreover, mysteries, observed, obtained, phenomenon, previously, proved, replace, revealed, safety, samples, suspect, unclear, varied

Words chosen by 3 of the 6 students	appropriate, bachelor, bulldozer, compromised, conflict, conviction, cooperate, dynasty, evidence, fetuses, fracturing, graze, hierarchy, necrosis, offspring, scholar, victim, violent, wounds
Words chosen by 4 of the 6 students	admitted, architect, arrested, committing, forensic, mediators, primates, robbery, scarce, urban
Words chosen by 5 of the 6 students	flock, leeches, orphan, perceived, suspect

Analysis Tools and Methods

The purpose of this study is to evaluate the usefulness of various online word list analysis tools in aiding the instructor in selecting appropriate vocabulary for the students. To provide quantitative data for this analysis, the author-chosen words and the student-chosen words were compared with various corpora, word lists, and proficiency levels. The online *Complet Web VocabProfiler* (Cobb, n.d.) was used to make comparisons with the following:

- First twenty-five 1,000-word frequency bands of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).
- General Service List (GSL) and Academic Word List (AWL)
- New General Service List (NGSL) and New Academic Word List (NAWL)

The final word list analysis was a comparison of the author-chosen and student-chosen word lists with the bands of the Common European Framework of Reference for Languages (CEFR). This analysis was conducted using the English Vocabulary Profile Online tool from Cambridge University Press (English, 2015).

Word lists such as the GSL and AWL are generally lemmatized. That is, they do not differentiate between different forms of the same root word. However, in the real-world classroom environment, students will likely only see one form of a lemma in a reading passage. Therefore, in order to provide more evidence to help the instructor choose the best words and most useful forms to present to the students, a part of speech analysis was also undertaken to compare the three sets of data (textbook reading passages, the textbook authors-chosen word lists, and the student-chosen word lists). In this analysis, the four textbook reading passages were analyzed using the online CLAWS part-of-speech tagger for English from the University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University (Garside, 1987). In addition to the overall scores for the full texts, the parts of speech of the 38 author-chosen and 34 student-chosen words for each text were collected, tabulated, and compared.

Results

As described in the previous section, there were two main types of analysis that were conducted. The first was to use online tools to compare the data with existing word lists. The second analysis was a part of speech analysis of the data. The following sections describe the results of these analyses.

Comparisons with Available Word Lists

In order to understand the appropriateness of word lists for the EAP class in this study, the textbook reading passages, the author-chosen words, and the student-chosen words were compared to various word lists using the online tools *Compleat Web VocabProfiler* from the LexTutor website and *English Vocabulary Profile Online*. In the case of the *Compleat Web VocabProfiler,* there are several different comparison tools that are available. For this study, the textbook reading passages were compared to the following existing word lists: 1) the British National Corpus and the Corpus of Contemporary American English (BNC/COCA), 2) the General Service List and Academic Word List (GSL/AWL), and 3) the New General Service List and New Academic Word List (NGSL/NAWL).

Table 4 contains the results of using the *Compleat Web VocabProfiler* tool to compare the data with the BNC and COCA word lists containing the first 25,000-word bands by frequency in those corpora. As can be seen in Table 4, the reading passages include words from all of the bands K-1 to K-12, one word from K-15, and two words from K-19. That is, almost all of the words in the texts were in the most common 12,000 English words. All of the author-chosen words, however, are in Bands K-1 to K-3 (the first 3,000 most frequent English words), with the majority being in the K-2 band. For the student-chosen words, the spread is much wider and almost half (47%) of the student-chosen words were in the 4,000-11,000 most common words. None of the students choose the rarest five words in the texts that were in the K-12 to K-24 bands.

