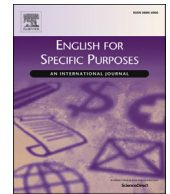




Contents lists available at [ScienceDirect](#)

English for Specific Purposes

journal homepage: <http://ees.elsevier.com/esp/default.asp>



Academic vocabulary in an EAP course: Opportunities for incidental learning from printed teaching materials developed in-house



Sophia Skoufaki ^{a,*}, Bojana Petrić ^b

^a Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

^b Department of Applied Linguistics and Communication, Birkbeck University of London, 26 Russell Square, London WC1B 5DQ, UK

ARTICLE INFO

Article history:

Available online 14 April 2021

Keywords:

EAP
Teaching materials
Vocabulary
Vocabulary learning
Vocabulary repetition

Crown Copyright © 2021 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Commonly defined as the words used more frequently in academic writing and speech across disciplines than in non-academic discourse (e.g., [Nation, 2013](#)), general English academic words pose challenges to both English as a second-language (L2) (e.g., [Evans & Green, 2007](#)) and first-language (L1) students (e.g., [Spencer, Clegg, Lowe, & Stackhouse, 2017](#)). However, academic words are not typically taught at school (e.g., [Beck, McKeown, & Kucan, 2013](#)) nor within subject area courses at university (e.g., [Mudraya, 2006](#)). Within EAP courses, given their typically short duration, multiple goals and students' varied vocabulary needs, few academic words can be taught explicitly. Researchers have therefore highlighted the value of incidental vocabulary learning through reading ([Gardner, 2013](#)) and teaching materials as sources of academic vocabulary exposure ([Stoller, 2016](#)) in EAP provision. Of particular relevance to promoting incidental vocabulary learning via

* Corresponding author.

E-mail addresses: sskouf@essex.ac.uk (S. Skoufaki), b.petric@bbk.ac.uk (B. Petrić).

reading are printed teaching materials, i.e., materials written and presented in paper-based or digital formats to be consumed via reading, such as textbooks and teacher-created handouts.

Research on the vocabulary input provided by teaching materials has focused on textbooks, particularly those for general English language teaching (e.g., Criado, 2009; Matsuoka & Hirsh, 2010), while EAP textbooks have been examined less commonly (e.g., Miller, 2011). However, studying textbooks alone provides only partial insights into students' academic vocabulary exposure through printed materials in EAP contexts, since EAP provision typically draws on materials developed in-house (Jones & Durrant, 2010; Stoller, 2016). Such materials combine parts of different textbooks and/or supplement textbooks with materials selected or created by EAP practitioners to respond to specific student needs, such as journal articles, locally produced course booklets, and teacher-created handouts with exercises or guidelines.

There is, therefore, both a research gap and a practical need to examine the extent to which printed EAP teaching materials developed in-house expose students to academic vocabulary, thus offering opportunities for incidental learning of academic vocabulary. In line with these goals, this exploratory study examines the occurrence of general English academic words, operationalised as Gardner and Davies's (2014) Academic Vocabulary List (AVL), in the printed EAP teaching materials created in-house for a preessional EAP course. It also examines how frequently general academic words are repeated in these printed materials to see whether this repetition rate aligns with the guidance provided by research into incidental vocabulary learning from reading.

1.1. Operationalising English academic vocabulary

Academic vocabulary is often operationalised as one of the wordlists which identify academic lexical items likely to be useful to students preparing for or already at university. We operationalised English academic vocabulary as the AVL for the following reasons.

The AVL lists the most frequent 3,014¹ academic lemmas in the Corpus of Contemporary American English (COCA). It consists of words from all frequency bands of COCA which occur at least 50% more frequently in the Academic section of COCA than would normally be expected, are evenly distributed across its disciplinary sections and occur in at least seven of COCA's eight disciplinary sections. This wordlist was preferred over wordlists which distinguish between very frequent and academic vocabulary, such as Coxhead's (2000) Academic Word List (AWL), because defining academic vocabulary as different from high-frequency vocabulary means that some words can be labelled 'general' or 'academic' depending on which words one considers 'high-frequency' (e.g., Masrai & Milton, 2018). The frequency-based distinction between general and academic vocabulary has also been contested because words in the most frequent 2,000 word families may have academic meanings, whereas less frequent words may not (e.g., Eldridge, 2008).

Another reason for choosing the AVL is that it uses the lemma (i.e., the root word form with a specific part of speech [POS] and its inflected forms) as its unit. The lemma seems to be a better unit than the word family because word-family wordlists assume that learners are able to infer the meanings of unknown words in their academic reading thanks to their knowledge of inflectional suffixes and derivational affixes (Coxhead, 2000), an assumption which has not been supported by research (e.g., Ward & Chuenjundaeng, 2009). Moreover, the lack of POS tagging in word-family wordlists means that words with the same spelling are counted together, although their frequency and meaning may differ, depending on POS. For example, the noun *group* is lemma number 2 in the AVL, whereas the verb *group* is lemma number 1,339.

In addition to the aforementioned reasons, we operationalised English academic words as the AVL because of research findings about its pedagogical relevance to students' reading and writing academic vocabulary needs, which is in line with our focus on academic vocabulary in printed EAP teaching materials. These findings are discussed in the rest of this section.

Text coverage (i.e., the percentage of running words a wordlist can cover) has been used to estimate how well an academic wordlist caters to students' reading vocabulary needs. The rationale is that wordlists covering a large percentage of text tokens include words worthy of learning for reading-comprehension purposes. The AVL provides nearly double the coverage of the academic sections of COCA and BNC (13.8% and 13.7%, respectively) than the AWL (7.2% and 6.9%, respectively) (Gardner & Davies, 2014).

As for students' writing academic vocabulary needs, Durrant (2016) examined the utility of the AVL for students' writing by locating AVL lemmas in the British Academic Written English (BAWE) corpus, a corpus of the written assignments of undergraduate and Masters level students at four British universities. The AVL provided high coverage of BAWE (16.82%), but this coverage was higher for Social Science than Hard Science writing. Moreover, 427 AVL lemmas – all from the most frequent 1,000 AVL lemmas – appeared more than 12 times per million tokens in 28 or more of the 31 BAWE disciplines. Durrant (2016) considers these 427 AVL lemmas useful for the writing needs of all university students. However, due to the low text numbers for some disciplines, BAWE is not equally representative of students' writing across disciplines (Durrant, 2016); thus, an AVL-search of a more representative student writing corpus may indicate a longer list of shared AVL lemmas. Moreover, most of the BAWE texts were written by undergraduate students, thus leaving the possibility for a longer list of AVL lemmas shared across disciplines in a corpus balanced between undergraduate and postgraduate writing samples. For these reasons and others (see Gardner & Davies, 2016), it is unclear whether only these 427 AVL lemmas or more are used across disciplines in

¹ The AVL really contains 3,014 lemmas because in the 3,015-lemma AVL provided as supplementary material in Gardner and Davies (2014) the entry *disproportionately* appears twice (Durrant, 2016).

students' university writing. Despite these limitations, this list is a good indication of the academic vocabulary students need for writing at university, particularly in the UK context.

In conclusion, we consider the AVL a relatively good operationalisation of English academic words because it provides high coverage of expert writing, which university students are likely to read, and the AVL lemmas which are frequent in BAWE provide a good estimate of the English academic words students across various disciplines are likely to use in their writing tasks.

The following section examines the role that vocabulary frequency can play in incidental vocabulary learning which can, in turn, support performance in reading and writing tasks at university.

