International Journal of TESOL Studies (2026)
Vol. 8 (3) 113-133 https://doi.org/10.58304/ijts.260211

Article

Lexical Profile of ChatGPT-generated Reading Materials Targeting EFL Learners Across the CEFR Levels

Nhat Ha Nguyen
The University of Danang - University of Science and Education, Vietnam
University of Leeds, UK

Xuechun Huang
Thi Ngoc Yen Dang*
University of Leeds, UK

Received: 29 October 2025/Received in revised form: 20 December 2025/Accepted: 1 February 2026/Available online: 10 February 2026

Abstract
With the capacity to produce a large number of texts within a short time, ChatGPT can be an effective tool to develop learning materials for extensive reading activities. However, no studies have tested its potential. To address this gap, this study examined the lexical profile of ChatGPT-generated texts targeting learners at six CEFR levels from A1 to C2. We created six corpora from the texts generated by ChatGPT targeting learners with each CEFR level and used RANGE (Heatley et al., 2002) and Nation's (2012) BNC/COCA 25,000-word family lists to analyse the lexical profile of these corpora. The results showed that regardless of the target CEFR levels, high-frequency words consistently constitute the largest percentage of ChatGPT-generated texts, followed by mid-frequency words and then low-frequency words. Moreover, ChatGPT-generated texts for lower levels (A1, A2 and B1) are less lexically demanding than those for higher levels (B2, C1 and C2). However, ChatGPT-generated texts for the A1 and A2 levels require the same vocabulary sizes as those targeting the B1 level, whereas the figures for C1 are slightly smaller than expected. Together, these findings indicate the potential of ChatGPT as a tool to create learning materials for learners with different proficiency but also highlight the importance of further checking and adjustment of ChatGPT generated texts to better tailor to learners' language abilities.

Keywords
Vocabulary, ChatGPT, extensive reading, lexical demand

*Corresponding author. Email: T.N.Y.Dang@leeds.ac.uk

1 Introduction

Extensive reading has been widely recommended by vocabulary researchers as an effective activity to learn vocabulary (e.g., Webb & Nation, 2017). For vocabulary learning to happen from extensive reading activities, learners need to read a large number of texts relevant to their current vocabulary levels (Nation, 2007). However, developing and selecting appropriate pedagogical materials require plenty of effort and time (Tomlinson & Masuhara, 2017), which may be challenging for teachers due to resource-and time-constraints. As ChatGPT can quickly generate a large number of texts replicating natural language use with the help of several prompts, this generative artificial intelligence (GenAI) tool may be useful to create materials for extensive reading activities. Yet very few studies have investigated its potential. The present study aims to address this gap by examining the lexical profile of texts generated by ChatGPT for learners at six CEFR levels from A1 to C2. It would provide useful information for researchers and teachers who would like to use ChatGPT-generated materials for extensive reading activities.

2 Literature Review

2.1 Designing materials for extensive reading activities

Extensive reading refers to the activity in which learners read a large number of materials at an appropriate level silently and independently (Nation & Waring, 2020). Unlike intensive reading, which aims to develop a detailed understanding of the texts under the guidance of the teacher (Carrell & Carson, 1997), extensive reading offers learners opportunities to read within their competence as much as possible for pleasure and comprehension (Day & Bamford, 1998). Extensive reading has been widely recognised for improving motivation, reading speed, fluency, and comprehension, as well as promoting vocabulary learning and writing development (Nakanishi, 2015; Suk, 2017; Zhou, 2024).

For the success of an extensive reading programme, reading materials should be easy and varied, and learners should have freedom to choose reading materials and read as much as possible with the purpose of general comprehension (Day & Bamford, 2002). In particular, Nation and Waring (2020) point out that these materials should meet several criteria. First, at least 98% of the words in the materials should be within learners' vocabulary levels so that they can easily understand these materials. Second, the reading activity should focus primarily on comprehending the content of the text rather than the linguistic features in the text so that it can effectively develop learners' reading comprehension skills. Third, learners should read a large quantity of reading materials to enhance their vocabulary size, improve other aspects of language proficiency, and develop greater reading fluency.

Despite the value of extensive reading, several challenges related to developing reading materials may hinder the chance of implementing it in English as a foreign language (EFL) contexts. The first challenge is the cost of accessing reading materials (Davis, 1995). As previously mentioned, in an effective extensive reading programme, learners should have the opportunities to select and read a large number of texts that are relevant to their interests. However, a book usually has limited copies in a school library; as a result, learners may need to purchase books themselves (Ramonda, 2020), which may not be affordable to all learners, especially for those from low socio-economic backgrounds. Another challenge is the selection of relevant materials. Developing materials that meet conditions suggested by extensive reading scholars is time- and resource-consuming, which may be challenging for many teachers given their busy schedules. As a result, teachers may go for available resources, which may not be appropriate for extensive reading programmes. For example, Puripunyavanich and Waring's (2024) survey of reading materials used in extensive reading programmes in Japan, Mongolia, Thailand, and Vietnam revealed that about 60% of the 259 surveyed teachers used non-graded readers (i.e., materials for intensive reading such as general English textbooks) or phonics readers (i.e., those for practicing pronunciation) for extensive reading activities. Puripunyavanich and Waring point out that the inappropriateness of these materials can make learners feel demotivated and do little reading in extensive reading activities, and therefore, call for further studies to make appropriate materials available. The next section will discuss the potential of ChatGPT in developing such materials.

2.2 ChatGPT and its application in language learning and vocabulary learning

Launched in November 2022, OpenAI's GenAI tool ChatGPT marked a breakthrough not only in AI technology but also in everyday life, including language learning (Pack & Maloney, 2023). Powered by a large language model called Generative Pre-trained Transformer, ChatGPT is a chatbot that can respond to specific prompts and produce human-like content (Yang & Li, 2024). While AI chatbots have been used for language learning since the 1970s, ChatGPT can foster a wider range of pedagogical activities, such as writing learning materials, providing personalised feedback, and acting as a conversation partner (Kohnke et al., 2023). However, there have been concerns regarding biased and unethical ChatGPT-generated content, hallucination (i.e., ChatGPT may generate seemingly convincing but in fact inaccurate information), overreliance on ChatGPT, and the threat of ChatGPT to academic integrity, critical thinking, privacy, and copyright (Barrot, 2023; Teng, 2024).

As for vocabulary learning, several studies have shown the affordances of ChatGPT-enhanced instruction for learning a set of target words (Abdelhalim & Alsehibany, 2025; Hao et al., 2025) and increasing vocabulary size (Mugableh, 2024; Zhang & Huang, 2024). Abdelhalim and Alsehibany (2025) conducted an experiment with 71 English language learners in Saudi Arabia to compare the effects of using ChatGPT versus a traditional teaching method. The experimental group interacted with ChatGPT for various vocabulary tasks (e.g., quizzes, sentence completion) while the control group was taught with the traditional method of teacher explanation, textbook exercise, and class discussion. It was found that the experimental group significantly outperformed the control group in terms of productive (measured by sentence writing of target words) but not receptive knowledge (measured by multiple choice questions). This finding indicates that interacting with ChatGPT benefits L2 vocabulary learning. Similarly, Hao et al. (2025) investigated relationships with self-regulation, motivation, and performance among Chinese private college students in GenAI chatbot-assisted vocabulary learning. One hundred and three students were randomly assigned to an experimental group or a control group. Both groups received the same list of target words to study for eight weeks. However, only the experimental group had access to Doubao, a popular GenAI chatbot in China, and was taught how to study these words with 28 vocabulary-learning prompts. Both groups took a vocabulary pretest and posttest at the beginning and end of the experiment. The experimental group participants also took a Vocabulary Learning Strategies questionnaire and a basic psychological needs questionnaire at the end of the experiment. Results showed that no significant differences were found between the groups in the pretest, but the experimental group outperformed the control group in the vocabulary posttest. In addition, perceived autonomy (one factor of basic psychological needs) mediated the relationship between pretest and posttest scores.

