https://doi.org/10.15625/2615-8965/22620114

Vocabulary Demands and Learning Opportunities in two EAP Themebooks: A Lexical Profiling Study

Quynh Le[1]✉, Averil Coxhead[2]

[1] quynhltn@ueh.edu.vn
School of Foreign Languages,
University of Economics, Ho Chi Minh City
279 Nguyen Tri Phuong street, Dien Hong ward,
Ho Chi Minh City,
(Vietnam)

[2] averil.coxhead@vuw.ac.nz
School of Linguistics and Applied Language Studies,
Faculty of Humanities and Social Sciences,
Victoria University of Wellington
21D Kelburn Parade, Wellington, 6012,
(New Zealand)

✉ Corresponding author

ABSTRACT: *In this study, we conducted two lexical profiling analyses to examine vocabulary demands and exposure to academic words in two English for Academic Purposes (EAP) in-house materials called Intermediate and Academic themebooks. Using Nation's (2017) BNC/COCA word family lists and AntWordProfiler (Anthony, 2021), the first analysis measured lexical coverage, or the number of word families needed to know 95% to 98% of words in the text. Research suggests that readers who know 95% and preferably 98% of the words in a text can comprehend it effectively. The first analysis reveals that the themebooks require knowledge of 3,000-word families for 95% coverage, and up to 5,000 - 6,000 word families for 98% coverage. The second analysis identified the repetition and distribution of pre-selected sets of 105 target academic words in each book. Data indicate that the Intermediate themebook offers better opportunities for consolidating and enriching vocabulary knowledge, with nearly 60% of target words from Coxhead's (2000) Academic Word List occurring ten times or more, and over half spread across three or more units. Lexical profiling can therefore guide teachers and textbook writers to adjust and supplement materials so that students read texts at a suitable level, and encounter target vocabulary frequently and across varied contexts.*

KEYWORDS: **Lexical profiling, vocabulary demands, academic vocabulary, EAP, repetition**

➔ Received 15/12/2025 ➔ Revised manuscript received 10/3/2026 ➔ Published 16/9/2026.

1. Introduction

In many English for Academic Purposes (EAP) programmes, especially those developed for a specific institution's educational context, in-house materials are often written to align with local curricular goals and student needs. In one such programme, two in-house theme-based EAP textbooks (themebooks henceforth) are currently used as the primary instructional materials to develop students' academic skills. However, it remained unclear whether these textbooks facilitate learning in ways that suit learners' needs. In this context, a textbook evaluation would be a necessary step to assess the suitability and make improvements accordingly.

When evaluating such EAP materials, one important consideration is the lexical demand of the input text. This characteristic of the textbooks determines the difficulty level of text, which has an important impact on vocabulary learning and on language learning in general. If a text contains many unfamiliar words, the task will likely shift from reading comprehension to a focus on word learning (Nation, 2022). Meanwhile, a text with all words familiar to learners would limit opportunities to increase their vocabulary size.

Another key consideration for textbook evaluation is the amount of academic vocabulary in the input text. EAP students need frequent meetings with useful vocabulary to develop their word knowledge (Nation, 2022). Besides technical words related to their respective major, general academic words are particularly essential for EAP students to learn. That is because these items (e.g., analyse, establish, link) are widely used across academic registers and disciplines, making them essential for EAP learners regardless of their major (Coxhead, 2000). The two EAP themebooks should therefore integrate this general academic vocabulary.

Despite the central role of the EAP themebooks in supporting vocabulary learning, the lexical features of the books have not yet been systematically examined. This leads to challenges

for curriculum design and practice, as EAP teachers lack empirical evidence to determine whether the in-house EAP themebooks are suitable to develop students' lexical knowledge. Although previous textbook studies (e.g., Lu & Dang, 2022; Yang & Coxhead, 2022) have examined the lexical demands and academic vocabulary of commercial EAP textbooks, little attention has been paid to in-house EAP materials designed for a specific educational context. Evaluating such materials can provide evidence not only for local curriculum refinement but also for how institution-specific EAP resources support word learning. To address this issue, the present study evaluates the vocabulary load and the presence of pre-selected academic words of these two EAP themebooks.