1.2. Vocabulary frequency and incidental vocabulary learning

Most studies indicate that the more frequent a word is in a learner's input, the more likely it is to be learned (Reynolds & Wible, 2014). As can be expected, however, other factors, such as learner characteristics and characteristics of the learning activity as well as the kind of vocabulary knowledge (recognition or recall) that is being tested, affect study results as well (Uchiyara, Webb, & Yanagisawa, 2019). Since studies on incidental vocabulary learning differ in terms of the aforementioned factors, they have, unsurprisingly, led to various estimates of the repetition rate that can ensure vocabulary learning for a large proportion of the learners participating in a study. This rate ranges from 8 to 10 occurrences (e.g., Webb, 2007) to at least 20 occurrences (e.g., Waring & Takaki, 2003).

Recent research on the role of vocabulary knowledge in reading comprehension indicates that the ability to recall word meaning (meaning recall) is a more reliable predictor of reading comprehension than the ability to recognise word meaning (meaning recognition) (e.g., McLean, Stewart, & Batty, 2020). These findings mean that the traditional view that vocabulary recognition knowledge is sufficient for good reading comprehension and vocabulary recall knowledge is necessary only for writing (e.g., see Paquot, 2010, pp. 15–16) needs to be abandoned. They also mean that helping students build their ability to recall knowledge of vocabulary, both academic and general, is advisable for the development not only of writing, but also reading performance. Therefore, a blanket recommendation could be to present words to learners as many times as required for recall knowledge to develop. As mentioned above, given the various factors that affect incidental vocabulary learning, no specific number of word occurrences can ensure the development of recall knowledge for word form, meaning, collocation and syntactic properties. However, if a threshold is to be specified for pedagogical purposes, at least 10 occurrences of a word are advisable in printed materials (e.g., Webb, 2007).

1.3. Academic vocabulary in EAP materials

Teaching materials include a broad range of written and audio(visual) texts, from teacher-designed worksheets to textbooks. Research on the linguistic content of EAP materials falls under two categories. One is based on comparisons of selected linguistic and discursive features in corpora of naturally occurring language with their occurrence in textbooks to determine whether textbooks accurately represent them. For example, Conrad (2004) found numerous discrepancies between linguistic patterns in a practice academic lecture in a textbook and a corpus of real-world university lectures, concluding that textbooks do not provide students with appropriate preparation for listening to lectures.

Researchers have also compared selected linguistic features of EAP textbooks and introductory university textbooks in specific disciplines to determine whether EAP textbooks provide students with appropriate preparation for reading subject textbooks. Of particular interest is Miller's (2011) comparison of academic vocabulary, among other features, in 75 passages from three textbooks commonly used in intensive English reading courses in US universities and in textbooks used in the first two years of undergraduate study in 18 disciplines. Miller (2011) found statistically significant differences in the percentages of AWL vocabulary in the two corpora, with AWL items constituting 4.78% and 8.8% of the total running words in the reading textbook and university textbook corpora, respectively. Miller (2011) concludes that the English reading skills textbooks provided low exposure to the academic vocabulary students would encounter in undergraduate subject textbooks and recommends the use of supplementary materials on the reading course.

Materials research has been criticised for predominantly focusing on textbooks (e.g., Harwood, 2014). This criticism is even more pertinent to EAP teaching materials, which are typically developed in-house; EAP practitioners tend to combine textbook and authentic materials and create their own materials in response to their students' needs (Jones & Durrant, 2010; Stoller, 2016). Although there are excellent accounts of in-house materials development for EAP and ESP courses (e.g., Feak & Swales, 2010), research on the linguistic content of in-house EAP materials is rare. The only such study, to our knowledge, is Jones and Durrant's (2010) small-scale analysis of academic vocabulary in a sample of in-house reading and writing EAP materials in a UK university. Using the non-academic parts of the BNC as their reference corpus, the researchers identified keywords in the materials (i.e., words uniquely frequent in the materials corpus in comparison to the reference corpus). The fifty most frequent keywords in the materials included the AWL words *academic* and *project*, with other AWL items featuring further down on the frequency list. They tentatively conclude that in-house materials 'may be suitable for the teaching of academic vocabulary' (p. 393) and offer recommendations for in-house materials development and for corpus-based approaches to vocabulary teaching such as awareness-raising tasks based on concordance lines.

In conclusion, research on academic vocabulary input provided by EAP materials, particularly those produced in-house, is limited. Given the widespread use of in-house materials in EAP courses, it is important to understand the extent to which such materials provide students with exposure to general academic vocabulary.

2. The present study

Given the central role materials play in language learning and the very limited research on the academic vocabulary students are exposed to through EAP materials developed in-house (see section 1.3), this study aims to examine general academic vocabulary in the in-house printed materials used in an EAP course. In addition to examining which general academic words appear in these materials, this study also examines whether their repetition rate is sufficient – according to research-based recommendations (see section 1.2) – for the development of word recall knowledge necessary for reading and writing tasks. *General academic vocabulary* is operationalised as the AVL (Gardner & Davies, 2014) (see section 1.1). The AVL lemmas which appeared more than 12 times per million word tokens in BAWE in 28 disciplines or more (Durrant, 2016), referred to as ‘AVL-in-BAWE’ henceforth, are the operationalisation for academic words likely to be useful for students’ academic writing (see section 1.1).

This study is guided by the following questions:

1. To what extent do the in-house printed EAP materials used in a UK university preessional EAP course expose students to general academic words, operationalised as the AVL?
2. To what extent do the printed EAP materials used in this course expose students to academic words which they may need to use in their academic writing, operationalised as the AVL-in-BAWE wordlist?
3. Are the AVL lemmas in the printed EAP materials repeated frequently enough for the incidental development of recall knowledge?

3. Method

3.1. Context of the study

This study identified AVL lemmas in the materials used in a preessional course at a British university. This course was offered to prospective students with an IELTS overall score of 5.5 or 6, who had applied for MA courses in Social science subjects for which the English language entry requirement was 6 or 7, respectively.

This preessional course was intensive (24 h per week). English for General Academic Purposes (EGAP) instruction was provided in the first five weeks, followed by English for Specific Academic Purposes (ESAP) instruction in the remaining 10 weeks. We focused on the EGAP course to obtain findings as relevant as possible to similar university contexts; different universities teach ESAP differently, depending on the disciplines taught, but EGAP instruction is offered across universities.

In this EGAP course, teaching was structured in terms of the language skills and language knowledge that students need to develop to perform well in tasks at university. Three modules were taught, each by a different teacher: Reading and Writing (Reading/Writing), Listening and Speaking (Listening/Speaking), and Vocabulary and Grammar (Vocabulary/Grammar).

Research ethics approval to collect data from the EAP teachers was granted by the university offering this course. Before the course started and after receiving permission from the course director, we emailed the teachers a call for participation detailing the study’s aims and what participating teachers would do. All teachers agreed to participate. At the end of the study they received £30 as compensation.

3.2. Printed teaching materials

The printed materials used during the EGAP preessional course were developed in-house over the years by the EAP team. The EAP provision at this university was accredited by BALEAP (<https://www.baleap.org>), attesting to its high quality. Teachers also had the autonomy to adapt the materials during the course. As we aimed to collect the materials actually used in the course, at the end of each week, each teacher provided us with photocopies of the printed materials used that week. Since the study examines the general academic vocabulary included in printed in-house EAP materials, audio(visual) materials were not collected.

The printed materials of the Reading/Writing module included (i) a course booklet, which contained information about the course (e.g., timetable), information on aspects of academic writing conventions (e.g., citing) and exercises based on set readings, and (ii) four research articles from academic journals of which three formed a thematic set (focusing on international students studying abroad). The Vocabulary/Grammar module used units from the textbook *Oxford EAP intermediate/B1+* (De Chazal & Rogers, 2013) and teacher-prepared handouts with tasks and assignment instructions. The printed materials of the Listening/Speaking module consisted of teacher-prepared handouts with tasks, brief explanations (e.g., of the phonetic alphabet), and presentation guidelines.