Regarding designing materials for language learners, researchers have examined the potential of using ChatGPT to assist teachers in simplifying reading materials (Koraishi, 2023; Liu et al., 2025). Moons and Van Bulck (2024) found that ChatGPT can improve the readability of written patient information materials in scientific journals, but cannot simplify them to the desired level of sixth-graders. Young and Shishido (2023a) suggested that ChatGPT can provide easier versions of EFL reading materials. Young and Shishido (2023b) further found that ChatGPT can generate appropriate EFL materials for learners at CEFR A2 and B1 levels.

However, studies evaluating the appropriateness of ChatGPT-generated materials for learners across all CEFR levels are particularly limited. One exception is Ramadhani et al. (2023), which analysed the length of texts, lexical density (the number of lexical items per clause), grammatical intricacy (the number of clauses divided by the number of sentences), and lexical variation (the ratio of unique lexical words to total lexical words) of 18 ChatGPT-generated materials and evaluated the results against the CEFR standard. They found that ChatGPT generated longer texts as the CEFR levels moved from A1 to C2, but these texts could not be clearly distinguished by their lexical density, grammatical intricacy, and lexical variation. Similarly, Uchida (2025) has evaluated the extent to which ChatGPT-generated materials, prompted to target a specific CEFR level, can be classified as belonging to that intended level by a CEFR-based Vocabulary Level Analyzer (CVLA). The results show that ChatGPT-generated texts differ by CEFR level, with A1 and A2 texts being simpler than B2, C1 and C2 texts. However, Uchida (2025) also found that only 5% of the total 180 ChatGPT-generated texts were classified by CVLA as the intended CEFR level. Moreover, ChatGPT-generated materials for A1 and A2 were too simple, while the ones for B2, C1, and C2 were too difficult. In addition, materials for B1 varied greatly. In terms of prompts, Ramadhani et al. (2023) did not mention the prompts used. Uchida (2025, p.3) used a prompt template for generating texts from A1 to C2:

> "*You are a proficient English writer. Generate a passage suitable for CEFR level {level} reading. When creating the passage, pay attention to the following points. (1) Ensure that the vocabulary and grammar used in the sentences are appropriate for {level}. (2) Keep the length of the passage suitable for {level}. (3) Choose a topic that is relevant and appropriate for {level}. (4) Ensure that the passage is designed for English language learners. (5) Return the output in the following JSON format: {"topic": "xxx", "title": "yyy", "content": ["sentence1", "sentence2", ...]".*

However, one limitation of this prompt is that it did not ensure the corpus was not strongly biased towards certain types of topics.

Together, Ramadhani et al.'s (2023) and Uchida's (2025) findings suggest that while ChatGPT has potential for generating materials across CEFR levels, these materials are far from appropriate. Despite this useful insight, it is important to note that Ramadhani et al. (2023) and Uchida (2025) had a small corpus size with only 18 and 300 texts respectively and thus failed to systematically include a wide range of topics. Moreover, these studies examined either lexical density, grammatical intricacy, and lexical variation (Ramadhani et al., 2023) or level estimation (Uchida, 2025). As the percentage of known words and the corpus frequency of words are important criteria for selecting extensive reading material (Nation & Webb, 2011), examining the vocabulary load and occurrence of high, mid, and low-frequency words in ChatGPT-generated texts would provide further insights into the potential of ChatGPT-generated text for extensive reading. The next section will review research on vocabulary load and the occurrence of high, mid, and low-frequency words in written texts in turn.

2.3 Vocabulary load of written discourse

Vocabulary load refers to the number of words required to successfully comprehend a text (Nation, 2006, 2013). Estimating the vocabulary load of texts allows teachers and educators to select suitable reading materials for their learners. To have a good degree of comprehension, learners need to possess sufficient vocabulary knowledge to reach a certain point of lexical coverage (i.e., the proportion of known words in a specific text) (Nation & Webb, 2011). Several lexical coverage points have been proposed for adequate comprehension of written discourse. Laufer (1989) suggested 95% coverage for efficient reading comprehension. However, Hu and Nation (2000) found that learners need to know 98% of the words in a written text for thorough comprehension of that text. Subsequent studies (e.g., Schmitt et al., 2011) found that comprehension increases with the amount of lexical coverage, but there was no lexical coverage threshold for reading comprehension. In fact, the lexical cut-off points depend on the desired degrees of comprehension. As suggested by Laufer and Revenhorst Kalovski (2010), while the 98% lexical coverage enables L2 learners to grasp reading materials independently, the 95% provides a minimal understanding of reading materials if learners have some assistance (e.g., a dictionary). Therefore, 95% and 98% have been widely adopted as the lexical coverage cut-off points in vocabulary load studies.

Table 1 presents an overview of the results of the vocabulary load of written texts. It shows that the number of words required to reach 95% and 98% coverages vary depending on types of reading materials, ranging from 2,000 – 3,000 word families for graded readers (Nation, 2006; Webb & Macalister, 2013) to up to 16,000 word families for subject-specialised or academic reading materials (e.g., Kaneko, 2014; Webb & Paripakht, 2015). Literary graded readers require the smallest vocabulary size because their vocabulary is simplified significantly to suit EFL learners' levels (Nation, 2006; Webb & Macalister, 2013). In contrast, subject-specialised or academic reading materials (subject-focused textbooks, English tests, scientific articles and scholarly abstracts) are the most lexically demanding. This is because these materials cover many technical terms and infrequently used words (Hsu, 2014). To the best of our knowledge, all of the previous research on the vocabulary load of written texts has been conducted on human-written materials. Given the potential of ChatGPT in generating written texts, further research on ChatGPT-produced texts allows teachers to determine the difficulty of these sources and thus select suitable reading materials for L2 English learners.

Table 1

The Number of Words Plus Complementary Wordlists to Reach 95% and 98% Coverage

Genres	Studies	95%	98%
Literary graded readers	Nation (2006) Webb & Macalister (2013)	2,000	3,000
TED transcripts	Hsu (2020)	3,000	5,000
Non-literary graded reading materials (news)	Wen & Yu (2025)	3,000	6,000
New Concept English textbooks	Yang & Coxhead (2022)	3,000 – 4,000	5,000 – 6,000
Chinese EFL textbooks (reading sections)	Sun & Dang (2020)	3,000	7,000
Online news (VOA news)	Hsu (2019)	4,000	6,000
Newspaper	Nation (2006)	4,000	8,000
Novels	Nation (2006)	4,000	9,000
Texts for English L1 speakers	Webb & Macalister (2013)	5,000	10,000
Popular science news	Yu & Wen (2025)	5,000	10,000
Engineering textbooks	Hsu (2014)	5,000	10,000
TOEFL	Webb & Paripakht (2015)	6,000	12,000
CANTest	Kaneko (2014)	6,000	14,000
Scientific articles	Huynh Le & Tan Ha (2023)	6,000	14,000 – 15,000
Scholarly abstracts	Nguy & Ha (2022)	7,000	15,000 – 16,000