Specifically, the study applies lexical profiling to estimate the vocabulary size required for 95% and 98% text coverage – commonly used thresholds for assisted and unassisted comprehension (Nation, 2017) – and analyses the repetition and distribution of selected academic words taken from Coxhead's (2000) Academic Word List (AWL). The study addresses the following research questions (RQs):

RQ1: What is the vocabulary load of the text in the selected EAP themebooks?

RQ2: What are the frequency and distribution patterns of the selected AWL word families in the themebooks?

2. Literature review

2.1. Lexical coverage and reading comprehension

Second language research has long established a strong link between reading comprehension and vocabulary size of language learners (Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011; van Zeeland & Schmitt, 2013). An indicator of this relationship is lexical coverage, defined as "the percentage of running words in the text known by the readers" (Nation, 2006, p. 61). Research suggests that 95% text coverage allows for assisted comprehension, while 98% coverage enables largely unassisted understanding (Nation, 2022). These two thresholds are widely applied in evaluating the suitability of texts for language learners and are therefore adopted in the present study to assess whether the selected instructional materials provide appropriate difficulty levels for learners.

2.2. Vocabulary load studies

Corpus-based studies have examined the vocabulary load of texts to determine how many word families learners need to know to achieve 95% to 98% lexical coverage (e.g., Coxhead et al., 2017; Dang & Webb, 2014; Hsu, 2011; Laufer & Ravenhorst-Kalovski, 2010). Such research provides a useful basis for evaluating whether learning materials are accessible for intended learners.

Studies of EAP textbooks have often found a mismatch between the lexical demands of materials and target EAP students' vocabulary knowledge. For example, Lu and Dang (2022) analysed textbooks used in an EAP programme in China and revealed that a knowledge of 4,000 word families was needed to reach 95% coverage of the materials, whereas nearly 80% of the 537 learner participants only had a vocabulary of the most frequent 1,000 words or less, according to their scores of Webb et al.'s (2017) Updated Vocabulary Levels Test. Similarly, Yang and Coxhead (2022) reported that between 3,000 and 6,000 word families were required for 95-98% coverage of a widely used textbook series, although their study relied on estimates of learner vocabulary size previously reported by Sun and Dang (2020).

Importantly, these two studies focused on commercial textbooks designed for heterogenous learner populations, where assumptions must be made about learners' proficiency. In contrast, in-house materials are typically written for more specific groups of learners within a particular institutional context, so information about their proficiency and learning needs may be available. So far, relatively few studies have examined the vocabulary demand of in-house EAP textbooks even though this kind of materials has been designed and used in many institutions. This gap is particularly significant in curriculum design, as in-house materials are expected to be tailored to specific learners' ability and needs, and should be subject to constant evaluation and adaptation. The

present study addresses this gap by analysing the vocabulary load of two in-house EAP themebooks used by a known group of learners.

2.3. Academic vocabulary in EAP

Academic vocabulary plays an essential role in EAP instruction, particularly general academic words that frequently appear across multiple disciplines (Coxhead, 2016). Because such vocabulary is not commonly used in non-academic language (Coxhead, 2000), learners may have limited opportunities for incidental learning outside academic settings. As a result, instructional materials are often expected to provide exposure to these words.

One way to evaluate this function of materials is by examining the frequency and distribution of academic vocabulary in texts. Previous research has shown that academic word lists such as the Academic Word List (AWL) account for up to 10% of academic discourse (Coxhead, 2000), indicating the significant relevance of items on this list for EAP learners. However, there remains scant research into the extent to which EAP textbooks, especially locally developed ones, provide repeated and spaced encounters with such vocabulary. In this context, the present study examined repetition and distribution of target academic words in the selected EAP themebooks.