Teachers were also interviewed at the end of each week of the course about the materials and tasks used that week. Here we only report on their decisions to supplement the existing materials regarding academic vocabulary. All three teachers

reported supplementing the in-house materials by the materials they selected or created; two such instances were reported that concerned academic vocabulary directly or indirectly. The Vocabulary/Grammar teacher, while mostly drawing on selected units from the *EAP Oxford EAP intermediate/B1+* (De Chazal & Rogers, 2013), which she considered ‘a good standard’ and ‘good as a base’, developed multiple worksheets with exercises to ‘scaffold a bit’, i.e. facilitate students’ engagement with textbook linguistic content. In week 2 she used a vocabulary building worksheet focusing on selected words from the textbook she believed students needed more practice with. The task required students to supply family members of the words, many of which were AVL lemmas (e.g., the adverb *theoretically* when given *theory* as a cue). Seeing students struggle with the jargon-heavy research article she had used, the Reading/Writing teacher decided to include two additional research articles in week 3, which she had selected as a more suitable model for teaching the components of research articles while using less technical vocabulary and addressing a topic closer to students’ experience (the experiences of international students studying abroad). The Listening/Speaking teacher also used his own materials; however, he did not refer to vocabulary when explaining their purpose.

It is important to highlight that although teachers’ materials selection was based on their ongoing assessment of the students’ needs and progress, they did not report having examined their chosen materials for academic vocabulary instances or repetition. The corpus in this study, therefore, allows us to examine materials resulting from experienced EAP teachers’ authentic context-driven and pedagogically-informed decisions. While the actual materials are specific to this course, the processes of their development and compilation are likely to be common across similar EAP programmes.

3.3. Corpus

Photocopies of the teaching materials were scanned and converted to plain text via optical character recognition software.

Table 1 summarises the breakdown of tokens among module subcorpora and weeks of teaching. As can be seen, the corpus contains 83,991 tokens. Most tokens (72.71%) appear in the Reading/Writing subcorpus, followed by the Vocabulary/Grammar subcorpus (22.36%) and the Listening/Speaking subcorpus (0.53%).

Table 1

Word tokens and lemmas in the corpus by module and week.

Module	Week					Total words
	1	2	3	4	5	
Reading/Writing	21,024	0	39,796	0	0	60,820
Listening/Speaking	841	915	658	1976	0	4,390
Vocabulary/Grammar	4880	2,844	2,756	4,170	4,131	18,781
Total words	26,745	3,759	43,210	6,146	4,131	83,991

Table 1 indicates that students were not given new materials in the Reading/Writing module in weeks 2, 4 and 5 and in the Listening/Speaking module in week 5. The interviews with the teachers reveal the reasons for this. In the Reading/Writing module, in week 2 students worked with an article they had been given in week 1; in week 4 they were engaged in reading and writing activities based on the two articles presented in week 3. In week 5, students did in-class writing in the Reading/Writing module and delivered oral presentations in the Listening/Speaking module.

3.4. Procedure

Files were converted into .txt files. They were then POS tagged via TagAnt (Anthony, 2015), a free POS tagger. TagAnt flags out potentially wrong POS codes in its output. These codes were manually checked and corrected where necessary by the first author.

The corpus was lemmatised because we needed to identify the lexical lemmas in general and the AVL lemmas in particular to answer Research question 1. To ensure correct lemmatisation, instead of following the common practice of lemmatising the raw corpus, we lemmatised the POS-tagged corpus using a POS-tagged lemma list.²

In particular, the AntBNC lemma list (available at <https://www.laurenceanthony.net/software/antcon/>) was modified by the first author in two ways:

- AVL lemmas were identified in the AntBNC lemma list. Then, the word forms which belonged to each lemma were identified in the 100,000 word COCA frequency lists (available at https://www.wordfrequency.info/100k_samples.asp). Word forms additional to those in the AntBNC lemma list were added to the list.
- POS tags matching those used in TagAnt were added to all word forms in the list.

Data were analysed with SPSS 25.

² To our knowledge, no existing lemmatiser software produces error-free results. Errors occur in terms of word stemming (e.g., *encouraging* may be an adjective in a sentence, but a lemmatiser may wrongly map it onto the verb *encourage*); word forms ambiguous between two lemmas may be assigned to the wrong lemma (e.g., *process* may be a noun in a sentence but a lemmatiser tags it as a verb).

4. Results

This section begins with addressing Research question 1 ('To what extent do the in-house printed EAP materials used in a UK preessional EAP course expose students to general academic words, operationalised as the AVL?') in section 4.1. The coverage that AVL tokens and lemmas provide to the printed materials corpus is examined first, followed by the AVL lemmas which appear in the corpus. Findings in relation to Research question 2 ('To what extent do the EAP materials used in this course expose students to academic words which they may need to use in their academic writing, operationalised as the AVL-in-BAWE wordlist?') are reported in section 4.2. Findings in relation to Research question 3 ('Are the AVL lemmas in the printed materials repeated frequently enough for the incidental development of recall knowledge?') are reported in section 4.3.

4.1. To what extent do the in-house printed EAP materials used in a UK preessional EAP course expose students to general academic words, operationalised as the AVL?

The AVL covers 5.3% of the word tokens in the corpus. This coverage is lower than that of the academic section of COCA (13.8%) and the BNC (13.7%) (see Section 1.1). This difference can be due to the different functions of these corpora: the preessional EAP teaching materials were created to help international students improve their English proficiency level and familiarise themselves with the demands of UK university study, whereas the reference corpora contain academic journal articles, a form of scholarly communication among peers. The more limited variety of topics addressed in the preessional EAP teaching materials compared to those in the reference corpora is another possible reason. Differences in topic variety are to be expected given not only the different functions of the texts in these two kinds of corpora, but also their different corpus sizes; the printed materials corpus is 83,991 tokens, whereas BNC academic is 32,828,961 tokens and COCA academic was 120,847,709 when the AVL was compiled (Gardner & Davies, 2014).

4.1.1. Lexical-word corpus coverage by the AVL

The AVL coverage of lexical word (i.e., noun, verb, adjective and adverb) tokens was examined because the AVL contains only lexical words (see also Durrant, 2016 for this approach to examining corpus coverage by the AVL). Table 2 shows the cumulative lexical coverage by AVL lemmas; the total lexical coverage (13.44%) is divided into 2% segments.

Table 2
Lexical coverage from AVL lemmas.

Cumulative coverage	AVL lemmas unique to each 2% lexical-token corpus increment	Counts in the materials of AVL lemmas unique to each 2% lexical-token corpus increment			Cumulative AVL lemmas
		Maximum	Minimum	Median	
2%	14	92	37	42	14
4%	15	37	25	30	29
6%	33	25	16	20	62
8%	52	16	10	13	114
10%	89	10	6	7	203
12%	173	6	3	4	376
13.44%	470	3	1	1	846

Table 2 indicates that, similar to previous research (e.g., Durrant, 2016), most of the coverage comes from a relatively small number of AVL lemmas. Out of the 846 AVL lemmas appearing in the materials, 62 (7.33%) account for nearly half (6%) of the coverage. The fact that most of the lexical coverage is taken up by a few AVL lemmas is reflected in the diminishing maximum, minimum and median AVL lemma counts (i.e., instances per AVL lemma unique to each 2% lexical-token corpus increment) as we move down the table. For example, there is a stark contrast between the maximum, minimum and median counts of the 14 AVL lemmas that account for the first 2% of the lexical coverage and these descriptive statistics for the 470 AVL lemmas which account for the last 2% increment.