Table 4: Comparison of word lists with BNC and COCA

	Reading Passages		**Author-Chosen Words**		**Student-Chosen Words**	
Freq. Level	**Types**	**Coverage**	**Types**	**Coverage**	**Types**	**Coverage**
K-1	578	64%	6	16%	1	3%
K-2	182	84%	20	68%	8	26%
K-3	84	93%	12	100%	9	53%
K-4	14	94%			6	71%
K-5	12	96%			2	76%
K-6	6	96%			3	85%
K-7	6	97%			2	91%
K-8	3	97%			1	94%
K-9	2	98%				94%
K-10	4	98%			1	97%
K-11	2	98%			1	100%
K-12	1	98%				
K-13		98%				
K-14		98%				
K-15	1	98%				
K-16		98%				
K-17		98%				
K-18		98%				
K-19	2	99%				

	Reading Passages		Author-Chosen Words		Student-Chosen Words	
Freq. Level	**Types**	**Coverage**	**Types**	**Coverage**	**Types**	**Coverage**
K-21		99%				
K-22		99%				
K-23		99%				
K-24	1	99%				
K-25		99%				

However, the frequency of the words chosen by students varied across texts. Texts 1, 3, and 4 had similar results, but Text 2 was an outlier. The frequencies for student-chosen words across the four texts are shown in Table 5. As can be seen from the table, the majority of words from Text 2 are in the K-1 to K-2 bands. Texts 1, 3, and 4 have a much wider spread.

Table 5: Frequency bands of student-chosen words across four reading passages

Freq. Level	Text 1	Text 2	Text 3	Text 4
K-1		1		
K-2		8		
K-3	2	1	2	4
K-4	4		1	1
K-5	2			
K-6	1	1	1	
K-7	1			1
K-8			1	
K-9				
K-10			1	
K-11		1		

An additional two analyses were done using the *Compleat Web VocabProfiler* tool: comparisons with the GSL/AWL word lists and with the NGSL/NAWL word lists. The results of these three analyses can be seen in Tables 6 and 7 respectively

The *Compleat Web VocabProfiler* tool for GSL/AWL separates the GSL into two groups each containing 1,000 words and also includes the AWL as the third word list. As seen in Table 6, the textbook reading passages have a very large percentage of the 1000 most common GSL words (GSL-1) at 65% of the total word types. The textbook authors seem to have selected a wider range of words, with GSL-1, GSL-2 and AWL containing 34%, 21% and 39% respectively. Students, on the other hand, didn't seem to have much trouble with GSL or AWL words, as most of their words were in the Off-list group (56%).

Table 6: Comparison of word lists with GSL and AWL

GSL / AWL	All texts		Author-Chosen		Student-Chosen	
	Types	% of total	Types	% of total	Types	% of total
GSL-1	589	65%	13	34%	4	12%
GSL-2	103	11%	8	21%	4	12%
AWL	101	11%	15	39%	7	21%
Off-list	116	13%	2	5%	19	56%

For the comparisons with NGSL and NAWL, shown in Table 7, the *Compleat Web VocabProfiler* tool divides the NGSL into three 1,000-word groups, and then includes a word list for the NAWL. As can be seen in Table 7, similarly to Table 6, the number of words in the most common 1,000-word NGSL list (NGSL-1) for the textbook reading passages is very high at 65%. The highest group for the author-chosen text is GNSL-2 at 47%, while the student-chosen word list is more evenly spread out in all categories, with the lowest being NGSL-1 at 9%. The largest group for the student-chosen text was again the Off-list group at 41%.