2.4. Repetition in vocabulary learning

Repetition, which often occurs when learners see a word repeatedly, is a well-established learning principle (Nation, 2024). Research consistently shows that repetition generally strengthens word retention. Many studies also suggest that spaced repetition (there are intervals between encounters) is often favourable to massed repetition (e.g., cramming) although there is inconclusive evidence about the effects of different spacing conditions, such as equal versus increased intervals, or the lag effect (short vs. long spacing).

Combining the spaced repetition principle with other favourable learning conditions, such as meeting words in new contexts, magnifies the positive impact on learning. Every time learners meet a word in original text, their attention to the word likely increases, and new information about the word (e.g., collocations, contextual meanings, grammar) is added to their vocabulary knowledge. In light of these research findings, an EAP textbook should feature multiple occurrences of target words and ideally distribute them across lessons.

2.5. Textbook studies on word repetition

Previous textbook-based studies have consistently shown that textbooks fell short of providing sufficient opportunities for repeated encounters with target vocabulary. For example, Matsuoka and Hirsh (2010) investigated frequency counts of the second 1,000 most frequent words, the AWL words and words of lower frequency levels in one textbook. Their analysis revealed that of the 271 words from the AWL (Coxhead, 2000) that appeared in the text, only 16% occurred five times or more, while over 40% occurred only once. This finding suggested a justified need for direct learning of academic vocabulary when using the examined textbook.

Similarly, Nguyen's (2021) examined word repetition in three English textbooks used in Vietnamese high schools, focusing on words beyond the most frequent 2,000 word families, which were considered novel for target learners. The analysis showed that very few of these words were repeated in the textbook, with only 17 word families appearing six times or more. Coupled with the findings of the vocabulary load, Nguyen (2021) concluded that while the textbooks contained an overwhelmingly large number of words that would probably be new to learners, the vast majority of these words would not be encountered often enough for incidental learning to occur.

Taken together, these studies suggest a persistent issue of limited opportunities for lexical recycling in English textbooks. Given that recurrent exposure is essential for vocabulary learning, especially for academic words, it is important to examine whether EAP materials provide sufficient encounters with target items.

In the present study, AWL words played a central role in the EAP programme. Students

were assigned to study a list of 105 AWL words. However, it was not clear regarding the extent to which the EAP themebooks offered chances for students to meet the target words again through reading. Thus, the study identified repetition and distribution of the target academic words in the themebooks.

3. Methodology

3.1. Context

The present study examined instructional materials in an EAP programme at a New-Zealand-based university. At the time of the study, the programme comprised two classes of two different proficiency levels: the Intermediate class for the Intermediate level, and the Academic class for the advanced level. The EAP students took the Updated Vocabulary Levels Test (UVLT) (Webb et al., 2017) as an entry test. The UVLT scores suggested that the students in the Intermediate class had a solid vocabulary of the 2,000 word families but their knowledge of the third 1,000 words onwards and academic vocabulary started to drop significantly. Meanwhile, the students in the Academic class had a partial mastery of 3,000 word families and better academic vocabulary knowledge. Yet, these students had a broad range of vocabulary sizes and many had yet to master the 3,000 word level and demonstrated very limited academic vocabulary.

3.2. Material

This study examined two in-house textbooks called themebooks used in the Intermediate class (the Intermediate themebook) and the Academic class (the Academic themebook). The selected text for this analysis contained 47,200 running words from the Intermediate themebook and 53,939 in the Academic one.

3.3. Instruments and wordlists

3.3.1. AntWordProfiler (Anthony, 2021)

AntWordProfiler (Anthony, 2021) is a free lexical analysis tool designed to estimate the vocabulary profile of texts by comparing them against user-defined word lists. Similar to the Range program (Nation & Heatley, 2002), it calculates the proportion of words belonging to different frequency bands, such as the BNC/COCA word lists (Nation, 2017) or the Academic Word List (Coxhead, 2000). It reports coverage, types, and tokens, enabling a systematic evaluation of lexical demands in teaching materials.