Table 3 provides more detail about corpus coverage by the AVL. It shows the lexical-word coverage of the tokens of lexical words in the materials of each module and in the whole corpus by each AVL frequency band³ and the AVL overall.⁴ It indicates that AVL lemmas from the first band provided nearly all of the AVL coverage of the materials used in every module.

³ Gardner and Davies (2014) do not divide the AVL list into bands. Throughout this paper, the 'first AVL band' and the 'second AVL band' each consist of 1,000 lemmas and 'the third AVL band' consists of 1,014 lemmas.

⁴ In all tables percentage numbers are rounded to the second decimal place.

Table 3

AVL corpus coverage per teaching module.

Module	Lexical word tokens	Tokens per AVL frequency band (percentage of lexical word tokens)			Total AVL tokens
		1	2	3	
Reading/Writing	23,244	1,829 (7.9%)	161 (0.69%)	28 (0.12%)	2,018 (8.68%)
Listening/Speaking	1,902	331 (17.4%)	32 (1.68%)	4 (0.21%)	367 (19.3%)
Vocabulary/Grammar	8,022	1,885 (23.5%)	159 (1.98%)	28 (0.35%)	2,072 (25.83%)
All modules	33,168	4,045 (12.2%)	352 (1.06%)	60 (0.18%)	4,457 (13.44%)

A comparison across subcorpora shows that AVL coverage was the highest for the Vocabulary/Grammar subcorpus (25.83%), second highest for the Listening/Speaking subcorpus (19.3%) and relatively low for the Reading/Writing subcorpus (8.68%). A chi-squared test indicates that these differences are significant overall, $\chi^2(2) = 1129.17$, $p < .001$. Pairwise comparisons via z-tests with Bonferroni correction indicate that AVL-to-lexical token proportions differed significantly in all pairs of subcorpora.

These significant differences among subcorpora in terms of AVL coverage may reflect differences in AVL lemma variation and/or repetition. The more varied the lemmas in a text/corpus are, the more likely they are to cover more text; conversely, the more lemmas are repeated in a text/corpus on average, the lower the coverage they offer. The roles that AVL lemma variation and repetition played in AVL coverage across subcorpora and the AVL lemmas in the materials per subcorpus are examined in section 4.2.

4.1.2. AVL lemmas per subcorpus

In total, 846 AVL lemmas appeared in the EAP printed materials. [File 1 in the Supplementary materials](#) lists the AVL lemmas in the teaching materials.

[Table 4](#) shows the coverage that AVL lemmas provide to all lexical lemmas (i.e., AVL and non-AVL lemmas) and to AVL tokens inside the materials of each module.

Table 4

Lexical lemmas and AVL lemmas per module and in all the materials.

Module	Lexical lemmas	AVL lemmas in the materials (percentage of lexical lemmas)	Coverage of AVL tokens by AVL lemmas
Reading/Writing	4,418	525 (11.88%)	26.02%
Listening/Speaking	858	171 (19.93%)	48.17%
Vocabulary/Grammar	2,512	573 (22.81%)	27.56%
All modules	5,896	846 (14.35%)	18.98%

The column 'AVL lemmas in the materials' in [Table 4](#) indicates that AVL lemmas in the Listening/Speaking and Vocabulary/Grammar subcorpora form a higher percentage of lexical lemmas than in the Reading/Writing subcorpus. A chi-squared test comparing the number of AVL and non-AVL lexical lemmas across subcorpora indicates that these proportions differed among subcorpora, $\chi^2(2) = 149.55$, $p < .001$. Z tests with Bonferroni correction indicate that the AVL to non-AVL lemma ratio a) was significantly lower in the Reading/Writing subcorpus than that in either of the other two subcorpora and b) did not differ significantly between the Listening/Speaking and Vocabulary/Grammar subcorpora. These results indicate that the lower coverage of lexical tokens by the AVL in the Reading/Writing subcorpus compared to that in the other two subcorpora (see section 4.1.1) is due, at least in part, to the more limited AVL lemma variation in this subcorpus than in either of the other subcorpora.

The last column in [Table 4](#) shows the coverage of AVL tokens from AVL lemmas per subcorpus. A chi-squared test indicates that AVL lemma-to-token proportions differed significantly among subcorpora, $\chi^2(2) = 32.65$, $p < .001$. Z tests with Bonferroni correction indicate that AVL lemma-to-token proportion is significantly higher in the Listening/Speaking subcorpus than in each of the other subcorpora. Therefore, higher AVL lemma repetition rate can explain the lower coverage of lexical tokens by the AVL in the Reading/Writing subcorpus when it is compared with the Listening/Speaking subcorpus, but not when it is compared with the Vocabulary/Grammar subcorpus. Differences in AVL lemma repetition among subcorpora are further explored in section 4.3.

4.1.3. AVL-lemma overlap among and between subcorpora

Half (423) of the AVL lemmas in the printed materials corpus are shared, some by pairs of subcorpora and others by all subcorpora. [Table 5](#) offers a detailed breakdown of AVL lemmas across the subcorpora. It first shows all the AVL lemmas in each subcorpus per AVL frequency band. It then shows how many of these AVL lemmas are shared between and across subcorpora and how many are unique to each subcorpus. Inside parentheses are the percentages of lemmas from each AVL frequency band, where the total of lemmas in each of the first two bands is 1,000 and in the third band 1,014.

Table 5

AVL lemmas per subcorpus, shared between and among subcorpora, and unique to each subcorpus.

AVL lemmas	Subcorpora	AVL frequency band			Total AVL lemmas
		1	2	3	
In	Reading/Writing	417 (41.7%)	84 (8.4%)	24 (2.37%)	525
	Listening/Speaking	151 (15.1%)	17 (1.7%)	3 (0.3%)	171
	Vocabulary/Grammar	472 (47.2%)	79 (7.9%)	22 (2.2%)	573
Shared	by all subcorpora	89 (8.9%)	1 (0.1%)	0	90
Shared only between	Reading/Writing and Listening/Speaking	16 (1.6%)	2 (0.2%)	1 (0.1%)	19
	Reading/Writing and Vocabulary/Grammar	186 (18.6%)	13 (1.3%)	0	199
	Listening/Speaking and Vocabulary/Grammar	25 (2.5%)	0	0	25
Unique to	Reading/Writing	126 (12.6%)	68 (6.8%)	23 (2.27%)	217
	Listening/Speaking	21 (2.1%)	14 (1.4%)	2 (0.2%)	37
	Vocabulary/Grammar	172 (17.2%)	65 (6.5%)	22 (2.17%)	259

As expected, the vast majority of AVL lemmas unique to a subcorpus occurred in the two large subcorpora, Vocabulary/Grammar and Reading/Writing. These subcorpora also mutually reinforced their AVL lemma repetition rate since they shared 199 AVL lemmas (i.e., 37.9% of the AVL lemmas in the Reading/Writing subcorpus and 34.73% of those in the Vocabulary/Grammar subcorpus). Conversely, the Listening/Speaking subcorpus overlapped to a small extent with each of the other subcorpora: The AVL lemma overlap between each of these subcorpora and the Listening/Speaking subcorpus accounted for about 4% of the total AVL lemmas in each of the large subcorpora (4.36% of the AVL lemmas in the Vocabulary/Grammar subcorpus and 3.62% of those in the Reading/Writing subcorpus) and for around 10% of the total AVL lemmas in the Listening/Speaking subcorpus (14.62% and 11.11%, respectively).

