

Article

Exploring Potential Factors Affecting Reading Comprehension in EAL Learners: A Preliminary Corpus-Based Analysis

Gavin Brooks ^{1,*} , Jon Clenton ²  and Simon Fraser ³¹ Faculty of Foreign Studies, Kyoto Sangyo University, Kita-ku, Kamigamo, Motoyama, Kyoto 603-8555, Japan² Graduate School of Humanities and Social Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan; jclenton@hiroshima-u.ac.jp³ Institute for Foreign Language Research and Education, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan; fraser@hiroshima-u.ac.jp

* Correspondence: gavinbrooks@gmail.com

Abstract: This article presents a study examining the vocabulary knowledge of English as an additional language (EAL) learners in two international schools in Japan in relation to the vocabulary profiles of the textbooks they are required to use in the classrooms. The vocabulary knowledge of 139 participants from two international schools was assessed using either the New Vocabulary Levels Test (NVLT) or the Updated Vocabulary Levels Test (UVLT). These results were compared to a 15 million-word corpus compiled from representative subject-specific textbooks to estimate the vocabulary coverage participants are likely to achieve. The findings revealed that EAL learners consistently scored lower than a combined group of their first-language English (FLE) and proficient L2 (PL2) peers, with fewer than 25% of EAL learners mastering the AWL before Grade 12. Furthermore, even the most frequent 5000-word bands provided only 91–93% coverage for subjects like biology and chemistry, leaving many EAL learners struggling to comprehend these texts. This analysis highlights the potential difficulties EAL learners may face in understanding the textbooks that are being used in EAL classrooms, underscoring the need for better vocabulary scaffolding and support for such learners in the international school context.



Academic Editor: Abdessattar Mahfoudhi

Received: 9 July 2024

Revised: 25 November 2024

Accepted: 29 January 2025

Published: 10 February 2025

Citation: Brooks, G., Clenton, J., & Fraser, S. (2025). Exploring Potential Factors Affecting Reading Comprehension in EAL Learners: A Preliminary Corpus-Based Analysis. *Languages*, 10(2), 30. <https://doi.org/10.3390/languages10020030>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: English as an additional language; English-medium instruction; vocabulary; corpus; coverage

1. Introduction

The growth of English as an additional language (EAL) education worldwide (observed by, e.g., [Smith, 2015](#)) has led to an increase in EAL learners in countries such as the UK ([Murphy & Unthiah, 2015](#)) and the US ([US Department of Education, National Center for Education Statistics, 2022](#)), highlighting the need for effective EAL language instruction. Previous studies that have focused on the needs of EAL learners ([Hessel & Murphy, 2019](#); [Spencer et al., 2017](#); [Townsend et al., 2016](#)) have shown that English academic vocabulary is one vital area where teachers can support EAL learners. While there have been fewer studies conducted in the international school context, existing research indicates that EAL learners in this context also often lack the vocabulary skills to engage with the academic material in their classes ([Coxhead & Boutorwick, 2018](#)).

The issue of providing research-backed support for learners studying in an English as a medium of instruction (EMI) context in Japan is urgent, given the significant increase in the number of international schools operating in the country. Globally, International Baccalaureate (IB) schools experienced a 33.3% increase from 2016 to 2022, with Japan

seeing commensurate levels of growth during this period (International Baccalaureate, 2023). International schools in Japan often comprise learners from diverse L1 backgrounds with varying degrees of English language proficiency (Brooks et al., 2021), making it challenging for teachers to support EAL learners in the classroom. In international schools, the need for this support becomes even more crucial. This is because the challenges posed by learners are compounded by the fact that they speak multiple languages, and some are required to learn both English and the language of the wider community where the school is located (Carder, 2007). Consequently, learning English becomes significantly more complex in this context. The English language proficiency of EAL learners can be greatly affected by this, causing them to struggle to understand classroom materials in Japan (Brooks, 2023).

To better support the needs of EAL learners in the classroom, it is necessary to understand both the range of vocabulary with which EAL learners are likely to be familiar across different educational stages as well as the vocabulary composition of academic materials these students are expected to engage with during their studies. This study investigates the gaps in vocabulary knowledge that EAL learners studying in the International School context may demonstrate across grade levels and subjects. To explore EAL learners' vocabulary proficiency across various grade levels, we compared the vocabulary knowledge of two cohorts of learners enrolled in the international school context in Japan against the vocabulary profiles of the textbooks prescribed for their classes.

2. Literature Review

2.1. The Importance of Vocabulary for EAL Learners

While not a homogeneous group, EAL learners are typically defined as students who study in classrooms where English is the medium of instruction and who speak a language other than English at home (Murphy, 2014; Murphy & Unthiah, 2015). Sharples (2021) highlights two key implications from this definition: (i) the mainstream classroom is the primary place of learning for these students, and (ii) for them, language learning is inseparable from content learning. Additionally, research indicates that there are numerous factors which can hinder the academic success of EAL learners in EMI classrooms (Afitska & Heaton, 2019). This includes a considerable number of studies that suggest language difficulties play a critical role in their academic struggles (e.g., Clegg & Afitska, 2011; Coxhead & Boutorwick, 2018). In short, it is crucial for EAL learners to receive sufficient language support in the classroom. Without this support, they can face challenges in both their academic and linguistic development.

One area of language that research has consistently shown EAL learners struggle with is vocabulary knowledge (Faitaki et al., 2022). This is significant because research also shows that vocabulary is a key predictor of reading comprehension, not only for EAL learners (e.g., Brooks et al., 2021; Melby-Lervåg & Lervåg, 2014) but also for first-language English (FLE) learners (e.g., Ouellette & Beers, 2010; Tunmer & Chapman, 2012). Research shows (NALDIC, 2015) that EAL learners typically enter the educational system with lower levels of vocabulary knowledge than FLE learners studying at the same grade level. EAL learners have also been shown to take longer to learn and master the vocabulary they need for the classroom (Coxhead & Boutorwick, 2018) and, as a result, often have lower vocabulary levels than their FLE classmates across all grade levels (e.g., August et al., 2005; Murphy & Unthiah, 2015). Limited vocabulary knowledge can pose difficulties for EAL learners when it comes to understanding what they read, particularly regarding the textbooks they are assigned to read for their courses (Brooks et al., 2021; Marianne & Coxhead, 2023). Leung (2014) notes that textbooks can be particularly challenging for EAL learners because they often make use of precise vocabulary and specialised terminology and discuss unfamiliar

content. While the exact amount of vocabulary EAL learners require for academic success in an EMI setting is uncertain, studies indicate that two types of vocabulary knowledge are crucial for this group. These include both general academic vocabulary as well as high-frequency words (Coxhead & Boutorwick, 2018).

2.2. Supporting EAL Learners in the Classroom

To ensure the success of EAL learners in the classroom, it is crucial that they understand the texts they are required to read. Due to the multitude of skills learners need to master to comprehend texts in English, reading comprehension can be a difficult skill to assess (Melby-Lervåg & Lervåg, 2014). Factors that can impact a learner's ability to understand a text include prior linguistic knowledge (Droop & Verhoeven, 2003) as well as reading ability in their L1 (Chuang et al., 2012). Researchers have also found a significant relationship between learners' phonemic skills and their reading comprehension (Melby-Lervåg et al., 2012). While other factors are undoubtedly important for comprehending classroom textbooks, as discussed above, vocabulary knowledge has consistently been shown to be one of the key predictors of reading comprehension (Brooks et al., 2021). Consequently, it is important that teachers provide EAL learners with the vocabulary support necessary to comprehend the texts they are being assigned to read.

Multiple researchers have highlighted a correlation between vocabulary knowledge and reading comprehension (Brooks et al., 2021; Laufer, 1989; Schmitt et al., 2011). These same studies provide evidence that there is a threshold of coverage that is necessary for textual understanding (Laufer & Ravenhorst-Kalovski, 2010). Put simply, if learners cannot comprehend above a minimum threshold of the vocabulary in a text, they will struggle to grasp its overall meaning. It is frequently cited that comprehending approximately 95% of the vocabulary is necessary for learners to have an adequate understanding of what they are reading (Laufer, 1989). Students who are familiar with 95% of the words in a text may still need assistance with unfamiliar vocabulary while reading, such as asking a fellow student or teacher for help or referring to a dictionary when needed. The likelihood of learners being able to read a text independently increases if they are familiar with 98% of the words. Although having a wide range of vocabulary (for example, knowledge of sufficient vocabulary to provide 98% to 100% coverage of the words in a text) does not guarantee complete understanding, research consistently shows that an understanding of over 95% of the words in a text is a necessary, if not sufficient, condition for understanding (e.g., Kremmel et al., 2023; Schmitt et al., 2011). Of course, the question then becomes, what words do learners need to know to be able to understand a certain text? To effectively provide the type of vocabulary support that EAL learners need, it is essential to have a clearer understanding of the words they require to comprehend the reading material they are given.