3.3.2. Nation's (2017) BNC/COCA wordlists

Nation's (2017) BNC/COCA wordlists consist of 25,000 word families organised into 25 frequency bands (1,000 per band), derived from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). They also include supplementary lists (e.g., proper nouns, marginal words, transparent compounds, acronyms). These supplementary categories ensure that items not central to general vocabulary knowledge can be distinguished during profiling rather than inflating frequency counts. Nation's BNC/COCA wordlists were used in the lexical profiling analysis of the present study.

3.3.3. Target vocabulary lists

The target vocabulary comprised two sets of 105 words from Coxhead's (2000) Academic Word List, one for the Intermediate EAP class and one for the Academic EAP classes. The lists consisted of the AWL words that were explicitly taught in each course and that also occurred in the reading texts of the corresponding themebooks. Students therefore were expected to learn these target academic words through both explicit vocabulary instruction and repeated exposure in the themebooks.

3.4. The lexical profiling procedures

Two analyses were conducted: (1) vocabulary load of the themebooks, and (2) repetition and distribution of target vocabulary. Prior to analyses, the corpora of the themebooks' text and Nation's (2017) word lists were prepared.

3.4.1. Preparation

The preparation process involved text cleaning and word list adaptation. The corpora comprised only written instructional materials in the themebooks, which included reading

texts and all language learning tasks written in the themebooks. Accordingly, transcripts of audio recordings were excluded because they represented spoken input. Research has shown that academic spoken language is generally less lexically demanding than academic written language (Biber et al., 1998; Nation, 2016); thus, combining transcripts with written texts would have underestimated the vocabulary demands of the written materials. Website links and typographical errors were also removed from the corpora. To compare the lexical features of the two themebooks and the progression of lexical demands within each book, we created two separate sub corpora for the two books and the sub corpus of each unit within a themebook.

Second, Nation's (2017) BNC/COCA word lists were adapted to include off-list words, which were words occurring in the themebooks but not included in Nation's word lists. Once identified, these items were then reassigned to appropriate BNC/COCA lists following these principles:

1. Derived forms were added to existing word families based on Bauer and Nation's (1993) levels of affixation (e.g., *decontextualised* was put in *context* word family).

2. Transparent compounds (e.g., *homestay* and *voluntarists)*, proper nouns, marginal words, and abbreviations were added to their respective supplementary lists. Although *voluntarists* is morphologically a blend of *volunteer* and *tourists*, it was treated as a transparent compound for the purposes of lexical profiling.

3. Hyphenated words (210 items across both themebooks) were treated following Nation (2016). They were split into separate tokens when functioning as modifiers (e.g., *test-taking, rapidly-growing, day-to-day*), or merged into single words if hyphenation was used to separate an affix (e.g., *post-reading, under-theorised*) or in established lexical items such as *stop-overs*, or *off-shore* (see Appendix D for the complete list).

4. Remaining items that could not be assigned to existing lists were grouped into the Other low-frequency words list. These include highly specialised words (e.g., *brominate*, *smoosh*), words containing combining forms (e.g., *megaships, bioenergy, allocentric*), and Māori words (e.g., *iwi)*.

Once all off-list words had been sorted into designated word lists, the lexical profiling task was carried out.

3.4.2. Analysis 1: Vocabulary load analysis

The first analysis aimed to identify the vocabulary load of the reading materials in the EAP themebooks, and the opportunities for the target vocabulary to be encountered in these materials. The vocabulary load was examined by calculating the text coverage for different frequent levels in the BNC/COCA word family lists (Nation, 2017) using the AntWordProfiler program (Anthony, 2021). The vocabulary load analysis (Webb & Nation, 2008) revealed what vocabulary size plus knowledge of proper nouns and marginal words would provide 95% and 98% lexical coverage, the two thresholds recommended by research. The 95% coverage allows acceptable level of comprehension with some assistance (e.g., dictionary) while 98% coverage enables readers to comprehend text unassisted (Laufer & Ravenhorst-Kalovski, 2010).