Table 5 indicates that most AVL lemmas come from the first AVL band, irrespective of whether we consider all AVL lemmas in each subcorpus, only those shared by two or all subcorpora or those unique to each subcorpus. This predominance of first-band AVL lemmas in the materials indicates that many AVL lemmas which occurred more than 12 times per million word tokens in BAWE in 28 disciplines or more (Durrant, 2016) are likely to appear in the materials. The next section examines this issue.

4.2. To what extent do the EAP materials used in this course expose students to academic words which they may need to use in their academic writing, operationalised as the AVL-in-BAWE wordlist?

Table 6 presents the breakdown of the 427 AVL-in-BAWE lemmas (Durrant, 2016) across subcorpora.

Table 6

AVL-in-BAWE lemmas per subcorpus, shared between and among subcorpora, and unique to each subcorpus. The percentage of AVL lemmas in all (175 lemmas), 30 (105 lemmas), 29 (77 lemmas) and 28 BAWE disciplines (70 lemmas) per subcorpus (combination) appears within parentheses.

Subcorpora		AVL lemmas shared by BAWE disciplines				Total
		All	30	29	28	
In	Reading/Writing	145 (82.86%)	60 (57.14%)	43 (55.84%)	29 (41.43%)	277 (64.87%)
	Listening/Speaking	73 (41.71%)	23 (21.91%)	8 (10.39%)	10 (14.29%)	114 (26.7%)
	Vocabulary/Grammar	142 (81.14%)	69 (65.71%)	53 (68.83%)	36 (51.43%)	300 (70.26%)
Shared by	Reading/Writing and Listening/Speaking	7 (4%)	4 (3.81%)	0	0	11 (2.58%)
	Reading/Writing and Vocabulary/Grammar	63 (36%)	33 (31.43%)	30 (38.96%)	15 (21.43%)	141 (33.02%)
	Listening/Speaking and Vocabulary/Grammar	2 (1.14%)	7 (6.67%)	1 (1.3%)	4 (5.71%)	14 (3.28%)
Unique to	All subcorpora	62 (35.43%)	9 (8.57%)	5 (6.49%)	5 (7.14%)	81 (18.97%)
	Reading/Writing	13 (7.43%)	14 (13.33%)	8 (10.39%)	9 (12.86%)	44 (10.3%)
	Listening/Speaking	2 (1.14%)	3 (2.86%)	2 (2.6%)	1 (1.43%)	8 (1.87%)
	Vocabulary/Grammar	15 (8.57%)	20 (19.05%)	17 (22.08%)	12 (17.14%)	64 (14.99%)

Since BAWE is a corpus of university students' written assignments, we were particularly interested in seeing how many AVL-in-BAWE lemmas were included in the Reading/Writing and Vocabulary/Grammar subcorpora, i.e., the materials used to foster students' academic reading and writing skills. High percentages of the lemmas in this list appear in the Reading/Writing and Vocabulary/Grammar subcorpora and one third of the list is shared between them.

The Listening/Speaking subcorpus, however, includes only 26.7% of the AVL-in-BAWE lemmas and overlaps very little with each of the other subcorpora. These findings are to be expected since academic vocabulary lists extracted from written corpora do not offer as good a coverage for corpora of academic spoken English as they do for corpora of academic written English (e.g., Dang & Webb, 2014).

An encouraging finding is that the coverage that the AVL-in-BAWE sublists (i.e., the lemmas shared by all, 30, 29 or 28 BAWE subdisciplines) offer to the materials subcorpora follows a falling trend as we move from the highest to the lowest

number of BAWE subdisciplines. This falling trend is disrupted only by a minor difference between the coverage of the 29-discipline (10.39% coverage) and 28-discipline (14.29% coverage) sublists in the Listening/Speaking materials.

After taking into consideration the lemmas that overlap between and among subcorpora, 363 AVL-in-BAWE lemmas appear in the teaching materials. This number represents 85.01% of all AVL-in-BAWE lemmas.

4.3. Are the AVL lemmas in the printed materials repeated frequently enough for the incidental development of recall knowledge?

This section reports on findings in relation to Research question 3. Section 4.3.1 addresses Research question 3 in terms of AVL-lemma repetition in the printed materials. Section 4.3.2 addresses it in terms of AVL-in-BAWE lemma repetition in the printed materials.

4.3.1. How many AVL lemmas receive the repetition level considered necessary for the incidental development of recall vocabulary knowledge through reading?

We examine how many of the AVL lemmas meet the 10-or-more occurrences requirement because this number of occurrences has been found to be necessary for form and meaning recall of at least about one third of the unknown words encountered in reading (e.g., Webb, 2007). Since the EAP course this study focuses on was an intensive one, it would be too demanding to expect students unfamiliar with the AVL words in the materials to learn more than about one third of them. File 2 in the supplemental materials provides a table with descriptive statistics about AVL lemma repetition in the printed materials.

AVL lemma counts in the materials were divided into three bands. First, a distinction was made between words which occurred in the printed materials 10 or more times and those which occurred less than 10 times because we aimed to see whether the printed materials provided students with enough lemma occurrences so that they were likely to be able to recall the meaning and form of at least one third of the academic lemmas in the materials. Second, a distinction was made between words which occur only once and those which occur more than once because the former are less likely to be learned than the latter (e.g., Waring & Takaki, 2003). Consequently, Band 1 included the AVL lemmas occurring once, Band 2 included those occurring 2–9 times and Band 3 included those occurring 10 or more times.

Figure 1 shows the percentage of AVL lemmas appearing inside the materials with each of these frequency levels. These percentages are organised per AVL frequency band. In addition to percentages, the bar labels show how many times lemmas inside each AVL frequency band appeared in the materials once, 2–9 times or 10 or more times.

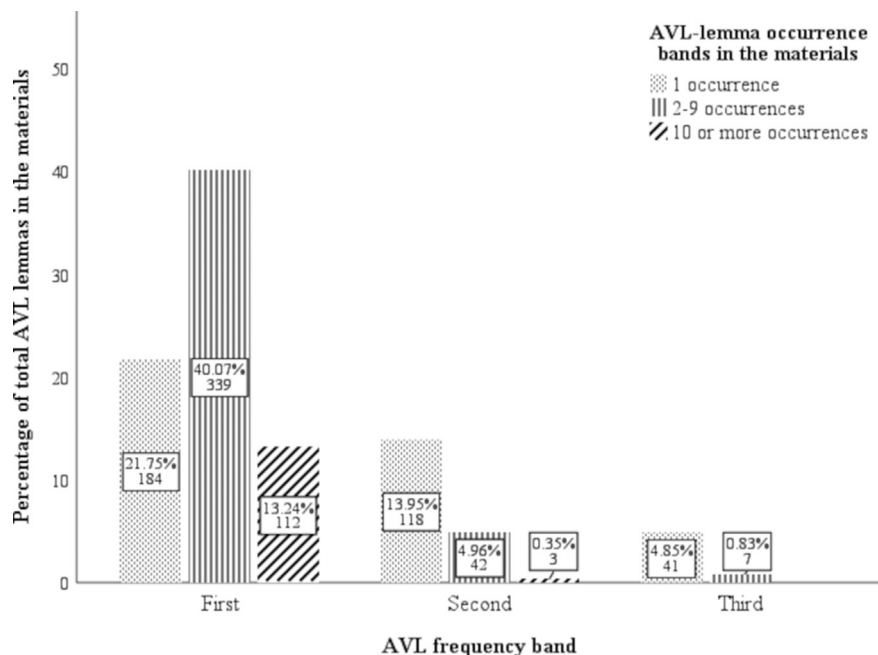


Figure 1. Percentage of AVL lemmas occurring once, 2–9 times and 10 or more times in the materials per AVL frequency band.