2.3. Investigating Vocabulary in Learning Materials

Word lists are one important tool that teachers can use to help them identify the vocabulary that would be most pertinent to a specific group of learners or a given learning context. The most common type of word lists being used in the classroom are frequency-based lists such as university-level academic word lists (e.g., the Academic Word List (AWL; Coxhead, 2000) and the Academic Vocabulary List (AVL; Gardner & Davies, 2014)), middle school word lists (e.g., the Middle School Vocabulary List (MSVL; Greene & Coxhead, 2015)), and word lists for specialised subjects such as medicine and pharmacology (e.g., the Pharmacology Vocabulary List (PVL; Fraser, 2010) and the Medical Academic Vocabulary List (MAVL; Lei & Liu, 2016)). Frequency is a useful criterion on which to base a word list because it allows educators and researchers to judge how likely learners are to encounter a

given word, and it has been shown to be a good proxy for word difficulty (Benjamin, 2011; De Clercq & Hoste, 2016; Hancke et al., 2012). Studies support the use of frequency as a measure of word difficulty. For example, Brysbaert et al. (2011) found that word frequency was the most important factor for determining difficulty and, along with similarity to other words and word length, accounted for 40.5% of the variance in lexical decision times, whereas all of the other variables looked at in the study accounted for less than 2%.

A commonly employed technique for arranging frequency-based word lists is to divide the vocabulary into bins or groups based on frequency, usually utilising bins of 1000 words (Laufer & Nation, 1995). Examples of such an approach can be seen in popular word lists, such as the BNC/COCA (Nation, 2020) and the JACET 8000 list (Uemura & Ishikawa, 2004). Directional binning approaches work well for general word lists because they usually clearly indicate what the most important high-frequency words are for beginning learners (Flor et al., 2024). However, depending on the corpus being used, the coverage of such lists may not be as effective for mid- to low-frequency or domain-specific and technical words (Dang, 2019). These words are of particular importance to EAL learners as they often comprise the words learners need to know to be able to understand the topics they are learning about in their textbooks (Marianne & Coxhead, 2023). However, because there are no EAL-specific word lists, EAL teachers often need to rely on word lists compiled for FLE learners (e.g., Living Word Vocabulary, Dale & O'Rourke, 1976; EDL Core Vocabulary, Taylor et al., 1989), academic word lists designed for university learners (e.g., the AWL, Coxhead, 2000), or general word lists (e.g., the General Service List (GSL; West, 1953)) in the classroom. To determine how effective such lists will be at providing support in the EAL context, it is first necessary to better understand what coverage they provide over textbooks EAL learners are being asked to read, as well as how many new words learners need to acquire to achieve this coverage.

Researchers have used corpora to identify the types of words learners need to engage with domain-specific texts in various contexts (Coxhead, 2017). For example, Coxhead et al. (2016) asked students what words were necessary to study carpentry and compared their responses with technical terms extracted from a corpus of written carpentry texts, emphasising the importance of specialised vocabulary for academic success. Similarly, Miller and Biber (2015) outlined techniques for developing subject-specific word lists using a corpus of psychology textbooks. Additionally, Coxhead (2012) explored how corpora can aid educators and researchers in identifying the common words and phrases high school students are likely to encounter in their textbooks, demonstrating the value of corpus analysis in the educational context.

2.4. Vocabulary in Textbooks

One important area of research related to supporting learners in the classroom is analysing the vocabulary load of textbooks to evaluate whether the vocabulary aligns with learners' needs and to determine the comprehensibility of texts for students in different settings. While there are different ways of approaching how to analyse a textbook, one of the most common is to look at the lexical coverage of the text (Sun & Dang, 2020). Vocabulary load is linked to coverage, meaning the percentage of words learners are likely to know, which affects their ability to comprehend the material (Nation, 2005; Webb & Nation, 2008). Most of this research has focused on textbooks that international publishers have produced for use in English as a Foreign Language (EFL) classrooms (Benson & Madarbakus-Ring, 2021; Hsu, 2009; Matsuoka & Hirsh, 2010; O'Loughlin, 2012).

While the lexical loads of textbooks used in the university EFL classroom can be expected to vary across levels, more recent studies have provided support for the findings of Matsuoka and Hirsh (2010) that 95% coverage of EFL textbooks can be achieved with around

two to three thousand-word families. In the Japanese context, [Benson and Madarbakus-Ring \(2021\)](#) found that 95% coverage of a commercial textbook used at the university level in Japan could be achieved with the first 3000 words of the BNC/COCA. [Yang and Coxhead \(2022\)](#) looked at textbooks in use in China and found that 95% coverage was provided by the first 4000 words of the BNC/COCA. [Hsu \(2009\)](#) examined 36 different commercially available textbooks used in classrooms in Taiwan and found that learners would need between 2500 to 13,000-word families to achieve 95% coverage of the texts. This highlights the inconsistencies between the lexical demands. These findings are important, as they show that, while mastery of the first 3000 words would be sufficient for understanding most EFL university textbooks, this is not true in all cases. Teachers should take their learner's lexical proficiency into account when selecting a textbook.

High school textbooks in EFL contexts have been studied to determine their vocabulary load and how well they align with students' knowledge. [Nguyen \(2021\)](#) found that high school students in Vietnam required 3000–5000-word families to comprehend most of their textbooks. [Sun and Dang \(2020\)](#) analysed Chinese high school textbooks, showing students needed up to 9000-word families to reach 98% coverage. In both studies, the vocabulary knowledge of the learners was compared to the vocabulary loads of the textbooks; both found that most, if not all, of the learners would be unlikely to know enough words to achieve 95% coverage. These findings highlight a significant gap between learners' vocabulary knowledge and the lexical demands of their textbooks.

While these studies do provide us with some insight into the types of vocabulary learners need to know in an EFL setting, it is important to remember the way that textbooks are used in an EMI or CLIL classroom is different from how they are used in an EFL classroom ([Sun & Dang, 2020](#)). Where the focus of the lesson is the content and not the language, the vocabulary demands of the textbooks differ as EFL classes tend to focus on basic interpersonal communicative skills (BICS). In contrast, EMI classes focus on cognitive academic language proficiency (CALP, Cummins, 1979). In EFL contexts, the goal of the textbook is to expose learners to everyday language (BICS) and provide support for activities and discussions carried out in the classroom. The textbooks used in EMI and content-based classes, on the other hand, use academic and subject-specific language, and learners need to understand the content of the textbook to answer domain-specific questions on exams and quizzes. This is supported by studies that have shown the degree to which students studying in an EMI context outperform their EFL counterparts is higher for academic vocabulary than for everyday language due to their exposure to academic and subject-specific language in both their textbooks and the classroom ([Castellano-Risco et al., 2020](#)).

Outside of the EFL context, there have been a number of studies that have looked at the relationship between vocabulary and textbooks in EMI classrooms, both in English-speaking countries (e.g., [Coxhead et al., 2015](#); [Luxton et al., 2017](#)) and in the international school settings ([Coxhead & Boutorwick, 2018](#)). There have also been a number of studies that have focused specifically on an analysis of the vocabulary found in textbooks written for content-based classes at both the junior high school ([Greene & Coxhead, 2015](#)) and high school levels ([Coxhead & Boutorwick, 2018](#); [Coxhead et al., 2010](#); [Green & Lambert, 2018](#)). Unlike the EFL studies discussed above, these studies focus on subject-specific textbooks written in English, such as those for science, mathematics, and literature, rather than textbooks that were written specifically for the purpose of learning English. [Coxhead et al. \(2010\)](#), for example, examined subject-specific textbooks (e.g., science textbooks in New Zealand secondary schools) and found that learners needed over 4000-word families for 95% coverage and over 14,000-word families for 98%. However, corpus-based studies that focused specifically on EAL learners and the unique vocabulary challenges they face remain under-explored in the literature. Expanding research in these areas would

provide a clearer understanding of the specific vocabulary demands and learning processes in EAL settings, ultimately contributing to more effective language instruction.

3. The Current Study

This study expands on prior research by examining the vocabulary knowledge of EAL learners in an international context in Japan and comparing it to the vocabulary found in the textbooks these learners use in class using the BNC/COCA word list (Nation, 2020). The aim is to better understand the vocabulary necessary for learners to comprehend these textbooks and assess the coverage provided by their current vocabulary knowledge. The specific goal is to gain insight into the vocabulary used in these textbooks to identify potential challenges EAL learners may face while reading them. As such, this study addresses four research questions:

1. To what extent does the vocabulary knowledge of English as an additional language (EAL) and first-language English (FLE) learners studying in an international school context vary across different grade levels?
2. To what extent does the vocabulary in subject-specific textbooks at an international school align with BNC/COCA frequency bands, and how does this alignment vary across International Baccalaureate (IB) subjects?
3. To what extent can EAL learners be expected to comprehend the vocabulary encountered in textbooks used for different IB subjects?
4. Using the BNC/COCA, how much additional vocabulary would EAL learners need to acquire to understand the language used in textbooks they would typically encounter within the international school context?

4. Methodology

4.1. Design

The study compared the vocabulary knowledge of EAL and FLE learners across seven different grade levels. A total of 139 participants were included in the final analysis, representing a diverse linguistic background. Each group's vocabulary knowledge was assessed using one of two versions of the Vocabulary Levels Test (VLT). The scores they received on the VLT were compared to an analysis of the vocabulary coverage in the textbooks used by the participants in their International Baccalaureate (IB) subjects.

The independent variables in the analysis were learner type (with two levels: (1) FLE and PL2 learners, and (2) EAL learners) and school level (with two levels: pre-IB, which covers Grades 6 to 9, and IB, covering Grades 10 to 12). The dependent variable was vocabulary knowledge, as measured by the VLT scores across the first five 1000-word bands. The analysis included an interaction term to examine whether the difference in vocabulary knowledge between FLE/PL2 and EAL learners varied across school levels.