3.4.3. Analysis 2: Repetition and distribution of the target words

The opportunities of seeing the target words in the reading materials of the EAP themebooks were examined by looking at the repetition and range of the target words and their word families. Repetition of words indicates how often they appear in new contexts, which provides opportunities for learners to encounter different aspects of these target words, such as different derivations, meanings, and collocations (Webb, 2020). Range of the target words shows opportunities for spaced retrievals throughout the EAP course, which contribute to word retention (Webb & Nation, 2017). Specifically, the target words that occurred once, two to five times, six to nine times, and ten times or more were identified and counted. The words that appeared in one reading lesson or across multiple reading lessons were counted.

4. Results

4.1. Vocabulary load of the themebooks

As shown in Table 1, both Intermediate and the Academic themebooks require knowledge of approximately 3,000 word families to reach 95% coverage. However, differences emerge at the 98% text coverage. In the Intermediate themebook, 4,000 word families were needed to reach 98% lexical coverage. The vocabulary load of the Academic themebook was heavier, involving 5,000 word families to reach 98% coverage.

Another noticeable finding relates to progression of lexical demands within the books. While the lexical demands at the 95% level remain relatively stable at around 3,000 word families in both themebooks, the lexical requirements for the 98% coverage show considerable fluctuation across lessons.

In the Intermediate themebook (Table 2), vocabulary demands do not follow a clear progression when reaching 98% coverage. Early lessons (Lessons 1 and 2) require approximately 4000 - 5,000 word families, whereas later units (e.g., Lesson 6 and 7) require fewer, at around 3000 and 4000 words. Variation is also evident in the middle of the book, with Lesson 3 requiring 3,000 word families for 98% coverage, compared to Lesson 5, which requires up to 6,000 words.

A similar lack of progression can be observed in the Academic themebook (Table 3). Over the first four reading lessons, the lexical demands increased gradually for a 98% coverage – 4,000 words in Lesson 1 to 6,000 words in Lesson 4. However, this progression is not sustained, as the lexical demands fluctuate between 4,000 and 5,000 word families in the later lessons.

In general, these findings suggest that both themebooks placed substantial vocabulary demands on learners in the programme. The

Table 1. Vocabulary loads of the intermediate and the academic themebooks (% Tokens)

	Intermediate themebook		Academic themebook	
Word lists	**Coverage (%)**	**Accumulated % with supplementary lists**	**Coverage (%)**	**Accumulated % with supplementary lists**
1st 1,000	77.11	80.73	70.77	73.81
2nd 1,000	9.49	90.22	12.69	86.5
3rd 1,000	7.00	97.22	9.72	96.22
4th 1,000	0.91	98.13	1.61	97.83
5th 1,000	0.64	98.77	0.86	98.69
6th – 25th 1,000	1.22	99.99	1.31	99.99

Table 2. Lexical coverage of individual reading lessons in the intermediate themebook

Word lists	**Cumulative coverage (% tokens) (plus supplementary lists)**						
	Lesson 1	**Lesson 2**	**Lesson 3**	**Lesson 4**	**Lesson 5**	**Lesson 6**	**Lesson 7**
1st 1,000	79.81	76.34	78.39	80.35	74.58	76.13	76.43
2nd 1,000	89.66	90.48	88.51	90.17	85.65	89.32	86.79
3rd 1,000	96.38	97.5	98.11	97.15	96.25	98.33	97.79
4th 1,000	97.27	98.24	98.61	98.09	97.22	99.3	98.52
5th 1,000	98.63	99.07	99.2	98.74	97.77	99.45	99.08
6th 1,000	99.06	99.16	99.79	99.14	98.74	99.67	99.45
7th – 25th 1,000	100	99.99	99.99	100	99.99	100	100

Table 3. Lexical coverage of individual reading sections in the academic themebook