Figure 1 indicates that AVL lemmas that occurred 2–9 times form the majority (45.86%) of AVL lemmas in the materials. Only 13.59% of all the AVL lemmas in the materials occur 10 or more times; all but three come from the most frequent 1,000 lemmas in the AVL. Therefore, an AVL lemma was more likely to appear 10 or more times in the materials if it was among the most frequent 1,000 AVL lemmas. To examine whether this tendency also exists in each subcorpus, Figures 2–4 summarise these findings per subcorpus.

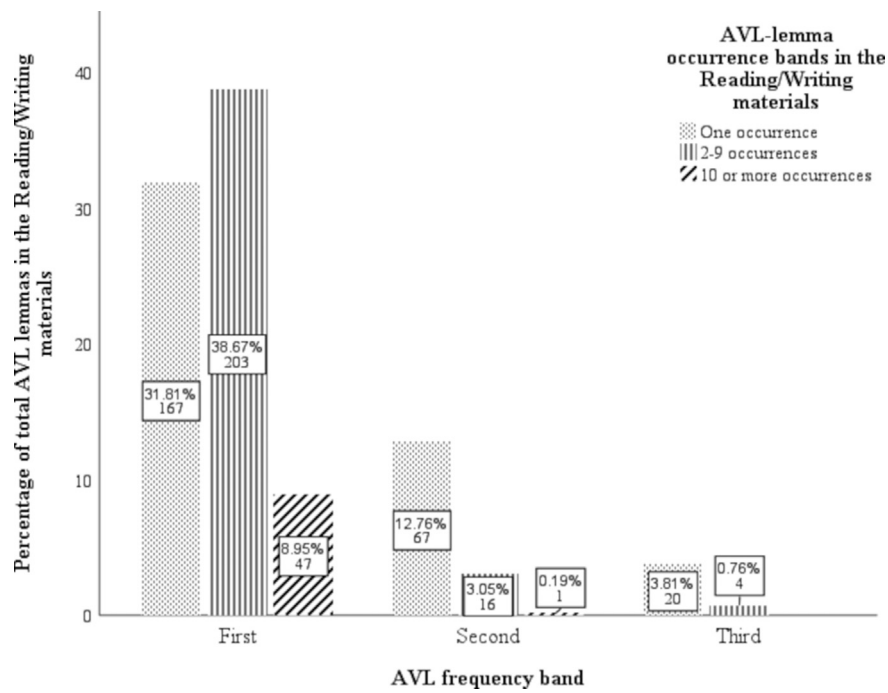


Figure 2. Percentage of AVL lemmas occurring once, 2–9 times and 10 or more times in the Reading/Writing materials per AVL frequency band.

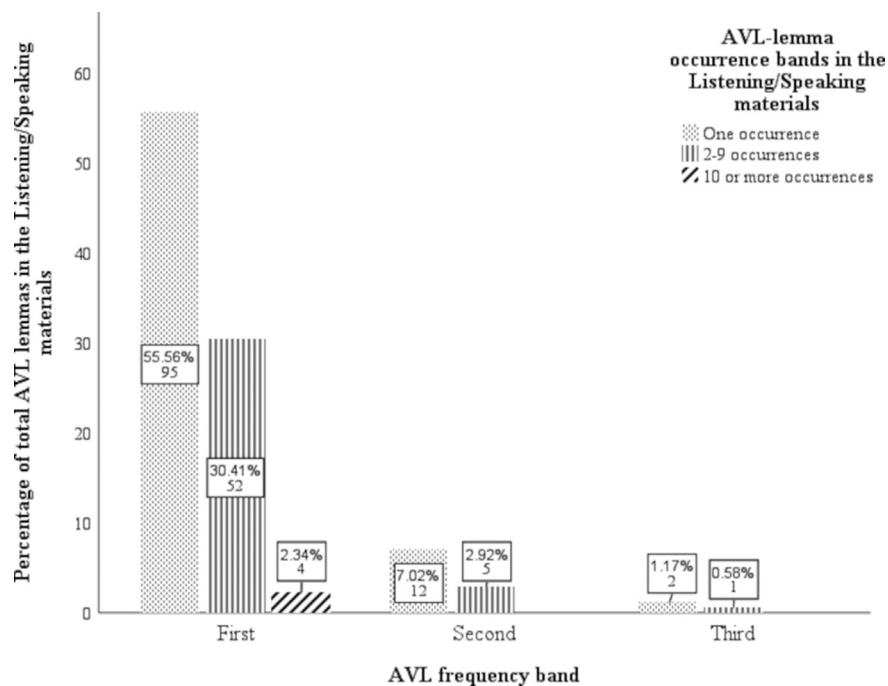


Figure 3. Percentage of AVL lemmas occurring once, 2–9 times and 10 or more times in the Listening/Speaking materials per AVL frequency band.

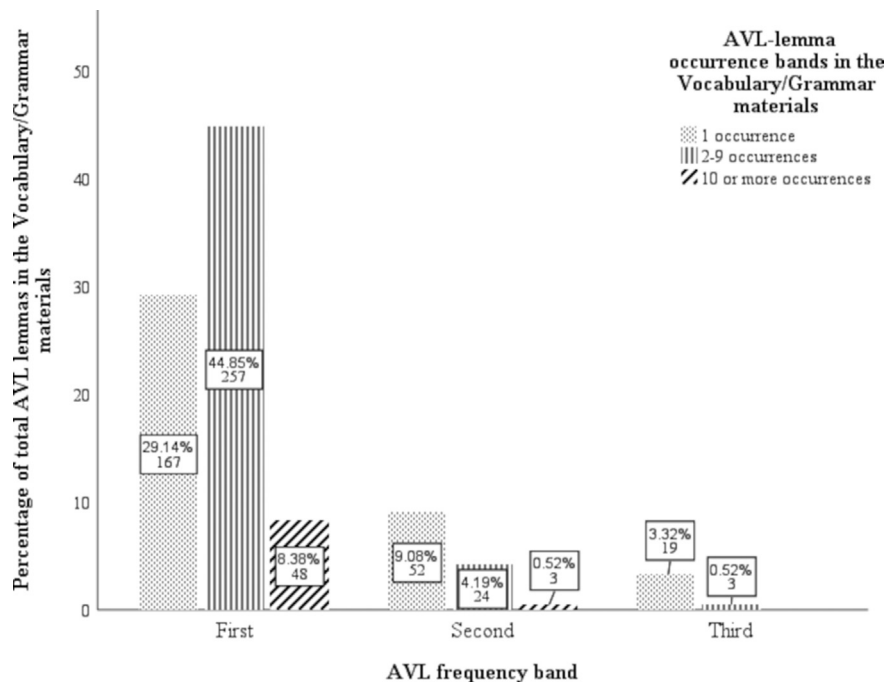


Figure 4. Percentage of AVL lemmas occurring once, 2–9 times and 10 or more times in the Vocabulary/Grammar teaching materials per AVL frequency band.

A comparison between Figures 2–4 on one hand and Figure 1 on the other indicates that the Reading/Writing and Vocabulary/Grammar subcorpora have the same pattern of AVL-lemma occurrence band frequencies as the whole corpus. In each of these subcorpora most of the AVL band-1 lemmas occur 2–9 times, the 2–9 occurrences band dominates overall (42.48% of the AVL-lemma occurrences in the Reading/Writing subcorpus and 49.56% of the AVL-lemma occurrences in the Vocabulary/Grammar subcorpus) and most lemmas from the other AVL bands occur only once. Conversely, in the Listening/Speaking subcorpus most AVL lemmas occur only once (63.75%) and single occurrences are predominant for lemmas from any AVL band.