The corpus analysis focused on five subject areas (literature, maths, physics, chemistry, and biology), examining the frequency and coverage of vocabulary in textbooks used in these subjects. Vocabulary comprehension was measured in terms of how well the learners' existing knowledge aligned with the vocabulary demands of these textbooks.

4.2. Participants

The study took place at two different International Schools in Japan. A total of 142 participants ($N = 142$) from diverse linguistic and cultural backgrounds participated in the study. While Japanese and English were the two most common languages spoken, 22 different languages were represented in our dataset. We used a Rasch analysis (Beglar, 2010) to evaluate participant scores on the uVLT and the NVLT. The majority of VLT scores demonstrated a good fit for the Rasch model. However, we removed three learners with

very high outfit scores ($Z_{std} > 8.76$) from the dataset. After the three participants were excluded from the final analysis, the final dataset consisted of 139 participants (68 male and 71 female learners ranging in age from 13 to 18 years old).

Following the procedure used in previous studies (Coxhead & Boutorwick, 2018), we categorised the participants as either FLE, Proficient L2 (PL2) learners, or EAL learners based on nationality, time spent in English-speaking countries, languages spoken at home, and teacher assessments of English proficiency. For this study, EAL learners were defined as those who did not speak English at home and required additional language support in the classroom. FLE learners were participants who were both proficient in English and either spoke English predominantly at home or had significant exposure to English-speaking environments. PL2 learners were proficient in English but primarily spoke a language other than English at home. The term PL2 was chosen over non-native speaker (NNS) to avoid the deficit model associated with native and non-native speaker labels. Table 1 details how many EAL, PL2, and FLE participants there were at each grade level.

Table 1. Breakdown of FLE, PL2, and EAL participants by grade.

Grade	6th	7th	8th	9th	10th	11th	12th	Total
FLE	0	3	2	1	1	3	1	11
PL2	2	2	4	6	3	6	2	25
EAL	4	13	15	17	20	28	6	103
Total	6	18	21	24	24	37	9	139

Given that our study was primarily concerned with the vocabulary sizes of EAL learners and investigating the relationship between their vocabulary knowledge and their ability to engage with and comprehend classroom readings, as highlighted in previous studies (Coxhead & Boutorwick, 2018; Marianne & Coxhead, 2023; Murphy & Unthiah, 2015), we made the decision to group the FLE and PL2 learners for the purpose of analysis. This was based on several factors. Firstly, this approach mirrored that of earlier studies (Brooks, 2023; Coxhead & Boutorwick, 2018) that have focused on EAL learners and compared their vocabulary knowledge and progress with those of their FLE and PL2 peers. Secondly, teachers considered both groups to be similar in the classroom context, and additional language support, such as pull-out classes, was reserved exclusively for EAL learners. Thirdly, demarcating between FLE and PL2 was challenging as the responses they gave to the questions on the surveys were very similar. For example, a Grade 10 student who had lived in the US for the first eight years of his life but spoke Japanese at home selected Japanese as his first language, while another learner who had bilingual parents but spoke predominantly English at home and had never lived overseas identified English as her L1. Despite initially being in different groups, the learners from the FLE and PL2 groups shared the most similarities in both their responses to the survey questions as well as how their language skills in the classroom were evaluated by their teachers. Given all these factors, it made sense to combine the FLE and PL2 learners, allowing us to focus on comparing the EAL learners with the rest of the cohort more effectively.

4.3. Materials

4.3.1. The Vocabulary Levels Tests

The learners' vocabulary knowledge was measured using the Vocabulary Levels Test (VLT). The VLT is a standardised assessment instrument designed to evaluate a learner's vocabulary knowledge across different frequency bands to gain a deeper understanding of their language proficiency (Schmitt et al., 2001). In this study, the participants were

given one of two different versions of the VLT test. Both tests cover the first five 1000-word bands of the BNC/COCA. These tests were chosen over previous VLTs (e.g., Nation, 1983; Schmitt et al., 2001) because they cover a greater number of frequency bands and are based on a more recently compiled set of word lists, the BNC/COCA (Nation, 2020). In total, 49 participants (31 EAL learners, 6 FLE learners, and 12 PL2 learners) took the new Vocabulary Levels Test (NVLT) and 90 participants (72 EAL learners, 5 FLE learners, and 13 PL2 learners) took the updated Vocabulary Level Test (uVLT).

We administered the new Vocabulary Levels Test (McLean & Kramer, 2016) to the first group. This vocabulary assessment tool is a multiple-choice test consisting of 24 questions for each of the first five bands. The test also includes an academic vocabulary section based on Coxhead’s (2000) AWL. Each of the individual items consists of the target word, which is given by itself and in context, along with four possible responses. The examinee has to select the correct response, which can be either a single word or a phrase closest in meaning to the target word (see Figure 1).

- 1. time: They have a lot of **time**.
- a. money
- b. food
- c. hours
- d. friends

Figure 1. A sample question from the NVLT. Adapted from McLean and Kramer (2016), page 4.

We administered Webb et al.’s (2017) updated Vocabulary Levels Test to the second group of participants. As with the NVLT, this test measures knowledge of the first five 1000-word bands. The uVLT also requires examinees to match a word with its definition or explanation. However, it differs from the NVLT in that the items are not presented individually but grouped into clusters, with each cluster containing three related words. There are 10 clusters for each of the first five 1000-word bands. The test taker must match each word to its corresponding definition or explanation (see Figure 2). The uVLT includes a representative portion of nouns, verbs, and adjectives (15, 9, and 6 items per level, respectively) selected from each of Nation’s (2020) BNC/COCA bands. Because the uVLT does not include a section for Coxhead’s (2000) Academic Word List, we supplemented the test with the AWL section of the NVLT.

	eye	father	year	van	voice	night
body part that sees	✓					
parent who is a man		✓				
part of the day with no sun						✓

Figure 2. A sample question from uVLT, along with the answers provided. Adapted from Webb et al. (2017), page 61.

We understand that it would have been better to use the same assessment tools for both groups of learners. However, the assessments were conducted as part of two larger studies (e.g., Brooks et al., 2021) and, in the context of this paper, we felt that it was important to include the data from both assessments as that allows us to examine larger trends within the population. The tests serve as a tool for assessing the learner’s comprehension of the frequency band, and the effectiveness of both assessments in demonstrating the mastery of frequency bands by learners has been proven in studies conducted by the authors of each assessment tool (McLean & Kramer, 2016; Webb et al., 2017) and have been used in this capacity by other researchers (e.g., Ha, 2021; Kremmel et al., 2023; Xodabande &

Hashemi, 2023). It is crucial to clarify that these scores are not being compared to any other scores or utilised as a measure of linguistic proficiency but are being used to provide us with a picture of the level of vocabulary knowledge EAL learners in this context are likely to possess. Given this, we feel that any potential variance between the tests would not invalidate the picture they are able to give us of the average level of vocabulary knowledge of the different groups of students in this study. Although there may be a slight variation between the tests, the benefits of gathering information about the vocabulary knowledge of a larger number of students in diverse settings outweigh any potential drawbacks related to differences between the assessments themselves.

4.3.2. The Corpus

To create a corpus representative of the textbooks used by learners in the classroom, we compiled a set of domain-specific corpora sourced from a variety of textbooks taken from across the different subjects the participants were studying at the two schools. The corpus comprises five subjects (Table 2): literature, maths, physics, chemistry, and biology. The corpus focuses on the textbooks used during the IB diploma and does not include the handouts and textbooks the participants were using at the middle school level. The rationale for this was two-fold. Initially, these textbooks aided in building a strong vocabulary foundation for EAL learners in the IB program. Secondly, teachers at the middle school level used handouts and teacher-created materials more frequently in the classroom rather than relying on prescribed textbooks. While there were no middle-school textbooks in the corpus, we feel that the vocabulary in the corpus is relevant to these grades because, in the IB setting, one of the primary goals of the middle-school grades is to prepare students for the IB program. Part of this preparation involves having students learn the vocabulary that they need to succeed in the IB context.

Table 2. Total number of texts and tokens for each subject.

Subjects	Total Textbooks	Total Tokens
Literature	7	1,122,212
Maths	8	942,784
Physics	5	1,135,845
Chemistry	5	1,315,730
Biology	5	941,215
Total	30	5,457,786

We digitised the textbooks by scanning them into the computer, then meticulously reviewed and refined the scanned texts to rectify any errors that might have occurred during the scanning process. This thorough process was conducted using a text editor, following the procedures suggested by Nation (2016). We utilised R and Excel to deal with errors not addressed during the initial cleaning phase. All of the errors that resulted in non-words were cleaned by exporting a csv of all the off-list words and manually checking those against the original PDFs. This was carried out by the primary researcher and a group of research assistants. However, despite the extensive data cleaning, it was not possible to completely eliminate all noise from such a large corpus, and some OCR (optical character recognition) errors remain, such as those with mathematical and chemical formulas, with numbers such as “1” being recognised as an “l”, or with punctuation. Although this may restrict the use of data to analyse features of the text such as paragraph length or mean sentence length, the focus on removing incorrectly scanned words means that the impact of this noise on vocabulary measures was minimal.