Table 7: Comparison of word lists with NGSL and NAWL

NGSL / NAWL	All texts		Author-Chosen		Student-Chosen	
	Types	% of total	Types	% of total	Types	% of total
NGSL-1	591	65%	11	29%	3	9%
NGSL-2	129	14%	18	47%	8	24%
NGSL-3	59	6%	4	11%	5	15%
NAWL	28	3%	2	5%	4	12%
Off-list	102	11%	3	8%	14	41%

The final analysis was performed using the *English Vocabulary Profile Online* tool, which contains word lists for all six CEFR levels A1, A2, B1, B2, C1, and C2. As seen in Table 8, the largest groups for all textbook reading passages is A1 and A2, for a total of 59%. This is not a surprising result, as the most common words are the words which are the most frequent. However, it is interesting to note that neither the authors nor the students chose any words from the A1 or A2 levels. The author-chosen words form a bell curve with the highest

point at B2 (58%) and just a few on each side of the peak at B1 and C1. The authors did not choose any C2 or off-list words. The student-chosen word list had a much wider range than the author-chosen list, but their highest point was also at B2 with 35%, but they reached this high point of 35% in the off-list category.

Table 8: Comparison of word lists with all CEFR levels

CEFR Level	All texts		Author-Chosen		Student-Chosen	
	Types	% of total	Types	% of total	Types	% of total
A1	656	43%	0	0%	0	0%
A2	245	16%	0	0%	0	0%
B1	202	13%	10	26%	3	9%
B2	166	11%	22	58%	12	35%
C1	33	2%	6	16%	3	9%
C2	38	2%	0	0%	4	12%
Off-list	186	12%	0	0%	12	35%

Comparisons to Parts of Speech

In order to provide more data to the instructor as to which vocabulary words should be chosen from the readings to the students, a part of speech analysis was conducted on the four textbook reading passages. The parts of speech of the 38 author-chosen words and the 34 student-chosen words were tabulated. The representation of each part of speech was then calculated as a percentage. For example, the vocabulary words chosen by the textbook authors included ten singular common nouns, or 26% of the total number of author-chosen words. In comparison, students chose 14 singular common nouns, which amounts to 41% of the total number of student-chosen words. The full results of the frequency of parts of speech, percentage of total, and the differences between the percentages for author-chosen and student-chosen words are shown in Table 9.

In this study, 12 different parts of speech were found in the text by using Part of Speech tagging. The twelve groups were: prepositions, adjectives, three types of nouns, adverbs, and six verb tenses. Most of the parts of speech between the author-chosen and student-chosen word lists were similar. For example, 13% of the words chosen by the authors were adjectives, which 12% of the student-chosen words were adjectives. The biggest differences between author-chosen words and student chosen words appeared in singular common nouns (26% and 41% respectively), adverbs (13% and 0%), and past tense verbs (13% and 0%).

As can be seen in Table 9, some parts of speech were never chosen by textbook authors or students, such as general prepositions (II) and -s form of lexical verbs (VVZ). Other parts of speech showed little to no differences between the chosen word lists, such as adjectives (JJ), common nouns (NN), -ing participle of lexical verbs (VVG), base forms of lexical verbs (VV0), infinitives (VVI), and past participles of lexical verbs (VVN).

Table 9: Part of speech analysis

	Author-Chosen	%	Student-Chosen	%	Difference
General preposition (II)	0	0%	0	0%	0%
General adjective (JJ)	5	13%	4	12%	-1%
Common noun, neutral for number (NN)	0	0%	1	3%	3%
Singular common noun (NN1)	10	26%	14	41%	15%
Plural common noun (NN2)	5	13%	6	18%	4%
General adverb (RR)	5	13%	0	0%	-13%
Base form of lexical verb (VV0)	1	3%	2	6%	3%
Past tense of lexical verb (VVD)	5	13%	0	0%	-13%
-ing participle of lexical verb (VVG)	2	5%	2	6%	1%
Infinitive (VVI)	3	8%	1	3%	-5%
Past participle of lexical verb (VVN)	2	5%	4	12%	7%
s form of lexical verb (VVZ)	0	0%	0	0%	0%

Since the largest differences were found in nouns, verbs, and adverbs, Table 10 shows the combination of words that might be found in noun phrases (nouns and adjectives) and verb phrases (verbs and adverbs). It can be seen that there is quite a large difference between these two sets. In general, students were more likely to choose words that might appear in a noun phrase, while the authors were more likely to choose words that would appear in a verb phrase.