As shown in Table 5, 90 lemmas were shared among all subcorpora. 62.22% appeared 10 or more times and the rest appeared 2–9 times. Conversely, most AVL lemmas shared by only two subcorpora – even when these were the large Reading/Writing and Vocabulary/Grammar subcorpora – occurred 2–9 times in total (see File 3 in the Supplementary materials). Therefore, for most AVL words, occurring in all three subcorpora was necessary to reach the 10-or-more-occurrences threshold.

4.3.2. AVL-in-BAWE lemma repetition in the materials

AVL-in-BAWE lemma repetition in the materials was examined because the number of times AVL-in-BAWE lemmas occur in EAP materials may affect students' ability to recall words. This ability in turn predicts performance not only in academic writing tasks but also in reading tasks (see section 1.3). This section reports on AVL-in-BAWE lemma occurrences in the materials as a whole and in each subcorpus, focusing specifically on the materials which aimed to foster the development of academic reading and writing skills. Descriptive statistics of the repetition rate of AVL-in-BAWE lemmas in the whole corpus and in the subcorpora are provided in File 4 in the Supplementary materials.

Figure 5 presents the percentage of AVL-in-BAWE lemmas which occurred less than or 10-or-more times in the corpus.

There are two patterns in Figure 5. First, as is the case for all AVL lemmas in the corpus (see Figure 1), AVL-in-BAWE lemmas that occurred 2–9 times form the majority (59.68%) of AVL-in-BAWE lemmas in the materials. Only about a quarter (25.62%) of all the AVL-in-BAWE lemmas in the materials occur 10 or more times. Second, there is a falling trend in the frequency of AVL-in-BAWE lemmas in the materials as we move from AVL lemmas shared across all 31 BAWE disciplinary subcorpora to those shared by 28 subcorpora; therefore, the higher the disciplinary range of an AVL lemma in BAWE, the more likely it was to be repeated in the materials.

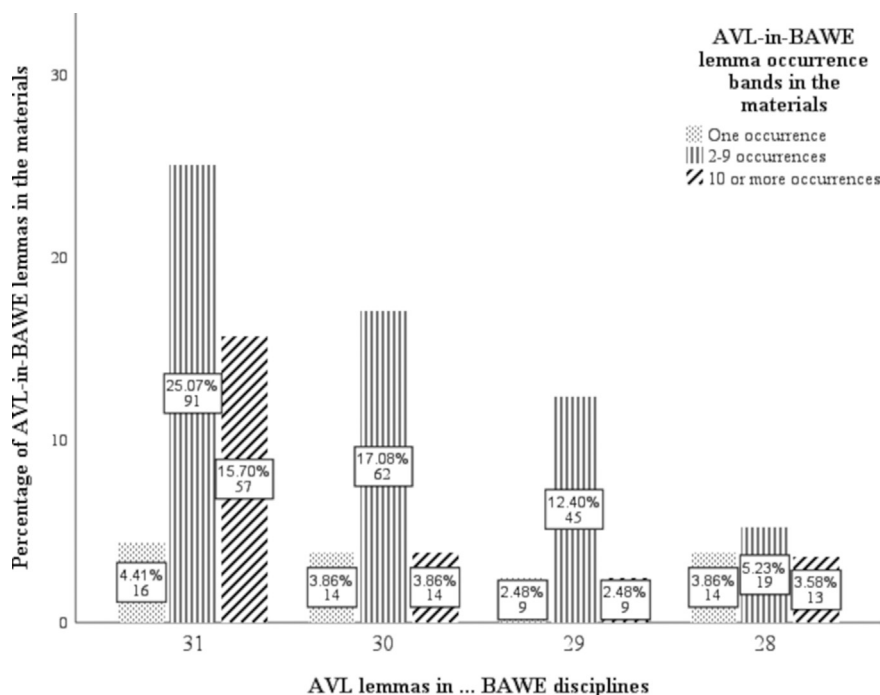


Figure 5. Percentage of AVL-in-BAWE lemmas occurring once, 2–9 times and 10 or more times in the materials per the number of BAWE disciplines which shared AVL lemmas.

Table 7 indicates that these two patterns exist also in each subcorpus, but in the Listening/Speaking subcorpus there is no appreciable difference in lemma frequency between AVL lemmas shared between 29 and 28 BAWE disciplinary subcorpora.

Table 7

Numbers and percentages of AVL-in-BAWE lemmas occurring once, 2–9 times and 10 or more times per teaching materials subcorpus and per the number of BAWE disciplines which shared AVL lemmas.

Subcorpus	Occurrences	AVL in BAWE disciplines			
		31	30	29	28
Reading/Writing	1	42 (15.16%)	20 (7.22%)	12 (4.33%)	10 (3.61%)
	2–9	82 (29.6%)	35 (12.64%)	25 (9.03%)	13 (4.69%)
	10 or more	21 (7.58%)	5 (1.81%)	6 (2.17%)	6 (2.17%)
Listening/Speaking	1	44 (38.6%)	15 (13.16%)	3 (2.63%)	4 (3.51%)
	2–9	27 (23.68%)	7 (6.14%)	5 (4.39%)	6 (5.26%)
	10 or more	2 (1.75%)	1 (0.88%)	0	0
Vocabulary/Grammar	1	32 (10.77%)	15 (5.05%)	12 (4.04%)	12 (4.04%)
	2–9	87 (29.29%)	47 (15.82%)	35 (11.78%)	18 (6.06%)
	10 or more	21 (7.07%)	7 (2.36%)	5 (1.68%)	6 (2.02%)

5. Discussion and conclusion

This study has examined general academic vocabulary occurrence and repetition rate in printed EAP course materials developed in-house, thus providing an initial insight into the academic-vocabulary exposure that students receive from such teaching materials.

In terms of the exposure to general academic vocabulary provided by the materials, nearly one third of AVL lemmas appear in the materials. Most of them are among the most frequent 1,000 AVL lemmas. This dominance of highly frequent AVL lemmas (e.g., *system*, *social*, *provide*, *however*, *include*) in the materials is evident in all module subcorpora and irrespective of whether AVL lemmas occurred in only one module subcorpus or were shared between/among subcorpora (see Table 5). This study also found that nearly all the AVL lemmas which commonly occur in the BAWE corpus occur in the materials.

These findings are encouraging because they indicate that even when EAP teachers do not specifically aim to include AVL lemmas when selecting and creating their printed materials (see Section 3.2), their in-house developed printed materials do include many. Further, the finding that the most frequent AVL lemmas are given prominence in the materials of all modules

indicates that the in-house materials examined in this study do not display the weaknesses often identified in textbooks, such as omitting features frequent in academic discourse or subject university textbooks while foregrounding infrequent ones (for a review of corpus-based textbook research, see [Harwood, 2014](#)).

The comparison of AVL coverage and lemmas across the subcorpora of the three modules in this EAP course indicated that the Vocabulary/Grammar printed materials included most AVL lemmas, closely followed by the Reading/Writing materials while their number in the Speaking/Listening materials was much lower. The lower number of AVL lemmas in the Speaking/Listening materials is to be expected because it is the smallest subcorpus; this small corpus size was, in turn, to be expected since the printed materials of a module aiming to develop students' listening and speaking skills were complemented with audio(visual) materials, which were not examined in this study. The higher number of AVL lemmas in the Vocabulary/Grammar subcorpus than in the Reading/Writing subcorpus is likely to be due to the different nature of the texts they contain; the brief texts and various activities in the Vocabulary/Grammar subcorpus are likely to expose students to more varied vocabulary than the full-length published articles and instructions to writing activities in the Reading/Writing subcorpus. As the main materials on these two modules were an EAP textbook (Vocabulary/Grammar) and research articles (Reading/Writing), the findings suggest that these types of materials may provide academic vocabulary input in complementary ways: while they contained similar numbers of AVL lemmas, their AVL coverage was significantly different, exposing students to different AVL density in texts. The fact that the different subcorpora complemented each other in terms of AVL exposure is also indicated by the small AVL overlap among subcorpora and the large number of lemmas unique to each subcorpus.