4.4. Procedures

As noted above, the participants for the study were recruited from two international schools in Japan. This was carried out through a collaboration with the EAL teachers and the principals at both schools, which helped to facilitate the recruitment process. All students in grades 6 through 11 were administered the Vocabulary Levels Tests (VLTs) as part of their regular classroom activities. Due to time constraints, not all grade 12 students were able to participate, as they were heavily engaged in preparing for their International Baccalaureate (IB) exams and final assessments.

At the start of the project, we obtained consent from both the participants and their parents or guardians. This process was carried out collaboratively with the research team, the students, the parents, and the school administration. Consent was required before the administration of the VLT, and only students from whom consent was received were included in the final dataset.

The VLT was administered in the classroom by the students' regular teachers. The test was paper-based, and students were given one hour to complete the VLT along with a short survey on their language background. All participants finished within the allotted time.

At the same time they took the VLT, participants also filled in a short survey which included questions on nationality, languages spoken at home, and time spent overseas. Additionally, students were asked to self-identify their first language (L1), with the option to select multiple L1s, acknowledging the number of bilingual and multilingual learners in the study.

5. Analysis

5.1. Analysing the VLTs

Following previous studies (Coxhead & Boutorwick, 2018), participants who scored above 86% on the words from a single frequency band were considered to have mastered that frequency band. The number of participants at each grade level that achieved mastery was used to determine how likely participants were to be able to read the required textbooks for each of the IB subjects. The overall VLT scores were used for the between-group comparisons at each grade level.

5.2. Analysing the Corpus

The texts that made up the corpus were initially imported and preprocessed in R (v4.2.1; R Core Team, 2021). After this, we lemmatised and identified proper nouns in the corpus using Python in conjunction with spaCy, a widely used natural language processing library that allows for efficient and accurate linguistic annotations, including tokenisation, part-of-speech tagging, lemmatisation, and named entity recognition. Following these processing steps, the annotated data were exported as a CSV file. The final analysis of the lemmatised corpus was conducted in R, utilising the tidyverse package.

Once we had a cleaned and lemmatised version of the corpus, we used a pre-prepared data frame created from Nation's (2020) BNC/COCA word family lists (from 1000 to 25,000) to determine the frequencies of word families in the text. During this stage, we also made use of four supplementary lists from Nation (2020) to identify proper nouns (e.g., Europe, David), abbreviations (e.g., pm, rpm, kwh), compounds (e.g., birthday, airbag), and marginal words (e.g., ah, mm, X, Y). These lists were supplemented with the list of proper nouns that spaCy had identified in the corpus during the previous stage.

In the next phase of our research, we further analysed the off-list words to ensure that all abbreviations, marginal words, and proper nouns from the text were accurately added to the appropriate supplementary list. To achieve this, we checked each of the off-list words against the original PDF to verify that the OCR had been carried out correctly on

that word and that there were no errors in the final text files. Any errors that we identified were manually corrected. Words that were correct and were either abbreviations, marginal words, or proper nouns were added to the appropriate list. After these amendments, we used the updated texts and lists to rerun our analysis. We then used this final analysis to determine what coverage the different frequency bands of the BNC/COCA could provide over the five subcorpora.

6. Results

Research Question 1: To what extent does the vocabulary knowledge of English as an additional language (EAL) and first-language English (FLE) learners in an international school context vary across different grade levels?

An examination of the VLT scores (Table 3) shows that the majority of the participants were not able to attain mastery of the AWL or the BNC/COCA mid-frequency bands. These findings reflect previous studies' findings (e.g., Coxhead & Boutorwick, 2018). Additionally, we noticed that a considerable portion of the participants struggled with learning the common high-frequency words from the 2000 range before reaching Grade 9, and less than half of both groups were able to grasp the 3000-word levels before Grade 12. The mastery of the AWL also posed a significant challenge for learners at all grade levels, with fewer than 25% of the participants being able to master these words until the 12th grade. Because of how important the general academic words found on the AWL are for comprehending academic texts such as school textbooks (Greene & Coxhead, 2015; Hu et al., 2021), the lower levels of proficiency the participants displayed on vocabulary items from the AWL would suggest that they would be likely to encounter difficulties reading texts appropriate for their grade level.

Table 3. Vocabulary mastery at each grade level (All participants).

Grade	1000	2000	3000	4000	5000	AWL
Grade 6 (<i>n</i> = 6)	100.00%	33.33%	16.67%	16.67%	16.67%	16.67%
Grade 7 (<i>n</i> = 18)	88.89%	50.00%	5.56%	27.78%	27.78%	5.56%
Grade 8 (<i>n</i> = 21)	95.24%	66.67%	33.33%	42.86%	38.10%	19.05%
Grade 9 (<i>n</i> = 24)	95.83%	83.33%	50.00%	45.83%	45.83%	20.83%
Grade 10 (<i>n</i> = 24)	87.50%	79.17%	25.00%	50.00%	45.83%	12.50%
Grade 11 (<i>n</i> = 37)	97.30%	86.49%	37.84%	48.65%	40.54%	27.03%
Grade 12 (<i>n</i> = 9)	100.00%	88.89%	55.56%	66.67%	44.44%	55.56%
Average (<i>n</i> = 139)	94.24%	74.82%	33.09%	44.60%	39.57%	20.86%

Note: AWL = Academic Word List.

We conducted separate analyses for EAL learners and PL2/FLE learners, highlighting that EAL participants showed significantly lower levels of proficiency with both high-frequency and mid-frequency vocabulary (Table 4) compared to PL2/FLE learners (Table 5). We found that the majority of PL2/FLE participants were able to master the first 5000 BNC/COCA frequency bands. In contrast, fewer than 50% of EAL participants could master beyond the 2000-word band. Even at grade 12, only 50% of EAL participants had mastered the 3000-word band, and just 33% had mastered the 5000-word band. However, despite their higher levels of vocabulary proficiency, we found that many of the PL2/FLE still struggled with the AWL, suggesting that vocabulary, particularly academic vocabulary, remains a challenge for even this more adept group of students.

Table 4. Vocabulary mastery at each grade level (EAL Learners).

Grade	1000	2000	3000	4000	5000	AWL
Grade 6 (<i>n</i> = 4)	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Grade 7 (<i>n</i> = 13)	84.62%	30.77%	0.00%	7.69%	0.00%	0.00%
Grade 8 (<i>n</i> = 15)	93.33%	53.33%	20.00%	26.67%	20.00%	0.00%
Grade 9 (<i>n</i> = 17)	94.12%	76.47%	29.41%	29.41%	23.53%	0.00%
Grade 10 (<i>n</i> = 20)	85.00%	75.00%	15.00%	40.00%	35.00%	10.00%
Grade 11 (<i>n</i> = 28)	96.43%	82.14%	17.86%	35.71%	25.00%	14.29%
Grade 12 (<i>n</i> = 6)	100.00%	83.33%	50.00%	50.00%	33.33%	50.00%
Average (<i>n</i> = 103)	92.23%	66.02%	18.45%	30.10%	22.33%	8.74%

Note: AWL = Academic Word List.

Table 5. Vocabulary mastery at each grade level (PL2/FLE learners).

Grade	1000	2000	3000	4000	5000	AWL
Grade 6 (<i>n</i> = 2)	100.00%	100.00%	50.00%	50.00%	50.00%	50.00%
Grade 7 (<i>n</i> = 5)	100.00%	100.00%	20.00%	80.00%	100.00%	20.00%
Grade 8 (<i>n</i> = 6)	100.00%	100.00%	66.67%	83.33%	83.33%	66.67%
Grade 9 (<i>n</i> = 7)	100.00%	100.00%	100.00%	85.71%	100.00%	71.43%
Grade 10 (<i>n</i> = 4)	100.00%	100.00%	75.00%	100.00%	100.00%	25.00%
Grade 11 (<i>n</i> = 9)	100.00%	100.00%	100.00%	88.89%	88.89%	66.67%
Grade 12 (<i>n</i> = 3)	100.00%	100.00%	66.67%	100.00%	66.67%	66.67%
Average (<i>n</i> = 36)	100.00%	100.00%	75.00%	86.11%	88.89%	55.56%

Note: AWL = Academic Word List.

The descriptive statistics of the scores for each of the Grade levels for the EAL and FLE/PL2 groups are given in Table 6.

Table 6. Descriptive statistics of the VLT scores for the EAL and FLE/PL2 across grade levels and frequency bands.