Previous studies also found that the texts from the same kinds of discourse but targeting different groups of learners may have different vocabulary loads. Wen and Yu (2025) analysed non-literary graded reading materials for students from Grade 1 to Grade 12 levels in the US and found that the lexical knowledge required for comprehension rises progressively from the lowest levels (levels 2 and 3) to the highest levels (11-12). Similarly, Yang and Coxhead (2022) investigated the vocabulary load of a set of New Concept English textbooks for Chinese EFL secondary and university students. They found that the vocabulary size required for both the 95% and 98% coverage of the advanced-level textbooks was 1,000-word families greater than that for the intermediate-level textbooks. By contrast, Sun and Dang's (2020) analysis of a set of general EFL textbooks for high school students in China revealed that written texts of lowest graded levels (Senior 1) are the most lexically demanding (3,000 word families for the 95% coverage and 12,000 word families for the 98%) while the Seniors 2 and 3 require the same vocabulary sizes for reading comprehension (3,000 word families for the 95% coverage and 6,000 word families for the 98%). Together, these findings indicate that while some textbooks did a good job of controlling the vocabulary load within the target learners' level, others did not. To assist EFL teachers in selecting suitable educational resources, Nation suggested the vocabulary sizes of learners at different CEFR levels (see Table 2) for teachers' reference. As many reading resources fail to align with learners' lexical knowledge, it is vitally important to check the lexical demand of all the potential learning materials, including ChatGPT-generated texts, a potential reading source, across different proficiency bands before using.

Table 2
Vocabulary Sizes at Six CEFR Levels (Adapted from Nation's Estimation)

Level	Suggested vocabulary size
C2	7,000-9,000 words
C1	5,000-6,000 words
B2	4,000 words
B1	2,000-3,000 most frequent high frequency words
A2	The most frequent 1000-word families
A1	120 words and phrases from the survival vocabulary

2.4 Occurrence of high, mid, and low-frequency words in written texts

According to Schmitt & Schmitt (2014), vocabulary can be classified into high, mid, and low frequency words. High-frequency words belong to Nation's (2012) 1st-3rd 1,000 BNC/COCA word lists. Mid-frequency words are those from the 4th to the 9th 1,000-word families, and low-frequency words are those beyond the 9th 1,000 word-families. Previous studies on vocabulary load of written discourse (e.g.,Nation, 2006; Hsu, 2020) have shown that high-frequency words constitute the largest percentage of lexical coverage. As knowledge of high-frequency words could improve comprehension significantly, these words should be the starting point of vocabulary teaching and learning in English language education (Nation, 2006; Webb & Nation, 2008). Mid-frequency words are also important for L2 learners because they allow readers to achieve a reasonable comprehension of most written texts (Nation, 2006; Schmitt & Schmitt, 2014), especially more authentic and demanding texts (Hsu, 2020). Unlike high and mid-frequency words, low-frequency words rarely occur in everyday life (Nation, 2013; Laufer, 2013). However, some researchers (Huynh Le & Tan Ha, 2023; Yu & Wen, 2025) advised that some technical terms and low-frequency words related to lesson topics should be taught in advance, so that learners are able to deeply grasp challenging written texts.

Given the relative value of high, mid, and low frequency words for L2 students, a number of studies on vocabulary load have implicitly and explicitly shown the occurrence of these words in learning materials (see Table 3). It can be seen that the vast majority of reading texts are generally high-frequency words (more than 80%) while mid-frequency words account for the second percentages (from around 3% to 8%), followed by the figures for low-frequency words (under 3%). In addition, according to Table 3, subject-centred and scientific materials have fewer high-frequency but more low-frequency words than the general reading sources. It is not surprising because academic written discourse requires learners to learn many technical and rarely used words in their specific areas and therefore is considered to be more challenging for readers than general discourse (Yu & Wen, 2025). As the coverage of high, mid, and low frequency words may indicate the difficulty of written texts, it is useful to examine the occurrence of these words as part of the evaluation of potential learning resources for EFL learners.

Table 3
Coverage of High, Mid and Low-Frequency Words in Written Texts (%)

Genres	Studies	High-frequency words	Mid-frequency words	Low-frequency words
Literary graded readers	Nation (2006) Webb & Macalister (2013)	94.21	1.18	
TED transcripts	Hsu (2020)	94.13	3.43	0.73
Novels	Nation (2006)	91.23	4.94	
Online news (VOA news)	Hsu (2018)	87.36	4.43	0.74
Texts for English L1 speakers	Webb & Macalister (2013)	88.22	4.68	
Popular science news	Yu & Wen (2025)	87.44	5.71	1.28
Engineering textbooks	Hsu (2013)	88.63	6.69	2.11
Scientific articles	Huynh Le & Tan Ha (2023)	85.65	6.85	2.09
Scholarly abstracts	Nguy & Ha (2022)	84.33	8.16	2.81

2.5 Present study and research questions

The literature review has shown that ChatGPT may be a potential tool to develop reading materials for extensive reading programmes. However, no studies have evaluated the appropriateness of ChatGPT-generated texts for extensive reading from the perspective of vocabulary load and occurrences of high, mid, and low-frequency words. The present study will address this gap by (a) calculating the number of words required for the comprehension of ChatGPT-generated texts and (b) examining the proportions of high, mid and low-frequency words in these reading materials. In particular, it will find the answer to the following research questions:

1. How many word families are needed to reach 95% and 98% of ChatGPT-generated texts targeting each CEFR level from A1 to C2?
2. What is the coverage of high, mid, and low-frequency words in these texts?

3 Methodology

3.1 Corpora

Two sets of corpora were created for the present study (see Table 4). The first set consisted of six corpora. Each corpus represented a CEFR level (A1, A2, B1, B2, C1, or C2). This set used a different reading speed for different CEFR levels. Previous research shows that beginning-level learners read about 70 words per minute (wpm) (Bell, 2001). After the reading speed training for one academic year, learners' reading speed can increase by about 30 wpm (McLean & Rouault, 2017). If learners aim to improve one proficiency level and if they also receive reading speed training every academic year, then when learners move to the next proficiency level, their reading speed may increase by 30 wpm. Therefore, in the first set, we chose 70 wpm, 100 wpm, 130 wpm, 160 wpm, 190 wpm, and 220 wpm to represent the reading speed for learners from A1 to C2, respectively. The size of each corpus represented the amount of reading that learners at a particular CEFR level are able to read for a period of time. As three months is a common length for extensive reading programmes (Nakanishi, 2015), following Dang & Long (2024), the corpus size of each CEFR level was about the amount that learners at that level could read five days a week and 40 minutes a day for three months (corpus size = reading speed × 5 days × 40 minutes × 12 weeks).