However, the low level of AVL-lemma overlap among subcorpora has impacted on the repetition rate of AVL lemmas in the printed materials of this course. Even when AVL lemmas were shared between the two largest subcorpora (Vocabulary/Grammar and Reading/Writing), the average repetition rate was below 10 occurrences, the repetition rate necessary for recall vocabulary knowledge to develop incidentally from reading (see [File 3 in the Supplementary materials](#)). Conversely, nearly two thirds of the AVL lemmas shared among all subcorpora appeared 10 or more times in the materials. The same pattern of findings appeared when we searched for AVL-in-BAWE lemmas in the materials. These findings mean that the repetition rate of AVL items was too low for recall vocabulary knowledge to develop incidentally for at least one third of the AVL words in the materials.

5.1. Pedagogical recommendations

Since EAP materials are typically developed in-house and further adapted and/or supplemented during the course in response to students' needs, as in this study, to help EAP practitioners make well-informed decisions regarding materials, they should be equipped with tools that can facilitate the selection of materials which include general academic vocabulary. Teachers can check whether the materials considered for inclusion contain academic vocabulary via freely accessible online tools such as Word and Phrase (<https://www.wordandphrase.info>). This software highlights all the AVL items in a text, provides information about their frequency based on the academic subcorpus of the COCA, and creates concordance lines for the words selected. When deciding which AVL items to focus on for teaching, teachers can use word-frequency information provided in Word and Phrase or consult [Durrant's \(2016\)](#) list of the 427 AVL lemmas commonly used in BAWE. These wordlists and tools should form part of EAP teacher training and development courses.

Since only a limited number of words can be taught explicitly during an intensive preessional course, exposure to academic vocabulary via materials leading to incidental learning is crucial. EAP practitioners should therefore aim to include materials containing a wide range of AVL lemmas. However, given the findings in this and previous studies about the low vocabulary repetition rate, tasks and activities that encourage vocabulary recycling should be included to maximise the vocabulary learning potential of materials and increase the number of encounters with target AVL lemmas. Examples include post-reading tasks requiring detailed re-reading of the texts through preparation of an oral or written summary, evaluation, comparison and commentary on the readings.

Students can also benefit from being familiarised with the AVL and the tools mentioned above so that they can check their own use of academic words in their writing. In addition, tasks involving the use of the AVL can be designed to encourage noticing and uptake. These types of tasks go particularly well with corpus-assisted and data-driven learning, encouraging students to independently exploit corpora for learning ([Jones & Durrant, 2010](#)). A small corpus of relevant texts can be created to support the course, with AVL-focused tasks designed by the teacher. Alternatively, students can be encouraged to create their own corpora of articles in their field or on a specific topic (see, e.g., [Charles, 2012](#)), which they can explore to complete AVL-focused tasks set by the teacher and report their findings to the class.

5.2. Limitations and implications for future research

As the first study to examine general academic vocabulary in in-house EAP teaching materials used in a preessional course, including those added by the teachers during the course, the present study is, necessarily, exploratory. It examines general academic vocabulary in printed only (not audio-visual as well) materials used in only one preessional EAP course. The exploratory nature of this study means that the generalisability of the findings remains to be tested in follow-up studies. In particular, since in-house materials are context-specific, research is needed to examine academic vocabulary in materials used in EAP courses at other universities. Moreover, to have a more complete picture of English academic vocabulary instruction in EAP courses, in addition to examinations of written teaching materials, such studies should explore the role of

academic vocabulary in teachers' materials selection and development and how teachers use these materials in the classroom. For the time being, EAP practitioners can estimate the relevance of our findings to their courses by comparing the context of our study (see section 3.1) to theirs.

In this study data were analysed quantitatively to identify AVL lemmas in the written materials and examine their repetition rate. Research into how many of these AVL lemmas were the focus of direct teaching in the EAP materials and what kinds of vocabulary knowledge (e.g., meaning, collocations, grammatical properties) were targeted in vocabulary activities is necessary to provide a more thorough examination of how well EAP teaching materials cater towards students' academic vocabulary needs.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.esp.2021.03.002>.

References

- Anthony, L. (2015). *TagAnt (Version 1.2.0)* [Computer Software]. Tokyo, Japan: Waseda University. Available from: <http://www.antlab.sci.waseda.ac.jp/>.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life. Robust vocabulary instruction*. New York: Guilford Press.
- Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it yourself corpus-building. *English for Specific Purposes*, 31, 93–102.
- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 67–85). Amsterdam: John Benjamins.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Criado, R. (2009). The distribution of the lexical component in ELT coursebooks and its suitability for vocabulary acquisition from a cognitive perspective. A case study. *International Journal of English Studies*, 9, 39–60.
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76.
- De Chazal, E., & Rogers, L. (2013). *Oxford EAP intermediate/B1+*. Oxford: Oxford University Press.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61.
- Eldridge, J. (2008). "No, there isn't an 'academic vocabulary', but...": A reader responds to K. Hyland and P. Tse's "is there an academic vocabulary?". *TESOL Quarterly*, 42, 109–113.
- Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6, 3–17.
- Feak, C. B., & Swales, J. M. (2010). Writing for publication: Corpus-informed materials for postdoctoral fellows in perinatology. In N. Harwood (Ed.), *English language teaching materials. Theory and practice* (pp. 279–300). Cambridge: Cambridge University Press.
- Gardner, D. (2013). *Exploring vocabulary: Language in action*. New York: Routledge.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35, 305–327.
- Gardner, D., & Davies, M. (2016). A response to "To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 62–68.
- Harwood, N. (2014). Content, consumption, and production: Three levels of textbook research. In N. Harwood (Ed.), *English language teaching textbooks: Content, consumption, production* (pp. 1–41). Basingstoke: Palgrave Macmillan.
- Jones, M., & Durrant, P. (2010). Corpora and vocabulary teaching materials. In A. O'Keefe, & M. McCarthy (Eds.), *The Routledge Handbook of corpus linguistics* (pp. 387–400). Abingdon: Routledge.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44–57.
- Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22, 56–70.
- McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, 37, 389–411.
- Miller, D. (2011). ESL reading textbooks vs. university textbooks: Are we giving our students the input they may need? *Journal of English for Academic Purposes*, 10, 32–46.
- Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25, 235–256.
- Nation, P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. New York: Continuum.
- Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *Tesol Quarterly*, 48, 843–861.
- Spencer, S., Clegg, J., Lowe, H., & Stackhouse, J. (2017). Increasing adolescents' depth of understanding of cross-curriculum words: An intervention study. *International Journal of Language & Communication Disorders*, 52, 652–668.
- Stoller, F. L. (2016). EAP materials and tasks. In K. Hyland, & P. Shaw (Eds.), *The routledge handbook of English for academic purposes* (pp. 577–591). Abingdon: Routledge.
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69, 559–599.
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37, 461–469.
- 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, 130–163.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65.

Sophia Skoufaki is Associate Supervisor at the University of Essex. She holds a PhD from the University of Cambridge. She specialises in second language vocabulary learning and teaching. Her other interest is written discourse coherence.

Bojana Petrić is Reader at Birkbeck, University of London. She has co-authored *Experiencing Master's supervision: Perspectives of international students and their supervisors* (Routledge, 2017) and has published articles on academic writing.