Grade and Group	<i>n</i>	VLT Score (Mean)	VLT Score (SD)	K1 (Mean)	K1 (SD)	K2 (Mean)	K2 (SD)	K3 (Mean)	K3 (SD)	K4 (Mean)	K4 (SD)	K5 (Mean)	K5 (SD)	AWL (Mean)	AWL (SD)
Grade 6 EAL	4	63.73	9.24	95.83	3.39	71.88	12.88	54.17	11.79	63.52	12.91	56.25	12.50	40.83	13.73
Grade 6 FLE/PL2	2	86.90	12.30	97.90	2.97	93.75	8.84	79.20	17.68	85.40	14.71	83.30	17.68	81.65	11.81
Grade 7 EAL	13	58.55	16.31	93.14	10.90	77.75	16.42	42.85	20.41	49.33	24.24	43.85	19.16	44.50	19.17
Grade 7 FLE/PL2	5	87.74	5.62	98.32	2.30	99.40	1.34	76.06	12.27	88.64	8.34	90.40	3.57	74.00	11.89
Grade 8 EAL	15	71.36	17.28	95.90	7.84	83.45	17.00	65.31	22.99	70.67	18.70	56.89	30.46	55.99	20.75
Grade 8 FLE/PL2	6	92.57	9.12	99.30	1.71	99.50	1.22	88.77	9.73	94.30	12.02	90.75	15.38	82.67	19.58
Grade 9 EAL	17	73.96	14.80	96.42	6.34	91.51	11.50	65.88	20.58	67.08	22.44	66.52	20.60	56.77	16.78
Grade 9 FLE/PL2	7	94.76	2.31	98.20	2.24	98.57	3.78	91.54	4.96	96.61	6.42	92.59	2.19	90.86	4.70
Grade 10 EAL	20	74.97	17.75	95.78	8.49	85.97	18.51	66.47	19.91	71.92	21.78	65.21	25.82	64.44	21.44
Grade 10 FLE/PL2	4	93.28	1.97	100.00	0.00	100.00	0.00	87.00	2.92	97.50	3.32	91.58	3.60	83.15	8.29
Grade 11 EAL	28	75.55	14.27	97.78	4.76	90.42	12.01	66.31	18.08	73.50	20.05	64.84	22.79	60.62	22.21
Grade 11 FLE/PL2	9	95.03	3.31	99.53	1.40	100.00	0.00	93.36	3.59	96.67	6.10	92.14	4.51	88.70	8.52
Grade 12 EAL	6	79.63	16.38	98.50	1.64	91.67	10.78	77.33	16.69	74.17	23.75	66.67	23.87	69.83	27.37
Grade 12 FLE/PL2	3	95.83	5.31	100.00	0.00	100.00	0.00	94.43	9.64	98.60	2.42	89.73	9.34	92.33	10.79

Note: K01 to K05 refer to the 1000 to 5000 word frequency bands; AWL refers to the Academic Word List.

A linear regression model was used to compare the vocabulary knowledge of EAL learners with that of FLE and PL2 learners across different grade levels (Table 7). Given the small number of participants at some grade levels, the participants were divided up into pre-IB, covering Grades 6 to 9, and IB, covering Grades 10 to 12. The results showed

that FLE/PL2 learners consistently outperformed EAL learners on the Vocabulary Levels Test (VLT). On average, FLE/PL2 learners scored about 19 percentage points higher than EAL learners. This significant difference in vocabulary knowledge was evident across both the pre-IB (Grades 6–9) and IB (Grades 10–12) levels. Additionally, EAL learners in pre-IB scored lower than those in the IB level, indicating an improvement in vocabulary knowledge as they advanced in grade levels. However, the gap between EAL and FLE/PL2 learners remained consistent across grade levels, suggesting that while all learners develop their vocabulary over time, EAL learners continue to lag behind their FLE/PL2 peers throughout their schooling.

Table 7. A summary of the linear regression model comparing vocabulary scores of FLE/PL2 and EAL learners across school levels.

	Estimate	Std. Error	t Value	Pr(> t)
Intercept	0.76	0.02	39.26	$<2 \times 10^{-16}$ ***
FLE/PL2	0.19	0.04	4.70	6.45×10^{-6} ***
Pre-IB	−0.08	0.03	−2.70	0.01 **
FLE/PL2 × Pre-IB	0.04	0.06	0.79	0.43

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Research Question 2: To what extent does the vocabulary in subject-specific textbooks at an international school align with BNC/COCA frequency bands, and how does this alignment vary across IB courses?

The vocabulary profiles of the domain-specific corpora (Table 8) show that the literature textbooks were likely to be the most accessible for EAL learners. Based on the coverage of the BNC/COCA over this corpus, we would expect learners to achieve the 95% coverage threshold required for comprehension if they could master the first five 1000-word bands from the BNC/COCA. The next most accessible group of textbooks were those from the maths corpus, where the coverage provided by the first five 1000-word bands was 94.28%, very close to the minimum threshold of 95%. However, for all the other corpora, the coverage provided by the first 5000 words fell significantly below the 95% mark.

Table 8. Coverage provided by the AWL and BNC/COCA by subject.

Frequency Band	Literature	Math	Physics	Biology	Chemistry	Average
2000	86.84%	83.79%	84.03%	77.08%	77.79%	81.91%
5000	96.55%	94.28%	91.48%	91.45%	92.60%	93.27%
AWL	6.87%	8.30%	7.36%	7.70%	8.60%	7.77%

Note: AWL = Academic Word List.

The first 2000 frequency bands provided notably low coverage, particularly in the chemistry and biology texts, where less than 78% of all tokens were from the first two bands. This is significantly lower than the coverage these bands have been found to provide over texts written specifically for English language learners, where the first 2000 words have been found to provide over 92% coverage (Sun & Dang, 2020). These findings suggest that EAL learners would struggle with these textbooks given their vocabulary mastery level.

Research Question 3: To what extent can EAL learners be expected to comprehend the vocabulary encountered in the textbooks used for different IB subjects?

We estimated the percentage of vocabulary in the textbooks that EAL learners would likely know by analysing their scores on the NVLT or uVLT. We then looked at the coverage provided by each of the bands of the BNC/COCA over the various domain-specific corpora (Table 9). By looking at the coverage provided by the words from the frequency levels that

an individual participant had displayed mastery of, we were able to determine the overall vocabulary coverage a learner would likely have over the texts from different subjects. Using these numbers, we were able to infer what percentage of words the participants would likely know from the different corpora.

Table 9. Expected coverage per discipline from the domain-specific subcorpora.

Category	Literature	Biology	Chemistry	Physics	Math
1000 word band	74.54%	65.51%	63.39%	70.88%	72.38%
2000 word band	9.56%	10.89%	12.94%	12.37%	9.96%
3000 word band	7.35%	9.23%	9.26%	7.58%	8.17%
4000 word band	1.52%	3.03%	3.78%	3.28%	2.71%
5000 word band	0.84%	2.15%	1.76%	1.28%	1.61%
6 to 25,000 word bands	2.33%	7.08%	5.96%	3.30%	3.05%
Other	3.86%	2.12%	2.90%	1.30%	2.12%

To assess whether the differences in coverage that learners were likely to achieve across subjects, as identified in the descriptive analysis, were statistically significant, a repeated-measures ANOVA was conducted, followed by pairwise comparisons using the Tukey adjustment to control for multiple comparisons. The ANOVA revealed a significant difference in coverage between academic subject, $F(4, 408) = 4858.89, p < 0.001, \eta^2 = 0.09$, showing that coverage learners would be likely to have over the different sub-corpora would vary between subjects. The subsequent post-hoc analysis showed that the largest difference was between biology and literature, with students achieving significantly higher coverage in literature ($M = 88.98, SE = 0.79$) than in biology ($M = 82.39, SE = 0.83$; Mean Difference = 6.58, $p < 0.001$). The only non-significant difference was between math ($M = 87.54, SE = 0.83$) and physics ($M = 87.54, SE = 0.85$; Mean Difference = 0.01, $p = 0.912$), suggesting learners would have similar vocabulary coverage across both of these two subjects. These findings suggest that, with one exception, the coverage provided by the BNC/COCA corpus is domain-specific, indicating that learners require varying levels of vocabulary support depending on the subject they are studying.

Figure 3 presents a histogram of the vocabulary coverage participants would likely know for each corpus. The data reveal that most EAL participants lack the vocabulary needed to read discipline-specific textbooks, with biology and chemistry being the most challenging subjects and maths and literature the easiest. Given the vocabulary levels displayed by EAL learners, even at the Grades 11 and 12 level, we expect them to struggle to fully comprehend the textbooks used in their classes.

Research Question 4: Using the BNC/COCA, how much additional vocabulary would EAL learners need to acquire to understand the language used in textbooks they would typically encounter within the international school context?

To answer this fourth and final question, we need to look at the gap that exists between the learners' current knowledge and the number of words they would need to learn, using existing word lists, to achieve 95% or greater coverage. From what we have discussed above, it is evident that learners often lack proficiency in frequency bands exceeding the first 2000 words. We also know that they are not likely to have mastered the Academic Word List. Since learners would have to master the 5000 to 9000 most frequently used words in the BNC/COCA to adequately cover various subjects, the number of words they would need to learn solely based on the BNC/COCA is likely too large for them to acquire within their available time (Table 10). For example, a typical 10th-grade EAL student would need to learn around 7000 new words to achieve the threshold necessary to comprehend a biology or chemistry textbook successfully.