Table 4
Corpus Structure and Size

Corpus version	CEFR level	Reading speed	Everyday topics	Concrete situations	Abstract topics	Literary topics	Total size
Set 1	A1	70	42,376	40,361	42,572	42,985	168,294
	A2	100	58,259	60,530	61,194	61,006	240,989
	B1	130	78,344	78,479	78,145	78,499	313,467
	B2	160	97,030	95,132	96,515	96,373	385,050
	C1	190	108,530	114,593	114,393	121,654	459,170
	C2	220	132,970	132,140	132,513	132,936	530,559
	Total						2,097,529
Set 2	A1	150	90,127	90,306	90,702	90,582	361,717
	A2	150	90,120	90,306	90,594	90,178	361,198
	B1	150	90,035	90,280	90,034	90,384	360,733
	B2	150	90,489	90,316	90,176	90,533	361,514
	C1	150	90,574	90,565	90,123	90,940	362,202
	C2	150	90,116	90,600	90,335	90,750	361,801
	Total						2,169,165

Table 5
Number of Texts in the Corpus

Corpus version	CEFR level	Reading speed	Everyday topics	Concrete situations	Abstract topics	Literary topics	Total size
Set 1	A1	70	302	420	384	210	1,316
	A2	100	372	350	270	300	1,292
	B1	130	318	284	202	190	994
	B2	160	186	186	235	222	829
	C1	190	160	160	199	160	679
	C2	220	178	166	191	173	708
	Total						5,818
Set 2	A1	150	1,100	745	412	2,916	1,100
	A2	150	556	507	415	432	1,910
	B1	150	365	321	233	217	1,136
	B2	150	176	176	221	213	786
	C1	150	125	118	148	108	499
	C2	150	120	118	129	119	486
	Total						5,917

In terms of the topics of the reading materials, the CEFR illustrative descriptor scales specify that learners are expected to read "from everyday topics (for example hobbies, sports, leisure activities, animals) and concrete situations to a full range of abstract and literary topics" as a leisure activity (Council of Europe, 2020, p. 58). Therefore, each of the six corpora had four sub-corpora representing four kinds of topics that learners at all CEFR levels may encounter when reading: (a) everyday topics (e.g., daily routine), (b) concrete situations (e.g., a visit to a doctor), (c) abstract topics (e.g., happiness), and (d) literary topics (e.g., a classic story). The size of the four sub-corpora was similar, so that the results were not biased towards one of the specific types of topics. The size of each sub-corpora was calculated by dividing the total corpus size by four.

Similar to the first set of corpora, the second set also consisted of six corpora. It had a similar structure as the first set. That is, each corpus represented each CEFR level and had four equally sized topical sub-corpora. The only difference is that this set used the same reading speed (150 wpm, following the recommendation by Nation (2014) for all CEFR levels. Therefore, the corpus size of each CEFR level for the second set was about 360,000 tokens (=150 wpm × 5 days × 40 minutes × 12 weeks). Examining two sets of ChatGPT-generated texts which reflect different and similar reading speed would provide a thorough evaluation of the potential of ChatGPT as a tool for generating extensive reading materials.

The reading materials for the corpora were generated by the second author with the 4o-mini model in ChatGPT from February to May 2025 through two steps. First, ChatGPT was prompted to generate a topic list. The prompt was: "*List [number, e.g., 100] specific everyday topics/concrete situations/abstract topics/literary topics for CEFR A1/A2/B1/B2/C1/C2 English language learners*". Second, ChatGPT was prompted to generate the reading materials one by one according to the topic list: "*Write a text for CEFR A1/A2/B1/B2/C1/C2 English language learners on an everyday topic/concrete situation/abstract topic/literary topic: [the specific topic]*". For example, "*Write a text for CEFR A1 English language learners on an everyday topic: A trip to the local library*". The texts generated by ChatGPT were then copied, pasted and saved as a single text file. Emojis as well as introductory (e.g., "*Here is a text on...*") and closing statements (e.g., "*Let me know if you'd like a different topic*") in ChatGPT responses were removed because they were not part of the reading material. The RANGE programme (Heatley et al., 2002) was used to check the corpus size from time to time. When more specific topics were needed, ChatGPT was prompted to list several more topics and generate more texts, until the target corpus size of each CEFR level was met. Table 5 shows the number of texts in each corpus.

3.2 Analysis

To answer the two research questions, the first author analysed the whole corpus and its subcorpora with the RANGE programme (Heatley et al., 2002). Nation's (2012) BNC/COCA wordlists were used as reference vocabulary lists. These lists consist of twenty-five 1,000 wordlists ranked from the most frequent to the least frequent word families extracted from two of the largest corpora in the world (BNC and COCA). In addition to the most widely-used 25,000-word families, Nation's lists also include additional wordlists of proper nouns, marginal words, compounds, and abbreviations. To begin with, we conducted the preliminary analysis on all the text files as a whole. Off-list words were then cleaned up and classified into relevant wordlists following Dang and Webb's (2014) guidelines.

Firstly, a range of proper nouns (e.g., *Instagram*), transparent compounds (e.g., *sandcastles*) and abbreviations (e.g., *FOMO*) were mistakenly put into NOT IN THE LIST by the RANGE programme. As a result, 5,892 tokens of proper nouns (0.14%), 2,166 tokens of transparent compounds (0.05%) and 942 tokens of abbreviations (0.02%) were manually recategorised by the first author into Nation's supplementary lists of proper nouns, compounds and abbreviations, respectively. The coverage provided by items in lists of proper nouns, compounds, abbreviations and marginal words was added in the cumulative coverage of the corpora because readers may face little difficulty in understanding these words in the texts (Nation & Webb, 2011).

Secondly, there are 17,095 hyphenated words (e.g., *well-being*) accounting for 0.4% of the corpus. For some words (e.g., *well-known*), the hyphen was deleted so that the smaller parts (e.g., *well*, *known*) could join and become a single word (e.g., *wellknown*). Only the words having their joined format already existing in the Nation's (2012) BNC/COCA wordlists could be classified in this way. Otherwise, the hyphens would be replaced with spaces (e.g., *problem-solving* becomes *problem solving*), and the hyphenated words would be categorised according to their smaller items (e.g., *problem*, *solving*).

Thirdly, the NOT IN THE LIST includes 4,054 words deriving from certain base words. These items, taking up 0.1 % of the corpus, were then added to the relevant wordlists containing their roots. For example, *influencer* was added to the wordlist that contains its base word *influence*. Finally, the RANGE programme was unable to deal with 1061 items (0.03% of the corpus) containing diacritical marks (e.g., *café*), so these marks had to be removed.

After the cleaning-up stage, the second analysis was performed on the subcorpora, each of which includes texts ranging from A1 to C2, with either the same or different speed. The data obtained from RANGE was put into Excel files for subsequent checking and calculation. For Research Question 1, the percentages of every 1,000-word family were counted cumulatively to reach 95% and 98% cut-off points. For Research Question 2, the coverages of high, mid and low-frequency words were determined by the total percentage of word families in three groups: the 1st – 3rd 1,000 BNC/COCA word lists, the 4th – 9th 1,000 BNC/COCA word lists and beyond the 9th 1,000 BNC/COCA word lists, respectively.

4 Results

4.1 RQ1: How many word families are needed to reach 95% and 98% of ChatGPT-generated texts targeting each CEFR level?

Table 6 presents the number of words needed to reach 95% and 98% coverage of each ChatGPT-generated corpus (see Appendices 1 and 2 for the detailed lexical profiles of each corpus). The results are consistent for two kinds of reading speed. The A1, A2 and B1 corpora were the least lexical demanding. Regardless of the reading speed, the most frequent 2,000-word families and 3,000-word families were required to achieve 95% and 98% coverage of the A1 corpus. The same number of word families is needed in the A2 and B1 corpora. However, it is important to note that the most 2,000-word families provided higher coverage in the A1 corpus (97.49%) than the A2 corpus (97.05%), which was higher than the B2 corpus (96.37%). This indicates that despite requiring the same number of words for the 95% lexical coverage, the A1 texts tend to be less lexical demanding than the A2 texts, which was less demanding than the B1 texts.