Table 10: Noun and verb combinations

	Author-Chosen	%	Student-Chosen	%	Difference
All Nouns	15	39%	21	62%	-22%
All Verbs	13	34%	9	26%	8%
All Nouns + Adjectives	20	53%	25	74%	-21%
All Verbs + Adverbs	18	47%	9	26%	21%

Discussion

Learning vocabulary in a foreign language is a daunting task, but there are several ways to narrow down the immense number of words for a learner to focus on. One approach that an instructor could take would be just to follow what is given in the course textbook. This is an easy solution, but it is not clear if it is the most appropriate to a specific class context. Instructors could also look to frequency lists from large corpora such as the BNC or COCA and try to focus their teaching on the most frequent words in English near the student's proficiency level. Or, instructors could turn to one of the many research-based word lists that have been created over the years. A final solution might be to follow the guidelines set out by the Common European Framework of Reference (CEFR).

But with so many potential solutions, it is difficult to know which would be the most suitable to support a specific textbook with a specific set of students. This research, therefore, was conducted to help instructors solve this puzzle. It is presented as a case study, and as with all case studies, the quantitative results are limited in that they are not generalizable to other contexts. That is, the results presented in the previous section are not generalizable to other instructors, other textbooks, or other classes. Every textbook is different, and every classroom has different students with different vocabulary knowledge. However, the methodology and analysis tools used in this study provided this specific instructor with evidence to better understand whether or not this textbook was a suitable match for these students. From the results shown above, this instructor was able to answer questions such as 1) whether or not the author-chosen vocabulary are enough in terms of quantity, 2) which word list would be the most useful in helping the students select words to learn, and whether there any other insights about student needs, and 3) what vocabulary would be most appropriate to learn in the specific context. The answers to each of these questions for this particular case study will be addressed in turn.

Is the author-chosen vocabulary enough in terms of quantity?

The first data point to consider is the coverage of provided vocabulary instruction as a percentage of the assigned reading passages, compared to the student's needs. If the number of words that the students have marked as problematic is high, then this might tell a teacher that the texts are too difficult, or that more vocabulary practice is needed, or else that the texts need to be shorter to allow students to focus on fewer words in one reading. In the case of this study, from the data gathered and shown in Table 1, it appears that the textbook authors did not select enough key vocabulary for the students in this course. For most texts, around 1% of the text was chosen as important for learning by the textbook authors. These students, however, agreed that there was an additional 4% - 7% of the words in the text that impeded their reading comprehension. This might imply that this particular set of students might need to study at least 5% - 8% percent of the words in the types of texts that are in this textbook. So, for an 800-word text from this specific textbook, the vocabulary list should be around 50 words for these students. This may be impractical due to time constraints in a classroom lesson (as well as constraints on the students' retention) so if the students are struggling with many words, then perhaps the texts should be shorter.

Which word list would be the most useful in helping these students to select words to learn?

Comparing the student's problematic word lists to available research-based word lists can also help the instructor gain understanding into what words are important for the students to learn. This study looked specifically at word lists from BNC/COCA bands, GSL/AWL, NGSL/AWL word lists, and CEFR levels, but there are many other word lists available, including many that are domain specific. But since this was a general first-year university course, the technical word lists were not considered for this study. If an instructor is teaching an ESP course in a specific field, however, then the technical word lists could be used in similar ways to the approaches in this study.

For the comparison to the BNC and COCA frequency bands, most of the words (53%) selected by students were in the K-1 to K-3 levels. This is not a surprising result, as the most common words will, by definition, appear in texts more often. The texts themselves have a

93% coverage through the K-3 band. In other words, most words in the text are in these first three bands, so it is perhaps understandable that students would pick more of them. The textbook authors have also chosen all of their words from these first three frequency bands, providing good coverage already. Therefore, since the other 47% of words that the students chose appear in the K-4 to K-11 bands, these less common bands might be a fruitful place to look for useful words to explicitly teach.