Word lists	Cumulative coverage (% tokens) (plus supplementary lists)							
	Lesson 1	Lesson 2	Lesson 3	Lesson 4	Lesson 5	Lesson 6	Lesson 7	Lesson 8
1st 1,000	77.76	73.78	68.91	69.57	72.6	67.32	68.89	73.62
2nd 1,000	88.1	87.41	84.95	83.1	85.67	84.15	83.48	86.7
3rd 1,000	96.26	96.51	96.64	95.12	94.89	95.17	95.68	96.19
4th 1,000	97.93	97.65	97.9	97.08	97.95	98.15	96.97	98.01
5th 1,000	99.19	98.44	98.89	97.76	98.42	98.67	98.35	98.85
6th 1,000	99.71	98.61	99.25	98.77	99.09	99.46	98.9	99.09
7th - 25th 1,000	100	100	100	100	100	100	100	99.98

95% lexical coverage suitable for assisted reading required knowledge of 3,000 word families for learners of both Intermediate and Academic EAP courses. For independent reading, learners needed greater vocabulary sizes, particularly 4,000 to 5,000 word families overall, with some individual lessons requiring up to 6,000 word families.

4.2. The repetition and distribution of the target words in the themebooks

As shown in Table 4, the Intermediate themebook provides substantial repetition of target vocabulary. Nearly 60% of the target words appear ten times or more, and all items appear at least twice. In terms of distribution, over half of the words occur in three or more lessons. These findings suggest that target vocabulary is distributed relatively well in this book, creating ample opportunities for learners to encounter most target words multiple times in different contexts.

Meanwhile, Table 5 shows more limited repetition in the Academic themebook. Only around 30% of the target words recurred ten times or more, while over half appear five times or fewer, including nine items that occur only once. Distribution is slightly more restricted than the Intermediate themebook, with well under half of the target items appearing across three or more lessons.

Table 4. Repetition and distribution of the 105 target words in the intermediate themebook

		Numbers of target words (max. 105 items)	% of the target words	Examples
Repetition	10 times or over	61	58.1	resource, motive, achieve, adapt, commit
	6 – 9 times	29	27.6	demonstrate, expert, function, element, inevitable
	2 -5 times	15	14.3	exceed, emphasis, response significance, facilitate
	Only once	0	0	N/A
Range (across 7 reading lessons)	Three or more	57	54.3	involve, generate, consume, adapt, focus
	Two	27	25.7	emerge, commit, evaluate, sector, impact
	One	21	20.0	perceive, participate, exceed, inevitable, fundamental

Table 5. Repetition and range of the 105 target words in the academic themebook

		Numbers of target words (max. 105)	% of the target words	Examples
Repetition	10 times or over	31	29.5	sustain, principle, promote, process, vary
	6 – 9 times	12	11.5	implicate, alternative, assign, notion, shift
	2 – 5 times	53	50.5	ongoing, allocate, initiate, assist, implement
	Only once	9	8.5	prohibit, consult, levy, widespread, ambiguity
Range (across 8 reading lessons)	Three or more	44	41.9	academy, positive, sustain, generate, emerge
	Two	35	33.3	input, justify, framework, advocate, reinforce
	One	26	24.8	shift, dispose, release hypothesis, attribute

In general, the Intermediate themebook offers more frequent and more widely distributed meetings with target vocabulary than the Academic themebook. This means that opportunities for incidental learning of academic words are likely more available in the EAP programme at the intermediate level.

5. Discussion

The current textbook analyses examine the lexical demands and opportunities for repeated encounters with target academic vocabulary in two EAP themebooks. Findings suggest a need to reconsider how lexical control, progression, and recycling are embedded in in-house EAP materials.