Table 6
The Number of Word Families Needed to Reach 95% and 98% Coverage of Each ChatGPT-Generated Corpus

CEFR level	**95% coverage**		**98% coverage**	
	Different speed	**Same speed**	**Different speed**	**Same speed**
A1	2,000	2,000	3,000	3,000
A2	2,000	2,000	3,000	3,000
B1	2,000	2,000	3,000	3,000
B2	3,000	3,000	4,000	4,000
C1	3,000	3,000	5,000	5,000
C2	4,000	4,000	6,000	6,000

Compared to the A1, A2 and B1 corpora, the B2 corpus was more lexical demanding, requiring the most frequent 3,000 word families for the 95% coverage and the most frequent 4,000 word families for the 98% coverage. The C1 corpus was slightly more demanding than the B2 corpus. Although 3,000 word families were also needed to reach 95% coverage of the C1 corpus, for the 98% coverage, a larger vocabulary was required (5,000 word families). Finally, the C2 corpus was the most demanding, requiring the most frequent 4,000 word families and the most frequent 6,000 words for the 95% and 98% coverage, respectively.

4.2 RQ2. What is the coverage of high, mid, and low frequency words in ChatGPT-generated texts?

The distribution of high, mid, and low-frequency words is illustrated in Tables 7, 8 and 9 (see Appendices 3 and 4 for the detailed coverage of each 1,000 word family band). Table 7 presents the coverage of high-frequency words in texts of six CERF levels at different and same reading speed, respectively. These words consistently accounted for the largest percentage in each corpus (more than 90% of all the texts). Regardless of reading speed, the B1 corpus always had the highest coverage of high-frequency words (nearly 97%). The ranking of the other corpora in terms of lexical coverage of high-frequency words aligned fairly well with the CEFR levels. The A1 and A2 corpora had higher coverage (more than 96%) than the B2 corpus (more than 95%), which had higher coverage than the C1 corpus (about 94%), and the C2 corpus had the lowest coverage (nearly 92%).

Table 7

Coverage of High-Frequency Words in ChatGPT-Generated Texts (%)

Corpus	**Different speed**	**Same speed**
B1	96.61	96.73
A1	96.35	96.32
A2	96.20	96.41
B2	95.31	95.35
C1	93.95	94.0
C2	91.74	91.67

Table 8

Coverage of Mid-Frequency Words in ChatGPT-Generated Texts (%)

Corpus	**Different speed**	**Same speed**
C2	5.46	5.45
C1	3.64	3.72
B2	2.43	2.40
A1	1.75	1.84
A2	1.54	1.52
B1	1.34	1.31

Table 8 illustrates the lexical coverage of mid-frequency words at different and same reading speed, respectively. The results were consistent between the two kinds of reading speeds. Mid-frequency words accounted for 1.31% to 5.46% of each corpus. Similar to high-frequency words, the ranking of lexical coverage for mid-frequency words was fairly in line with the CEFR level. The C2 corpus had the highest coverage (more than 5%). The C1 corpus ranked second (nearly 4%) and the B2 corpus ranked third (more than 2%). Next came the A1 and A2 corpora (nearly 2%). The B1 corpus ranked last with a lexical coverage of more than 1%.

Table 9 presents the coverage of low-frequency words in each corpus. This word group took up less than 0.5% coverage of texts at different CERF levels. As with high and mid-frequency words, the ranking of the coverage for low-frequency words was largely consistent with the CEFR levels. The C2 corpus had the highest coverage (about 0.4%), which was higher than the C1 corpus (about 0.3%), which was higher than the B2 corpus (about 0.2%), which was higher than the A2 corpus (about 0.1%), and the A1 corpus had the lowest percentage of low-frequency words (0.07%). The only exception is the B1 corpus, which had lower coverage than the A2 corpus.

Table 9
Coverage of Low-Frequency Words in ChatGPT-Generated Texts (%)

Corpus	**Different speed**	**Same speed**
C2	0.43	0.43
C1	0.31	0.32
B2	0.20	0.21
A2	0.11	0.10
B1	0.08	0.08
A1	0.07	0.07

5 Discussion

The first research question is about the vocabulary size needed to reach 95% and 98% coverage of ChatGPT-generated texts for learners at different CEFR levels. In answer to this question, irrespective of the reading speed, the number of word families needed to reach 95% and 98% lexical coverage increased from 2,000-3,000 word families for texts targeting the A1, A2, and B1 levels to 3,000-4,000 word families for texts targeting the B2 levels, 3,000-5,000 word families for text targeting the C1 level, and 4,000-6,000 word families for text targeting the C2 level. It means that ChatGPT-generated texts for the higher CEFR levels tend to be more lexically demanding than those for the lower levels. Together, these findings indicate that ChatGPT can generate texts for learners with different English language proficiency levels. The current study is in line with Milton's (2009) recommendation that L2 learners need to know more words if they would like to comprehend texts of higher CEFR levels. It is also consistent with Uchida's (2025) finding that on the surface, ChatGPT-developed texts vary by level, with A1 and A2 texts being simpler than B2, C1 and C2 texts. However, the results are somehow contrary to the research carried out by Ramadhani et al. (2024). They found that texts generated by ChatGPT can be distinguished across six CEFR levels according to word count and length of sentences, but there is no significant difference among texts in terms of lexicon-grammatical features (lexical density, lexical variation and grammatical structures). This implies that the criteria for evaluating text difficulty should be given full attention when teachers use this AI tool to compose learning materials. Moreover, our finding suggests that the difficulty of ChatGPT-generated texts can be distinguished more clearly according to CEFR levels if 98% coverage is set as the learning goal. Specifically, there is little disparity in the number of words required for the minimal comprehension of B2 and C1 texts (the most frequent 3,000 word families). By contrast, a larger vocabulary size is required to achieve optimal comprehension of C1 texts than B2 texts, with 4,000 word families for the former and 3,000-word families for the latter. Therefore, the level of learners' expected text comprehension (minimal or optimal) should be taken into consideration when ChatGPT is employed in developing reading materials.

While the findings of the vocabulary load analysis suggest that ChatGPT may be able to generate texts for learners with different English language proficiency levels, they also revealed that texts generated by ChatGPT do not always match the vocabulary sizes of learners at certain CEFR levels. The number of word families needed for the optimal comprehension of the B1, B2, and C1 texts is relevant to the vocabulary size of learners at these levels as estimated by Nation. However, the number of known words for optimal comprehension of ChatGPT-generated texts targeting the A1 and A2 levels (3,000 word families) was much larger than the vocabulary sizes of A1 and A2 learners as estimated by Nation (120 words and phrases and 1,000 word families). In the meantime, the number of words for optimal comprehension of ChatGPT-generated texts targeting the C2 level (6,000 word families) was smaller than the vocabulary size of C2 learners as estimated by Nation (7,000-9,000 word families). These findings indicate that from the vocabulary load perspective, ChatGPT may be able to generate extensive reading materials for B1, B2, and C1 learners. However, it is probably incapable of generating materials for students at lower language proficiency levels (A1 and A2) or higher language proficiency level (C2). This finding contradicts Uchida's (2025) findings. When comparing her ChatGPT-generated texts with the CEFR framework, she discovers the different patterns when A1 and A2 texts are too simple, whereas B2, C1 and C2 texts are overly complicated. The difference between the current study and Uchida's may be because the size of Uchida's corpus (300 texts, around 47,000 words) was smaller than those used in the present study, which may limit the topics of her texts. In particular, Uchida's A1 and A2 texts are mainly restricted to personal experiences and everyday life, whereas texts with higher CERF levels revolve around more abstract topics. In contrast, in our study, a far wider range of topics appeared in all six CEFR levels. This suggests that further checking and modification should be performed on ChatGPT-generated texts to suit learners' language competence. The differences in topics are highly relevant to prompts. The prompts may have affected the discrepancies in findings with previous studies. The present study prompted ChatGPT to generate four types of topics: everyday topics, concrete scenarios, abstract topics, and literary topics. Each type of topic took up about 25% of the corpus at each CEFR level. However, Ramadhani et al. (2023) did not mention the prompt and only reported the topics when the topics matched the CEFR standard. Their results reported that three texts at A1 or A2 were about family, hobbies, and personal life. Three texts on B1 or B2 were on finance, sport, and technology. CI and C2 texts were all about astronomy and computer science. In Uchida (2025), over 30% of the texts were on daily life (56 out of 180), followed by topics on AI (20 out of 180) and environment (19 out of 180). In other words, while Ramadhani et al. (2023) and Uchida (2025) focused on several specific topics, the topics in the present study were broader and more diverse because of the prompts (i.e., we specified the types of topics needed in the prompts).