Looking at the comparisons of the student-chosen words to the GSL/AWL and NGSL/NAWL word lists, it is perhaps surprising that, for this English for Academic Purposes class, these general and academic word lists were not useful for selecting additional vocabulary for these students to learn, as most of the student-chosen words were off-list in both comparisons. This means that the students were not very interested in learning words from these word lists. The CEFR level analysis, however, did provide useful results, as the students showed great interest in the B2 level of the CEFR. This indicates that students might actually be at the B2 level as expected. This also implies that the CEFR B2 level is a good place to look for additional useful vocabulary for the students to learn.

Can the analysis provide any additional insights?

Going back to the analysis with the BNC/COCA frequency bands, it was shown in Table 5 that Text 2 had a different vocabulary profile than Texts 1, 3, and 4. Specifically, the words that the students chose in Text 2 were more common (that is, easier words) than the words chosen in the other texts. Looking at these words shows that most of them are all related to the same topic — *arrested*, *suspect*, *admitted*, *conviction*, *victim*, *forensic*, *robbed*, *wounds*, and *violent* are all about police investigations of crimes, which is also the topic of the text. It is interesting to realize that these students probably never had a reason to learn words about crime scenes in their previous English courses. Therefore, this analysis could provide insights into where pre-reading vocabulary exercises and games on the reading topics might help prepare the students to better comprehend the assigned reading passage.

Part of speech taggers might also be able to play a role in streamlining the vocabulary that is presented to the students to learn. An analysis such as the one conducted in this study can help to show which parts of speech are most common in the textbook reading passages, which are most likely to be highlighted by the textbook authors, and which are most likely to be problematic for the students. If the parts of speech that the students are struggling with is different than what the textbook provides, then this might point to areas where instructors can bring in additional exercises, and teaching. For example, in this particular study, there was a big difference in the parts of speech selected by the students compared to those selected by the textbook authors. In the student-selected list, 62% of the words were nouns while only 39% were nouns in the author-chosen list.

The analyses in this study can also show whether or not the textbook and the authors' chosen words are a good match for the students. In this particular case study, the CEFR levels appeared to be the most useful analysis tool as mentioned above. The textbook is rated CEFR B2, and B2 is also the category with the highest number of author-chosen words and student-chosen words. This would imply that the B2 level textbook focuses on the B2 level as advertised and that the students have felt that B2 words were important for them to learn (even though they had no idea that they were B2 words).

Conclusion

In classrooms around the world, English instructors are teaching EAP courses. One difficult task for these instructors is to determine which vocabulary words they should teach their students. Many of these instructors are using a textbook which contains reading passages along with short lists of selected vocabulary for each passage, but it is difficult for these teachers to know if the texts and vocabulary are enough, or if they are at the right proficiency level for their students. If the instructors do want to teach additional words, what words should they teach? One solution would be to employ word lists to help prioritize the words which should be taught. However, there are multitudes of potential word lists to choose from, such as general and academic word lists, technical word lists, opaque word lists, and CEFR-based word lists. The methodology used in this study was an attempt to shed light on this complex issue. This study compared textbook author-chosen words, student-chosen words, and various word list tools to narrow down the most useful words for students in their courses. The results from this study showed that for this case study, the CEFR B2 textbook for this course was appropriate for these students but learning additional CEFR B2 vocabulary would be help them understand the textbooks reading passages better. This was especially true if the additional B2 words were nouns, as well as including words about police investigations. Although these results are not generalizable to other contexts, the methodology used in this study can provide an analytical approach that other instructors can use in their own classrooms. It is hoped that this case study will provide some ideas for instructors about how they could conduct similar analyses to help guide their context-specific vocabulary teaching practices.