First, the vocabulary load analysis indicates that both themebooks require substantial vocabulary knowledge to reach 95% to 98% coverage, the benchmarks associated with assisted and unassisted comprehension (Hu & Nation, 2000; Nation, 2022). The Intermediate themebook requires 3000 word families to reach 95% text coverage, which aligns with Nation's (2006) estimation that 3,000–4,000 word families are typically needed for general text comprehension. The majority of students in the Intermediate class, who had not yet mastered the third 1,000 words onwards, would find the themebook challenging with many unfamiliar words. The load of 3,000 to 5,000 words to meet 95-98% coverage of the Academic themebook also exceeded the vocabulary levels of many EAP students. In fact, the UVLT scores revealed wide variations in knowledge of 3,000 to 5,000 words among students upon entering the Academic class. While many learners might be able to reach 95% coverage of text, the majority did not have sufficient vocabulary to attain the 98% threshold for independent reading and incidental learning of vocabulary in the Academic themebook (Webb & Nation, 2017). These findings point to limitations in lexical control, as the difficulty of the texts appears to exceed learners' vocabulary levels. Moreover, the fluctuations in lexical demands across lessons suggests a lack of systematic progression, with no clear or consistent development in difficulty throughout the materials.

Second, the repetition analysis reveals clear differences between the themebooks, highlighting the importance of recycling as a deliberate design feature. The Intermediate themebook offered more favourable conditions for incidental vocabulary learning, as most AWL items occur frequently and are distributed across multiple lessons. This enables more spaced retrievals and exposure to different aspects of word knowledge (e.g., collocations, grammatical use, contextual meanings) (Nation, 2022; Webb, 2020). In contrast, less frequent academic vocabulary in higher-level materials, such as those in the Academic themebook, may need more intentional recycling to ensure sufficient exposure.

Overall, the findings indicate that effective EAP materials need to take into account lexical

demand in relation to learners' vocabulary knowledge, and ensure a more stable progression of difficulty. Importantly, the materials also need to incorporate systematic recycling of target vocabulary. In contexts where such conditions are not met, additional support, such as explicit vocabulary learning or extensive reading, may be necessary to facilitate vocabulary growth (Nation & Waring, 1997; Schmitt, 2010).

6. Conclusions and recommendations

6.1. Summary

The present study adopted lexical profiling analyses to examine the vocabulary demands and learning opportunities provided by two EAP themebooks. The analyses revealed that both themebooks required considerable vocabulary knowledge for adequate comprehension. Specifically, learners would need around 3,000 word families to reach 95% lexical coverage and 4,000 - 5,000 families to achieve the 98% coverage typically associated with unassisted comprehension (Nation, 2022). These findings suggest that, despite being developed by teachers who targeted particular learners, the two themebooks were still lexically demanding for many students in the EAP programme, whose vocabulary knowledge was below these thresholds. This mismatch between text difficulty and learner lexical competence mirrored previous findings from commercial EAP textbook analyses (e.g., Dang & Webb, 2014; Lu & Dang, 2022; Yang & Coxhead, 2022), which reflects the ongoing challenge of adapting EAP materials to suit students' vocabulary knowledge.

In terms of learning opportunities, the repetition analysis showed that both themebooks included recurring exposure to target academic words, though to varying degrees. The Intermediate themebook provided more frequent and distributed encounters with Coxhead's (2000) Academic Word List items, potentially supporting lexical learning through repeated and spaced retrievals (Nation, 2024; Webb & Nation, 2017). The Academic themebook also recycled a subset of target words, but for a large number of items, opportunities for spaced repetitions were quite limited.

6.2. Pedagogical implications

The findings in this study have some important implications for designing and using instructional materials. First, vocabulary demands should be aligned with learners' vocabulary levels. As a guideline, materials should include texts that can reach at least 95% coverage with learners' existing vocabulary, while gradually building towards 98%. This requires regular assessment of learners' vocabulary size, for example by using Webb et al.'s (2017) Updated Vocabulary Levels Test, and the use of profiling tools such as AntWordProfiler (Anthony, 2021) during material development. The teacher could use these data to identify gaps in learners' vocabulary knowledge, provide timely support (e.g., pre-teaching of words, glossaries, or graded readers), and guide learners to personalise their vocabulary learning goal in each lesson.