The second research question concerns the coverage of high-, mid-, and low-frequency words in ChatGPT-generated texts for learners with different proficiency levels. In answer to this question, regardless of reading speed and targeted CEFR levels, high-frequency words always made up the largest proportion of ChatGPT-generated texts, followed by mid- and then low-frequency words. This proportion reflects the typical lexical profile of texts produced by humans (Nation, 2022). Moreover, ChatGPT-generated texts targeting higher CEFR levels had greater percentages of low and mid-frequency words, but a lower percentage of high-frequency words than texts targeting lower CEFL levels. According to Nation (2006), texts with more high-frequency words and fewer low-frequency words can be more easily understood. Therefore, this finding suggests that ChatGPT are relatively good at generating texts aligning with L2 learners' language proficiency levels.

However, our analysis of high, mid, and low-frequency words also revealed that texts generated by ChatGPT did not always match the target level. First, texts targeting the B1 level should have a moderate level of difficulty in terms of the high, mid, and low-frequency word proportion. However, among the corpora targeting different CEFR levels, the B1 corpus had the largest percentage of high-frequency words, the lowest percentage of mid-frequency words, and the second lowest percentage of low-frequency words. Moreover, texts targeting the A2 level should have lower percentage of high-frequency words and lower percentage of mid-frequency words than those targeting the A1 level. However, our finding shows that the ChatGPT-generated texts for the A2 level had a higher percentage of high-frequency words (same reading speed) and lower percentage of mid-frequency words (both kinds of reading speed).

Taken together, the results of the vocabulary load and high, mid, and low-frequency analysis provide useful insights into the potential of ChatGPT in generating texts for extensive reading programmes. ChatGPT was able to generate texts whose lexical demand and proportion of mid- and low-frequency words increased, while the proportion of high-frequency words decreased according to language proficiency level. This indicates that it may be a useful tool for teachers to generate graded reading materials for learners at different language proficiency levels. However, the fact that the lexical demand of texts generated by ChatGPT did not always match the vocabulary sizes of learners at certain CEFR levels and the occurrences of high, mid and low frequency words in texts targeting certain CEFR levels did not always align with the ranking of the CEFR levels indicate that ChatGPT-generated text is not perfect. Further checking from humans is needed to ensure that ChatGPT generated text is relevant to the target learners.

The present study has several limitations, which can be explored further in future studies. First, it used ChatGPT 4o-mini to generate the materials because it was the most recent version at the time the project was conducted. Future studies can use other GenAI tools or a more advanced version of ChatGPT to test the generalisability of our findings. Second, this study used word families as the counting unit for analysis. Future research using other lexical units (e.g., lemmas) would provide further insights.

This study has several pedagogical implications related to the use of ChatGPT to generate extensive reading materials for L2 learners. To begin with, the findings indicate that ChatGPT has the capability to generate differentiated reading materials targeting learners of different proficient levels. As ChatGPT has the capacity to generate large amounts of reading output within a short time for free, teachers can save a great deal of their time and efforts in preparing extensive reading materials. Therefore, they should consider using ChatGPT to assist in the design of these materials. However, the lexical discrepancy in the ChatGPT-generated texts for some CEFR levels, as found in this study, suggests that before introducing these texts to their students, teachers should carefully check and evaluate these texts to ensure they are relevant to their students' levels. They can use Nation's figures as the vocabulary standard for CEFR levels and check the lexical demand of ChatGPT-generated text with RANGE or Lextutor to see whether the demand of the text matches the vocabulary size needed for a certain CEFR level. If there are mismatches, they should tailor the texts further by instructing ChatGPT to revise the texts to incorporate words at either higher or lower frequency levels so that the lexical difficulty of the text is appropriate for learners at a particular CEFR level (see Dang, 2022 for further instruction on how to adapt vocabulary in texts).

6 Conclusion

This corpus-driven study investigates the lexical profile of ChatGPT-generated texts across six CEFR levels. The findings indicate that ChatGPT is useful for creating learning materials for learners with different proficiency. In particular, ChatGPT can differentiate texts according to learners' language abilities to some extent. First, texts at lower levels (A1, A2 and B1) are less lexically demanding than those for higher levels (B2, C1 and C2) regardless of reading speed. Second, in all cases, high-frequency words consistently constitute the largest percentage, followed by mid-frequency words and then low-frequency words. This pattern aligns with the lexical profile of texts written by humans (Nation, 2013). However, the findings of the present study also indicate that texts of several levels do not adhere to CEFR standards. Specifically, A1 and A2 texts generated by ChatGPT require the same vocabulary sizes as B1 texts, whereas the figures for C1 are slightly smaller than expected. Given the shortcomings of ChatGPT in generating texts, teachers and learners should check the lexical features of these sources carefully before applying them into practice.

Appendices

Appendix 1

Cumulative Coverage of ChatGPT-Generated Texts Targeting Six CEFR Levels at Different Reading Speed (%)

Vocabulary levels	A1	A2	B1	B2	C1	C2
1,000	92.97	91.4	89.13	76.87	70.79	66.74
2,000	97.49[a]	97.05[a]	96.37[a]	90.08	85.52	81.45
3,000	98.17[b]	98.32[b]	98.56[b]	97.35[a]	96[a]	94.02
4,000	98.71	98.81	99.09	98.52[b]	97.87	96.72[a]
5,000	99.34	99.36	99.47	98.99	98.53[b]	97.74
6,000	99.72	99.63	99.73	99.4	99.03	98.53[b]
7,000	99.82	99.73	99.81	99.58	99.35	99.08
8,000	99.87	99.83	99.87	99.72	99.56	99.36
9,000	99.92	99.86	99.9	99.78	99.64	99.48
10,000	99.95	99.88	99.92	99.82	99.7	99.55
11,000	99.96	99.9	99.93	99.86	99.75	99.65
12,000	99.97	99.92	99.95	99.88	99.8	99.73
13,000	99.99	99.95	99.96	99.92	99.86	99.78
14,000	99.99	99.95	99.97	99.94	99.88	99.8
15,000	99.99	99.95	99.97	99.95	99.9	99.83
16,000	99.99	99.96	99.97	99.95	99.91	99.86
17,000	99.99	99.97	99.97	99.96	99.93	99.89
18,000	99.99	99.97	99.98	99.97	99.94	99.9
19,000	99.99	99.97	99.98	99.97	99.95	99.9
20,000	99.99	99.97	99.98	99.98	99.95	99.9
21,000	99.99	99.97	99.98	99.98	99.95	99.91
22,000	99.99	99.97	99.98	99.98	99.95	99.91
23,000	99.99	99.97	99.98	99.98	99.95	99.91
24,000	99.99	99.97	99.98	99.98	99.95	99.91
25,000	99.99	99.97	99.98	99.98	99.95	99.91