References

Baker, W., & Jarunthawatchai, W. (2017). English language policy in Thailand. *European Journal of Language Policy*, 9(1), 27-44.

Browne, C., Culligan, B., & Phillips, J. (2013). The new general service list: A core vocabulary for EFL students and teachers. *JALT's The Language Teacher*, *34*(7), 13-15.

Chen, H., & Lei, G. (2019). Developing a technical words list for research articles in computer science discipline. *English Language Teaching*, *12*(10), 131-141.

Cobb, T. (n.d.). *Compleat Web VocabProfiler v.2.1*. Accessed 5 Jan 2020 from https://www.lextutor.ca/vp/comp/.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-238.

Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly*, *45*(2), 355-362.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4*(2), 139-155.

English Vocabulary Profile Online. (2015). English Profile: The CEFR for English. Accessed 5 Jan 2020 from http://www.englishprofile.org/wordlists/evp.

Franz, J., & Teo, A. (2018). 'A2 is Normal' – Thai Secondary School English Teachers' Encounters with the CEFR. *RELC Journal*, 49(3), 322-338.

Gardner, D., & Davies, M. (2014). A new vocabulary word list. *Applied Linguistics*, *35*, 305-327.

Garside, R. (1987). The CLAWS Word-tagging System. In R. Garside, G. Leech and G. Sampson (Eds.), *The Computational Analysis of English: A Corpus-based Approach* (pp. 30-41). London: Longman.

Hsu, W. (2011). A business word list for prospective EFL business postgraduates. *Asian ESP Journal*, *7*(4), 63-99.

Intathep, L. (2014) European standards set for language. *The Bangkok Post*. 10 March 2014.
Kanchai, T. (2019). Thai EFL university lecturers' viewpoints towards impacts of the CEFR on their English language curricula and teaching practice. *NIDA Journal of Language and Communication*, 24(35), 23-47.
Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, *39*, 1-11.
Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*(3), 183-198.
McNamara, D. S., Crossley, S.A., & McCarthy, P.M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57-86.
Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*, 235-256.
Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91-108.
Pojanapunya, P. (2019). Tailoring academic words to multidisciplinary EAP classes. *TESOL Journal*, *10*(3), 1-16.
Poonpon, K., Honsa Jr, S., & Cowan, R. A. (2001). The teaching of academic vocabulary to science students at Thai universities. *Studies in Languages and Language Teaching*, 51-63.
Sukying, A. (2010). An analysis of headwords in ELT coursebooks and Ordinary National Educational Tests in Thailand. *LSCAC 2010 Proceedings*, 188.
Tangpijaikul, M. (2014). Preparing business vocabulary for the ESP classroom. *RELC Journal*, 45(1), 51-65.
Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, *12*, 248-263.
Van Lier, L. (1996). *Interaction in the Language Curriculum: Awareness, Autonomy and Authenticity*. London: Routledge.
Wan-a-rom, U. (2008). Comparing the vocabulary of different graded-reading schemes. *Reading in a Foreign Language*, 20(1), 43-69.
Wan-a-rom, U. (2012). Lexical evaluation of teacher-made coursebooks: Thai case studies of Foundation English Courses at Tertiary Level. *English Language Teaching*, 5(8), 146-156.
Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*(4), 442-458.
Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170-182.
Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31-39.
West, M. (1953). *A General Service List of English Words*. London: Longman.
What is the CEFR? (2015). *English Profile: The CEFR for English*. Accessed 5 Jan 2020 from http://englishprofile.org/the-cefr
Wiriyakarun, P. (2018). Examining Thai EFL learners' knowledge of academic English vocabulary. *Liberal Arts Journal*. 1(1), 119-132.
Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*, 27-38.

About the Author

Stuart G. Towns is an instructor at King Mongkut's University of Technology Thonburi (KMUTT) where he also received his PhD in Applied Linguistics. His PhD thesis investigated the role of linguistic features in writing quality. Other interests include the use of computers for teaching, learning, and researching language.