Second, minimum recycling thresholds could be built into materials design. Target vocabulary should ideally occur at least six to ten times across the course, and be distributed across three or more units, to increase opportunities for retention through spaced retrievals (Webb, 2020; Nation, 2022). Materials writers could integrate this recycling feature by reintroducing target items in later texts, exercises, and consolidation exercises. Importantly, teachers and learners should be informed of these recurrences of target words so that they do not go unnoticed.

Third, lexical profiling should be introduced in the materials development workflow. Profiling tools can be used to check vocabulary load of drafted texts, monitor repetition and distribution of target items, and modify text input to meet predefined thresholds for coverage and recycling. This process can help materials writers monitor and manage lexical demands, progression, and more systematic and deliberate exposure of target vocabulary.

6.3. Limitations

This study has two notable limitations. First, the analyses were limited to two in-house EAP themebooks used in a single institutional context. Although the findings provide useful insights into these materials, they should not

be generalised to other EAP textbooks without further investigation. Second, the study evaluated the potential opportunities for vocabulary learning based on lexical profiling, repetition, and distribution, but it did not examine whether learners actually noticed, learned, and retained the target vocabulary during classroom instruction. As a result, the findings indicate conditions that may facilitate vocabulary learning rather than direct evidence of vocabulary gains.

6.4. Directions for future research

Future research could extend this study in several directions. One is to examine how learners actually interact with the vocabulary in the themebooks, for example, whether frequent words are noticed and retained through classroom exposure. Longitudinal classroom studies combining lexical profiling with pre- and post-course vocabulary testing could shed light on the relationship between textbook input and measurable vocabulary growth.

7. Acknowledgements/Funding

The author gratefully acknowledges the EAP programme director for granting access to the EAP themebooks analysed in this study.

This research project was funded by Victoria University of Wellington, New Zealand.

References

Alexander, L. G., & He, Q. X. (Eds.). (1997). *New Concept English* (Revised ed.). Foreign Language Teaching and Research Press.

Anthony, L. (2021). AntWordProfiler (Version 1.5.1) [Computer software]. <https://www.laurenceanthony.net/software>

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253–279.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly*, *34*(2), 213–238.

Coxhead, A. (2016). Reflecting on Coxhead (2000), "A New Academic Word List." *TESOL Quarterly*, *50*(1), 181–185.

Coxhead, A., Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes, 30*, 66–78.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66–76.

Gardner, D., & Davies, M. (2014). A new Academic Vocabulary List. *Applied Linguistics, 35*(3), 305–327.

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes, 30*(4), 247–257.

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

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.

Lu, C., & Dang, T. N. Y. (2022). Vocabulary in EAP learning materials: What can we learn from teachers, learners, and corpora? *System*, *106*, 102791.

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language, 22*(1), 56–70.

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

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Nation, I. S. P. (2017). *The BNC/COCA word family lists (Version 1.0.0)* [Data set]. Victoria University of Wellington. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nation, I. S. P. (2024). Re-thinking the principles of (vocabulary) learning and their applications. *Languages*, *9*(5), 160.

Nation, I. S. P., & Heatley, A. (2002). *Range: A program for the analysis of vocabulary in texts* [Computer software]. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

Nation, I. S. P., & Waring, R. (1997). Vocabulary size,

text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 6–19). Cambridge University Press.

Nguyen, C. D. (2021). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal, 52*(3), 509–522.

Schmitt, N. (2010). *Researching vocabulary: A Vocabulary research manual*. Palgrave Macmillan.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43.

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System, 93*, 1–13.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., & Nation, I. S. P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal, 16*, 1–10.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics, 168*(1), 33–69.

Yang, L., & Coxhead, A. (2022). A corpus-based study of vocabulary in the New Concept English textbook series. *RELC Journal, 53*(3), 597–611.