[a] Obtaining 95% coverage

[b] Obtaining 98% coverage

Appendix 2

Cumulative Coverage of ChatGPT-Generated Texts Targeting Six CEFR Levels at Same Reading Speed (%)

Vocabulary levels	A1	A2	B1	B2	C1	C2
1,000	92.84	91.6	89.15	76.89	70.6	66.78
2,000	97.35[a]	97.14[a]	96.37[a]	90.11	85.38	81.44
3,000	98.08[b]	98.35[b]	98.59[b]	97.38[a]	95.94[a]	94.06
4,000	98.63	98.85	99.09	98.53[b]	97.86	96.76[a]
5,000	99.28	99.37	99.46	99	98.54[b]	97.76
6,000	99.69	99.64	99.72	99.4	99.02	98.52[b]
7,000	99.8	99.75	99.8	99.58	99.35	99.09
8,000	99.87	99.84	99.87	99.72	99.58	99.38
9,000	99.92	99.87	99.9	99.78	99.66	99.51

10,000	99.95	99.89	99.92	99.82	99.73	99.59
11,000	99.96	99.91	99.93	99.86	99.78	99.69
12,000	99.97	99.93	99.95	99.88	99.83	99.77
13,000	99.99	99.95	99.96	99.92	99.89	99.82
14,000	99.99	99.95	99.97	99.94	99.91	99.84
15,000	99.99	99.96	99.97	99.95	99.93	99.86
16,000	99.99	99.96	99.97	99.95	99.94	99.89
17,000	99.99	99.97	99.97	99.96	99.96	99.92
18,000	99.99	99.97	99.98	99.98	99.97	99.93
19,000	99.99	99.97	99.98	99.98	99.98	99.93
20,000	99.99	99.97	99.98	99.99	99.98	99.93
21,000	99.99	99.97	99.98	99.99	99.98	99.94
22,000	99.99	99.97	99.98	99.99	99.98	99.94
23,000	99.99	99.97	99.98	99.99	99.98	99.94
24,000	99.99	99.97	99.98	99.99	99.98	99.94
25,000	99.99	99.97	99.98	99.99	99.98	99.94

[a] Obtaining 95% coverage

[b] Obtaining 98% coverage

Appendix 3

Lexical Coverage of Each 1,000 Word Families Level at Different Reading Speed (%)

Vocabulary levels	A1	A2	B1	B2	C1	C2
1,000	91.15	89.28	87.18	74.83	68.74	64.46
2,000	4.52	5.65	7.24	13.21	14.73	14.71
3,000	0.68	1.27	2.19	7.27	10.48	12.57
4,000	0.54	0.49	0.53	1.17	1.87	2.7
5,000	0.63	0.55	0.38	0.47	0.66	1.02
6,000	0.38	0.27	0.26	0.41	0.5	0.79
7,000	0.1	0.1	0.08	0.18	0.32	0.55
8,000	0.05	0.1	0.06	0.14	0.21	0.28
9,000	0.05	0.03	0.03	0.06	0.08	0.12
10,000	0.03	0.02	0.02	0.04	0.06	0.07
11,000	0.01	0.02	0.01	0.04	0.05	0.1
12,000	0.01	0.02	0.02	0.02	0.05	0.08
13,000	0.02	0.03	0.01	0.04	0.06	0.05
14,000	0	0	0.01	0.02	0.02	0.02
15,000	0	0	0	0.01	0.02	0.03
16,000	0	0.01	0	0	0.01	0.03
17,000	0	0.01	0	0.01	0.02	0.03
18,000	0	0	0.01	0.01	0.01	0.01
19,000	0	0	0	0	0.01	0
20,000	0	0	0	0.01	0	0
21,000	0	0	0	0	0	0.01
22,000	0	0	0	0	0	0
23,000	0	0	0	0	0	0
24,000	0	0	0	0	0	0
25,000	0	0	0	0	0	0

Appendix 4

Lexical Coverage of Each 1,000 Word Families Level at Same Reading Speed (%)

Vocabulary levels	A1	A2	B1	B2	C1	C2
1,000	91.08	89.66	87.29	74.86	68.66	64.39
2,000	4.51	5.54	7.22	13.22	14.78	14.66
3,000	0.73	1.21	2.22	7.27	10.56	12.62
4,000	0.55	0.5	0.5	1.15	1.92	2.7
5,000	0.65	0.52	0.37	0.47	0.68	1
6,000	0.41	0.27	0.26	0.4	0.48	0.76
7,000	0.11	0.11	0.08	0.18	0.33	0.57
8,000	0.07	0.09	0.07	0.14	0.23	0.29
9,000	0.05	0.03	0.03	0.06	0.08	0.13
10,000	0.03	0.02	0.02	0.04	0.07	0.08
11,000	0.01	0.02	0.01	0.04	0.05	0.1
12,000	0.01	0.02	0.02	0.02	0.05	0.08
13,000	0.02	0.02	0.01	0.04	0.06	0.05
14,000	0	0	0.01	0.02	0.02	0.02
15,000	0	0.01	0	0.01	0.02	0.02
16,000	0	0	0	0	0.01	0.03
17,000	0	0.01	0	0.01	0.02	0.03
18,000	0	0	0.01	0.02	0.01	0.01
19,000	0	0	0	0	0.01	0
20,000	0	0	0	0.01	0	0
21,000	0	0	0	0	0	0.01
22,000	0	0	0	0	0	0
23,000	0	0	0	0	0	0
24,000	0	0	0	0	0	0
25,000	0	0	0	0	0	0

Statement for the use of AI tools

ChatGPT was used to generate the texts for analysis, as this is the purpose of the article. The use of ChatGPT was described in the "Methodology" section. Other than this, AI was not used in the preparation of this material. The authors wrote, reviewed and edited the content as needed and took full responsibility for the content, accuracy and originality of the article.

References

Abdelhalim, S. M., & Alsehibany, R. (2025). Integrating ChatGPT for vocabulary learning and retention: A classroom-based study of Saudi EFL learners. *Language Learning & Technology*, *29*(1), 1–24. https://doi.org/10.64152/10125/73635

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, *1*(1). https://readingmatrix.com/articles/bell/

Carrell, P. L., & Carson, J. G. (1997). Extensive and intensive reading in an EAP setting. *English for Specific Purposes*, *16*(1), 47-60. https://doi.org/10.1016/S0889-4906(96)00031-2

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*, Council of Europe Publishing: Strasbourg, available at https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4.

Dang, T. N. Y. (2022). Using VocabProfilers to select texts for extensive reading activities. In V. Viana (Ed.), *Teaching English with corpora: A resource book* (pp. 69–73). London, UK: Routledge

Dang, T. N. Y., & Long, X. (2024). Online news as a resource for incidental learning of core academic words, academic formulas, and general formulas. *TESOL Quarterly*, *58*(1), 32-62. https://doi.org/10.1002/tesq.3208

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

Davis, C. (1995). Extensive reading: an expensive extravagance?. *ELT Journal*, *49*(4), 329-336. https://doi.org/10.1093/elt/49.4.329

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.