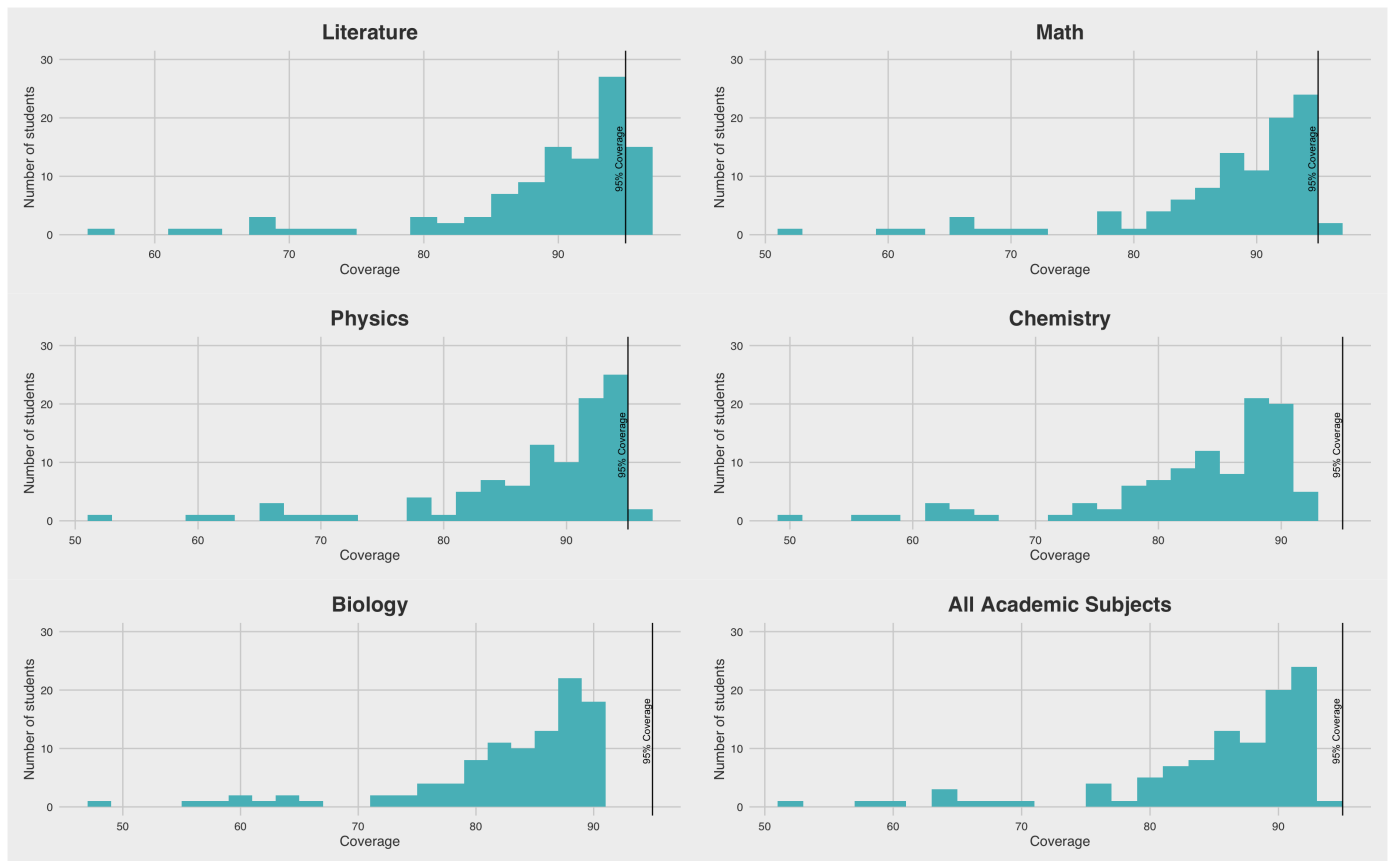


Figure 3. A histogram depicting the percentage of words EAL participants would likely understand in each of the subject-specific corpora.

Table 10. Mastery of the word frequency bands for individual participants.

Participant	1000	2000	3000	4000	5000	AWL
S04	100%	97%	80%	87%	87%	67%
S09	100%	100%	67%	77%	90%	63%
S12	100%	97%	70%	87%	83%	77%
S28	100%	100%	70%	93%	87%	73%
S32	100%	97%	67%	100%	73%	67%
S51	100%	100%	87%	83%	87%	77%

Note: AWL = Academic Word List. Bold text indicates participants who have shown mastery of that frequency band.

Examples from our study further illustrate this issue: participant S04 was able to master the 4000- and 5000-word bands but not the 3000-word band. Participant S09 mastered the 5000-word band but not the 3000- or 4000-word bands. According to previous studies (e.g., Schmitt & Schmitt, 2014; Teng, 2019), learners are likely to acquire words based on the number of exposures to those words. Given this fact, our finding suggests that, beyond the first 2000 high-frequency words, the BNC/COCA frequency list may not accurately represent the frequency of the vocabulary items in the textbooks that EAL learners are being asked to read for their classes. This misalignment calls into question the effectiveness of using traditional frequency lists for supporting EAL learners’ vocabulary acquisition. Given this fact and the strong and significant correlation between vocabulary knowledge and reading comprehension shown by other studies (e.g., Brooks et al., 2021; Laufer & Ravenhorst-Kalovski, 2010), more research is needed regarding EAL learners’

vocabulary knowledge and needs in an EMI context. Additionally, there is a pressing need to develop tools that effectively meet these vocabulary needs.

7. Discussion

The current study was designed to examine the vocabulary knowledge of EAL learners in an international context in Japan and compare it to the vocabulary found in the textbooks these learners use in class, using the BNC/COCA word list. The aim was to gain insight into the vocabulary used in these textbooks to identify potential challenges EAL learners may face while reading them. The study was driven by four research questions, each of which we discuss in turn below.

Regarding Research Question 1, the data strongly support the findings of previous research (Coxhead & Boutorwick, 2018), indicating a widespread challenge among learners, especially those with English as an additional language, in mastering academic vocabulary. Similar challenges were noted in studies by Dixon et al. (2020), which found that while EAL learners do acquire vocabulary, they are often not able to do so with sufficient speed to keep up with their FLE-speaking peers. The finding that fewer than 25% of participants could master the AWL before Grade 12 is concerning, as academic vocabulary is essential for reading comprehension and success in the classroom (Brooks, 2023; Green & Lambert, 2018; Greene & Coxhead, 2015). Studies conducted by Coxhead and Boutorwick (2018) and Marianne and Coxhead (2023) yielded similar findings, indicating low levels of academic vocabulary mastery among EAL learners in comparable educational contexts. Given their difficulties with academic texts, it is probable that EAL learners will face substantial challenges in their academic pursuits, particularly as they progress to higher grades where the academic demands intensify.

For Research Question 2, the results reinforce the need for domain-specific vocabulary lists, a need that has been extensively discussed in earlier research (Green & Lambert, 2018; Greene & Coxhead, 2015). In line with previous studies on vocabulary (Coxhead, 2017), it appears that the difficulty of the vocabulary found in a corpus can vary markedly across domains. While more proficient learners may possess sufficient vocabulary resources to handle content in subjects like literature and mathematics, which were found to be relatively accessible with approximately 95% coverage achieved using the first 5000 BNC/COCA word bands, specialised subjects such as biology and chemistry present greater lexical challenges, resulting in a notable drop in coverage. This builds upon previous studies that have shown both the difficulty of the vocabulary in textbooks from a specific domain, such as the sciences (Hu et al., 2021), as well as the importance of this vocabulary because of the strong connection between content knowledge and technical vocabulary (Coxhead, 2017; Woodward-Kron, 2008). The results of our research corroborate the findings of previous studies and underscore the distinct obstacles posed by subject-specific terminology, emphasising the necessity of providing tailored lexical assistance to students grappling with such domains.

Although the 5000 most frequent words offer a wide range of coverage, the BNC/COCA corpus falls short in adequately capturing the specialised terminology prevalent in the fields of biology, chemistry, and physics. This deficiency echoes similar findings from prior studies on the vocabulary demands of English as an additional language (EAL) learners, highlighting the need for a more comprehensive approach to vocabulary acquisition in these subjects (Coxhead & Boutorwick, 2018; Coxhead et al., 2010). The low coverage in these specialised domains emphasises the difficulties faced by EAL learners, especially considering that only a minority (e.g., 33% of Grade 12 participants) demonstrated mastery of the requisite vocabulary bands. This further underscores the importance of developing

targeted vocabulary lists to support comprehension in more specialised academic texts, as earlier studies have suggested (Green & Lambert, 2018; Greene & Coxhead, 2015).

Research Question 3 investigates the vocabulary coverage of discipline-specific textbooks, analysing the expected coverage based on English as an additional language (EAL) learners' Vocabulary Levels Test (VLT) scores. This analysis reveals a concerning trend: most EAL learners possess insufficient vocabulary to fully comprehend the content within these textbooks. This supports earlier research, which has shown that these learners tend to struggle with reading comprehension in an academic setting (Brooks, 2023; Coxhead et al., 2010; Murphy & Unthiah, 2015).

While the EAL learners in this study demonstrated improvement in vocabulary knowledge across grade levels, the rate of this growth did not match the rapid gains observed in other researchers, such as Coxhead et al. (2015), who found that 15–16-year-old FLE speakers enrolled in New Zealand high schools learned an average of over 1300-word families per year. While our study indicated vocabulary knowledge improvements across grade levels, these improvements were less substantial than those observed in a study by Coxhead and Boutorwick (2018), where some EAL learners achieved comparable vocabulary knowledge levels to their native-speaking peers within a shorter timeframe—four years for high-frequency vocabulary and five years for academic vocabulary. Our research reveals a persistent gap in vocabulary knowledge, even among upper-grade learners, particularly in subjects like biology and chemistry. Even Grade 12 English as an additional language (EAL) learners show a significant lack of proficiency in essential vocabulary bands, highlighting a substantial discrepancy between their vocabulary comprehension and the demands of these disciplines. This presents a significant barrier to their academic success in these subjects (Marianne & Coxhead, 2023; Woodward-Kron, 2008).

Although the BNC/COCA list could potentially help address these vocabulary gaps, our analysis for Research Question 4 reveals that EAL learners' vocabulary profiles deviate significantly from the typical profile predicted by this frequency list (Nation, 2016), and further suggests that the number of words they would need to learn based on these lists is too extensive to be realistically managed. These lists organise words by how often learners encounter them in real-world contexts, with higher-frequency words appearing more frequently in texts. As a result, learners are expected to learn higher-frequency words sooner due to repeated exposure. However, previous studies have shown that textbooks in subjects like maths and science can have very different vocabulary profiles and loads than other texts, such as novels or English language textbooks (Groves, 2016; Hu et al., 2021), which could result in EAL learners acquiring the vocabulary in a different order than ESL or EFL learners. The results of our study, which support previous research, indicate a notable disparity in vocabulary knowledge growth, particularly within the 3000-word band. In this band our participants demonstrated a mastery level that was noticeably lower (by more than 10%) compared to the mastery they displayed in the 4000-word band. Table 9 highlights this pattern with examples from individual learners. The misalignment between the expected and actual vocabulary acquisition sequences, where learners are expected to acquire bands in order, underscores the challenges associated with using the BNC/COCA for teaching mid- to low-frequency vocabulary to this particular group of learners.

Furthermore, given that previous studies have shown that vocabulary acquisition takes place at a relatively steady pace (Dixon et al., 2020; Webb & Chang, 2012), it is unlikely that learners would be able to acquire the vocabulary knowledge required to gain the 95% to 98% coverage necessary for comprehension during their time at school using the BNC/COCA alone. This supports calls from researchers (Coxhead & Boutorwick, 2018; Nation, 2016) that highlight the need to develop domain-specific word lists for EAL learners in order to help them acquire the vocabulary knowledge they need to succeed academically.

8. Implications and Limitations

The current study has several limitations that should be considered when interpreting its findings. One potential issue, as previously discussed, is the use of two different vocabulary assessment tools to measure the vocabulary knowledge of the participants in the study. In the future, it would be preferable to conduct a similar study with the same types of assessment for both groups of learners. However, it is important to keep in mind that both assessment tools used in this study were developed for adult EFL learners, and neither may be entirely suitable for EAL learners. Unfortunately, there are limited assessment tools for use with this group of learners. While tools do exist for younger EAL learners, such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 2012), developed for young FLE speakers, and the English Picture Vocabulary Test (Güngör & Önder, 2023), developed for young L2 learners of English, there are no tools that have been specifically designed to measure the vocabulary knowledge for the age group of EAL learners participating in this study. The development of such a test is important, based on both the findings of this study as well as those of previous research (Green & Lambert, 2018; Greene & Coxhead, 2015), which have shown that the vocabulary needed in the EMI classroom is different from the vocabulary found in frequency lists developed for adult learners. A further limitation of the current study relates to the content of the corpus. The corpus used for this study was composed entirely of textbooks developed for the classes being taught. While it would have been beneficial to include a spoken component in the corpus, the logistics of making recordings in the classroom and transcribing a sufficient number of those recordings to develop a spoken corpus of a comparable size to the written corpus made it impossible to undertake as part of this study. Given the acknowledged differences between spoken and written English (Dang et al., 2017), we feel that it would be beneficial to conduct a follow-up study using a corpus of spoken English. A final limitation that needs to be addressed is the effect of L1 backgrounds, which the literature has shown can affect learners' English vocabulary knowledge (Booth & Clenton, 2020). While our dataset includes assessments from learners with diverse linguistic backgrounds, we were unable to evaluate how these backgrounds might impact their knowledge of academic English. We would like to pursue this in the future should we be able to include a sufficient number of learners from similar L1 backgrounds. We would also like to expand the scope of our study to include EAL learners studying in schools outside of Japan.

Despite the limitations discussed above, the current study offers some important implications with regard to supporting EAL learners in the classroom. First, as suggested by previous studies (Coxhead & White, 2012; Green & Lambert, 2019; Greene & Coxhead, 2015), educators looking to support EAL should prioritise enhancing their students' vocabulary knowledge. The data from this study highlight a pressing need to enable learners to expand their vocabulary knowledge in order to comprehend the textbooks they are asked to read. Additionally, there is an urgent requirement to develop word lists specifically tailored for EAL learners. While lists such as the AWL (Coxhead, 2000) and the BNC/COCA (Nation, 2020) are able to provide EAL learners studying in the international school context some support in acquiring the vocabulary they need to understand the textbooks that they are being asked to read, they are often either too broad (in the case of the BNC/COCA) or too limited (in the case of the AWL) in scope to provide learners with the support they require. While including important words, tools like the AWL were developed using a corpus of articles written for adult learners and may not include all of the words that EAL learners need in their context. On the other hand, the BNC/COCA contains too many words, requiring EAL learners to acquire upwards of 7000 new words before they are able to achieve sufficient coverage to understand the textbooks from subjects such as biology or chemistry.

Furthermore, both sets of word lists, by design, cover vocabulary from multiple domains, making it difficult for teachers to select the words most appropriate for a single subject. This is important because research shows us (e.g., Teng, 2019; van Zeeland, 2013) that it is more effective to teach words in context. A set of domain-specific word lists focusing on the various subjects that EAL learners studying in the international school context would be likely to study would allow teachers to focus on a manageable list of words and teach them within the context of the subject where those words are required.

9. Conclusions

Our research emphasises the vocabulary gaps among EAL learners, highlighting the need for targeted vocabulary support in the classroom. The crucial role of vocabulary in EAL learners' reading comprehension shown by previous studies (e.g., Brooks et al., 2021; Marianne & Coxhead, 2023), along with the gaps in this knowledge illustrated by the current study, further underscores the need to prioritise vocabulary acquisition. Equipping EAL learners with the academic vocabulary they need can be achieved through classroom instruction and additional support outside of the classroom. This study highlights why placing greater emphasis on vocabulary instruction is vital for EAL learners to effectively comprehend the texts they encounter in their coursework.

Nonetheless, viewing vocabulary in the broader context of the types of texts EAL learners are being asked to read for their class is also essential. Therefore, teachers within the EAL setting, where content learning often takes precedence over language acquisition, need to be given the tools to identify the most relevant vocabulary for their learners. This will allow them to integrate activities that enhance the vocabulary knowledge and overall English language skills that learners need for the classes that they are enrolled in. We hope that this study will provide a starting point for such research in the future.

Author Contributions: Conceptualization, G.B., J.C. and S.F.; Data curation, G.B. and J.C.; Formal analysis, G.B.; Funding acquisition, G.B., J.C. and S.F.; Investigation, G.B. and J.C.; Methodology, G.B., J.C. and S.F.; Project administration, G.B., J.C. and S.F.; Resources, G.B., J.C. and S.F.; Supervision, J.C.; Validation, G.B.; Visualization, G.B.; Writing—original draft, G.B.; Writing—review & editing, G.B., J.C. and S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by two Grants-in-Aid for Scientific Research (No. 17K03035 and No. 20K00793) from the Japan Society for the Promotion of Science. The authors are very grateful for this support.

Institutional Review Board Statement: The study has been approved by the Research Ethics Committee of the Graduate School of Humanities and Social Sciences, Hiroshima University (approval number: HR-HUM-000762).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: We extend our sincere gratitude to the two anonymous reviewers and the *Languages* production team for their valuable comments and suggestions. Their insightful feedback has greatly improved the quality of our manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Afitska, O., & Heaton, T. J. (2019). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education*, 103(6), 1396–1422. [CrossRef]

- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice: A Publication of the Division for Learning Disabilities, Council for Exceptional Children*, 20(1), 50–57. [CrossRef]
- Beglar, D. (2010). A rasch-based validation of the vocabulary size test. *Language Testing*, 27(1), 101–118. [CrossRef]
- Benjamin, R. G. (2011). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1), 63–88. [CrossRef]
- Benson, S., & Madarbakus-Ring, N. (2021). A comparison of textbook vocabulary load analysis. *Vocabulary Learning and Instruction*, 10(2), 9–17. [CrossRef]
- Booth, P., & Clenton, J. (2020). *First language influences on multilingual lexicons*. Routledge. [CrossRef]
- Brooks, G. (2023). *Bridging the vocabulary gap for English as an additional language learners* [Ph. D. thesis, Hiroshima University].
- Brooks, G., Clenton, J., & Fraser, S. (2021). Exploring the importance of vocabulary for English as an additional language learners' reading comprehension. *Studies in Second Language Learning*, 11(3), 351–376. [CrossRef]
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, 58(5), 412–424. [CrossRef]
- Carder, M. (2007). Bilingualism in international schools. *Multilingual Matters*. [CrossRef]
- Castellano-Risco, I., Alejo-González, R., & Piquer-Piriz, A. M. (2020). The development of receptive vocabulary in CLIL vs EFL: Is the learning context the main variable? *System*, 91, 102263. [CrossRef]
- Chuang, H., Joshi, R. M., & Dixon, L. Q. (2012). Cross-language transfer of reading ability. *Journal of Literacy Research*, 44(1), 97–119. [CrossRef]
- Clegg, J., & Afitska, O. (2011). Teaching and learning in two languages in African classrooms. *Comparative Education Review*, 47(1), 61–77. [CrossRef]
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. [CrossRef]
- Coxhead, A. (2012). Researching vocabulary in secondary school English texts: 'The hunger games' and more. *English in Aotearoa*, 78, 34–41.
- Coxhead, A. (2017). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives* (Routledge research in English for specific purposes). Routledge. [CrossRef]
- Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, maths, and science. *TESOL Quarterly*, 52(3), 588–610. [CrossRef]
- Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of Carpentry: Loads, lists and bearings. *TESOLANZ Journal*, 24, 38–71.
- Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, 50(1), 121–135. [CrossRef]
- Coxhead, A., Stevens, L., & Tinkle, J. (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics*, 16(2), 37–52.
- Coxhead, A., & White, R. (2012). Building a corpus of secondary school texts: First you have to catch the rabbit. *New Zealand Studies in Applied Linguistics*, 18(2), 67–73.
- Dale, E., & O'Rourke, J. P. (1976). *The living word vocabulary: The words we know, A national vocabulary inventory*. Field Enterprises Educational Corp.
- Dang, T. N. Y. (2019). Corpus-based word lists in second language vocabulary research, learning, and teaching. In *The Routledge handbook of vocabulary studies* (pp. 288–303). Routledge. [CrossRef]
- Dang, T. N. Y., Coxhead, A., & Webb, S. A. (2017). The academic spoken word list. *Language Learning*, 67(4), 959–997. [CrossRef]
- De Clercq, O., & Hoste, V. (2016). All mixed up? Finding the optimal feature set for general readability prediction and its application to English and Dutch. *Computational Linguistics (Association for Computational Linguistics)*, 42(3), 457–490. [CrossRef]
- Dixon, C., Thomson, J., & Fricke, S. (2020). Evaluation of an explicit vocabulary teaching intervention for children learning English as an additional language in primary school. *Child Language Teaching and Therapy*, 36(2), 91–108. [CrossRef]
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, 38(1), 78–103. [CrossRef]
- Dunn, L. M., & Dunn, L. M. (2012). *Peabody picture vocabulary test* (3rd ed.). American Guidance Service: PscTESTS Dataset. APA PscTests. [CrossRef]
- Faitaki, F., Hessel, A., & Murphy, V. A. (2022). Vocabulary and grammar development in young learners of English as an additional language. In M. Schwartz (Ed.), *Handbook of early language education* (pp. 428–444). Springer International Publishing. [CrossRef]
- Flor, M., Holtzman, S., Deane, P., & Bejar, I. (2024). Mapping of American English vocabulary by grade levels. *ITL—International Journal of Applied Linguistics*, 175(1), 25–45. [CrossRef]
- Fraser, S. (2010). *The lexis of pharmacology texts: A corpus linguistic analysis* [Ph.D. thesis, Swansea University].
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. [CrossRef]

- Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes*, 35, 105–115. [CrossRef]
- Green, C., & Lambert, J. (2019). Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through secondary phrase lists. *English for Specific Purposes*, 53, 1–12. [CrossRef]
- Greene, J. W., & Coxhead, A. (2015). *Academic vocabulary for middle school students: Research-based lists and strategies for key content areas*. Paul H. Brookes Publishing.
- Groves, F. H. (2016). A longitudinal study of middle and secondary level science textbook vocabulary loads. *School Science and Mathematics*, 116(6), 320–325. [CrossRef]
- Güngör, B., & Önder, A. (2023). Development of English picture vocabulary test as an assessment tool for very young EFL learners' receptive and expressive language skills. *Early Education and Development*, 34(2), 572–589. [CrossRef]
- Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11(1), 20. [CrossRef]
- Hancke, J., Vajjala, S., & Meurers, W. D. (2012, December 8–15). *Readability classification for German using lexical, syntactic, and morphological features*. International Conference on Computational Linguistics (pp. 1063–1080), Mumbai, India.
- Hessel, A. K., & Murphy, V. A. (2019). Understanding how time flies and what it means to be on cloud nine: English as an Additional Language (EAL) learners' metaphor comprehension. *Journal of Child Language*, 46(2), 265–291. [CrossRef]
- Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage. *Electronic Journal of Foreign Language Teaching*, 6(1), 42–62.
- Hu, J., Gao, X., & Qiu, X. (2021). Lexical coverage and readability of science textbooks for English-medium instruction secondary schools in Hong Kong. *SAGE Open*, 11(1), 215824402110018. [CrossRef]
- International Baccalaureate. (2023). *International baccalaureate facts and figures*. Available online: <https://www.ibo.org/about-the-ib/facts-and-figures/> (accessed on 6 September 2023).
- Kremmel, B., Indrathne, B., Kormos, J., & Suzuki, S. (2023). Unknown vocabulary density and reading comprehension: Replicating Hu and Nation (2000). *Language Learning*, 73(4), 1127–1163. [CrossRef]
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. In C. Lauren, & M. Nordman (Eds.), *From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.
- Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307–322. [CrossRef]
- 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.
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42–53. [CrossRef]
- Leung, C. (2014). Researching language and communication in schooling. *Linguistics and Education*, 26, 136–144. [CrossRef]
- Luxton, J., Fry, J., & Coxhead, A. (2017). Exploring the knowledge and development of academic English vocabulary of students in New Zealand secondary schools. *Set: Research Information for Teachers*, 1, 12–22. [CrossRef]
- Marianne, & Coxhead, A. (2023). Getting to know your learners in an EAL context. In *EAL research for the classroom* (pp. 185–208). Routledge. [CrossRef]
- 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., & Kramer, B. (2016). The creation of a new vocabulary levels test. *Shiken*, 19(1), 1–9.
- Melby-Lervåg, M., & Lervåg, A. (2014). Effects of educational interventions targeting reading comprehension and underlying components. *Child Development Perspectives*, 8(2), 96–100. [CrossRef]
- Melby-Lervåg, M., Lyster, S. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352. [CrossRef] [PubMed]
- Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30–53. [CrossRef]
- Murphy, V. A. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford University Press.
- Murphy, V. A., & Unthiah, A. (2015). *A systematic review of intervention research examining English language and literacy development in children with English as an Additional Language (EAL)*. Educational Endowment Foundation.
- NALDIC. (2015). *EAL achievement: The latest information on how well EAL learners do in standardised assessments compared to all students*. NALDIC.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, P. (2005). *Teaching and learning vocabulary*. Taylor Francis.
- Nation, P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company. [CrossRef]

- Nation, P. (2020). *The BNC/COCA word family lists*. Available online: https://www.wgtn.ac.nz/__data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf (accessed on 14 August 2022).
- Nguyen, C. (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. [CrossRef]
- Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, 23(2), 189–208. [CrossRef]
- O’Loughlin, R. (2012). Tuning In to vocabulary frequency in coursebooks. *RELC Journal*, 43(2), 255–269. [CrossRef]
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- 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. [CrossRef]
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. [CrossRef]
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. [CrossRef]
- Sharples, R. (2021). Teaching EAL: Evidence-based strategies for the classroom and school. *Multilingual Matters*. [CrossRef]
- Smith, B. (2015). *US, UK set to lose market share of mobile students in next decade*. Available online: <https://thepienews.com/news/us-uk-set-to-lose-market-share-of-mobile-students-in-next-decade/> (accessed on 21 February 2020).
- Spencer, S., Clegg, J., Stackhouse, J., & Rush, R. (2017). Contribution of spoken language and socio-economic background to adolescents’ educational achievement at age 16 years. *International Journal of Language & Communication Disorders/Royal College of Speech & Language Therapists*, 52(2), 184–196. [CrossRef]
- Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93, 102279. [CrossRef]
- Taylor, S. E., Frackenpohl, H., & White, C. E. (1989). *EDL core vocabularies in reading, mathematics, science, and social studies*. Steck-Vaughn Company.
- Teng, F. (2019). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *Language Learning Journal*, 47(2), 145–158. [CrossRef]
- Townsend, D., Bear, D., Templeton, S., & Burton, A. (2016). The implications of adolescents’ academic word knowledge for achievement and instruction. *Reading Psychology*, 37(8), 1119–1148. [CrossRef]
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45(5), 453–466. [CrossRef]
- Uemura, T., & Ishikawa, S. (2004). JACET 8000 and Asia TEFL vocabulary initiative. *Journal of Asia TEFL*, 1(1), 333–347.
- US Department of Education, National Center for Education Statistics. (2022). *English learners in public schools*. Available online: <https://nces.ed.gov/programs/coe/indicator/cgf/english-learners> (accessed on 13 November 2022).
- van Zeeland, H. (2013). L2 vocabulary knowledge in and out of context. *Australian Review of Applied Linguistics*, 36(1), 52–70. [CrossRef]
- Webb, S. A., & Chang, A. C. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126. [CrossRef]
- Webb, S. A., & Nation, P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1–10.
- Webb, S. A., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *International Journal of Applied Linguistics*, 168(1), 33–69. [CrossRef]
- West, M. (1953). *A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longman.
- Woodward-Kron, R. (2008). More than just jargon—The nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes*, 7(4), 234–249. [CrossRef]
- Xodabande, I., & Hashemi, M. R. (2023). Learning English with electronic textbooks on mobile devices: Impacts on university students’ vocabulary development. *Education and Information Technologies*, 28(2), 1587–1611. [CrossRef] [PubMed]
- Yang, L., & Coxhead, A. (2022). A corpus-based study of vocabulary in the New Concept English textbook series. *RELC Journal*, 53(3), 597–611. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.