Day, R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, *14*(2), 136-141. https://www2.hawaii.edu/~readfl/rfl/October2002/day/day.html

Hao, C., Xu, W., Halim, H. B. A., & Hao, M. (2025). AI chatbot-assisted vocabulary learning: Relationships with self-regulation, motivation, and performance among Chinese private college students. *Language Teaching Research*. https://doi.org/10.1177/13621688251352595

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range: A program for the analysis of vocabulary in texts*. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes (New York, N.Y.)*, *33*(Jan), 54–65. https://doi.org/10.1016/j.esp.2013.07.001

Hsu, W. (2019). Voice of America News as Voluminous Reading Material for Mid-frequency Vocabulary Learning. *RELC Journal*, *50*(3), 408–421. https://doi.org/10.1177/0033688218764460

Hsu, W. (2020). Can Ted Talk transcripts serve as extensive reading material for mid-frequency vocabulary learning? *TEFLIN Journal*, *31*(2), 181–203. https://doi.org/10.15639/teflinjournal.v31i2/181-203

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

Huynh Le, N., & Tan Ha, H. (2023). Lexical Demands of Academic Written English: From Students' Assignments to Scholarly Publications. *SAGE Open*, *13*(4). https://doi.org/10.1177/21582440231216292

Kaneko, M. (2014). Is the vocabulary level of the reading section of the TOEFL Internet-based test beyond the lexical level of Japanese senior high school students? *Vocabulary Learning and Instruction*. doi: http://dx.doi.org/10.7820/vli.v03.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology*, *3*(1). https://langedutech.com/letjournal/index.php/let/article/view/48

Laufer, B. (1989) What percentage of text-lexis is essential for comprehension? In: Lauren, C. and Nordman, M. (Eds.), *Special language: From human thinking to thinking machines* (*pp.* 316-323). Multilingual Matters.

Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *TESOL Quarterly*, *47*(4), 867–872. https://doi.org/10.1002/tesq.140

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30. https://doi.org/10.64152/10125/66648

Liu, F., Lu, X., Jin, T., Kang, M., & Zhang, H. (2025). Does ChatGPT simplify texts like expert teachers? Linguistic features of simplified texts. *Reading and Writing*, 1-21. https://doi.org/10.1007/s11145-025-10676-2

McLean, S., & Rouault, G. (2017). The effectiveness and efficiency of extensive reading at developing reading rates. *System*, 70, 92-106. https://doi.org/10.1016/j.system.2017.09.003

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Moons, P., & Van Bulck, L. (2024). Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept. *European Journal of Cardiovascular Nursing : Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *23*(2), 122–126. https://doi.org/10.1093/eurjcn/zvad087

Mugableh, A. I. (2024). The impact of ChatGPT on the development of vocabulary knowledge of Saudi EFL students. *Arab World English Journal*, 265–281. https://dx.doi.org/10.24093/awej/ChatGPT.18

Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, *49*(1), 6-37. https://doi.org/10.1002/tesq.157

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 1–12.

Nation, I. S. P. (2012). The BNC/COCA word family lists. http://www.victoria.ac.nz/lals/about/staff/paul-nation.

Nation, I. S. P., & Waring, R. (2020). *Teaching extensive reading in another language*. Routledge.

Nation, I.S.P. & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Nation, I.S.P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I.S.P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language, 26*(2), 1-16. https://nflrc.hawaii.edu/rfl/item/303

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. http://dx.doi.org/10.1017/9781009093873

Nation, I.S.P. (n.d). *Vocabulary, the CEFR levels, and word family size*. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists/vocabulary-cefr-and-word-family-size/vocabulary-and-the-cefr-docx

Nguy, N. L. Q., & Ha, H. T. (2022). Lexical profile of academic written English revisited: What does it take to understand scholarly abstracts? *SAGE Open*, *12*(3). https://doi.org/10.1177/21582440221126342

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571-1582. https://doi.org/10.1002/tesq.3253

Puripunyavanich, M., & Waring, R. (2024). Investigating the implementation of extensive reading in four Asian countries. *Reading in a Foreign Language*, *36*(1), 1-34. https://hdl.handle.net/10125/67475

Ramadhani, R., Aulawi, H., & Ulfa, R. L. (2023). Readability of Reading Texts as Authentic Materials Issued by ChatGPT: A Systemic Functional Perspective. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *8*(2), 149-168. http://dx.doi.org/10.21093/ijeltal.v8i2.1546

Ramonda, K. (2020). Extensive reading and class readers: The case for no choice. *ELT journal*, *74*(3), 277-286. https://doi.org/10.1093/elt/ccaa017

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal (Boulder, Colo.)*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, *52*(1), 73-89. https://doi.org/10.1002/rrq.152

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System (Linköping)*, *93*, 102279-. https://doi.org/10.1016/j.system.2020.102279

Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, *6*(3). https://doi.org/10.58304/ijts.20240304

Tomlinson, B., & Masuhara, H. (2017). *The complete guide to the theory and practice of materials development for language learning*. John Wiley & Sons.

Uchida, S. (2025). Generative AI and CEFR levels: Evaluating the accuracy of text generation with ChatGPT-4o through textual features. *Vocabulary Learning and Instruction*, *14*(1), 2078. https://doi.org/10.29140/vli.v14n1.2078

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

Webb, S., & Macalister, J. (2013). Is Text Written for Children Useful for L2 Extensive Reading? *TESOL Quarterly*, *47*(2), 300–322. https://doi.org/10.1002/tesq.70

Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes (New York, N.Y.)*, *38*, 34–43. https://doi.org/10.1016/j.esp.2014.11.001

Wen, J., & Yu, H. (2025). The lexical profile of online graded reading materials in English language teaching: A corpus-based study. *Language Teaching Research*. https://doi.org/10.1177/13621688251320465

Yang, L., & Coxhead, A. (2022). A corpus-based study of vocabulary in the New Concept English textbook series. *RELC Journal*, *53*(3), 597–611. https://doi.org/10.1177/0033688220964162

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, *124*, 103351. https://doi.org/10.1016/j.system.2024.103351

Young, J. C., & Shishido, M. (2023a). Evaluation of the potential usage of ChatGPT for providing easier reading materials for ESL students. In *EdMedia+ innovate learning* (pp. 155-162). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/222496/.

Young, J. C., & Shishido, M. (2023b). Investigating OpenAI's ChatGPT potentials in generating Chatbot's dialogue for English as a foreign language learning. *International Journal of Advanced Computer Science and Applications*, *14*(6). http://dx.doi.org/10.14569/IJACSA.2023.0140607

Yu, H., & Wen, J. (2025). Lexical coverage in science popularization discourse: The case of popular science news from Scientific American. *English for Specific Purposes (New York, N.Y.)*, *77*, 45–55. https://doi.org/10.1016/j.esp.2024.10.001

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, 10(3). https://doi.org/10.1016/j.heliyon.2024.e25370

Zhou, J. (2024). *Extensive reading*. Cambridge University Press.

Nhat Ha Nguyen is a recent graduate from the University of Leeds. She has presented at several national and international conferences in Vietnam and the UK. Her research interests include corpus linguistics, vocabulary studies, and TESOL materials. Currently, she is an lecturer at University of Science and Education - The University of Danang, Vietnam.

Xuechun Huang is a PhD candidate at the School of Education, University of Leeds. Her research interests include vocabulary studies, technology-enhanced language learning, and translanguaging. She is a CELTA-qualified English language teacher and a teaching assistant for undergrad and graduate modules at the University of Leeds. She is the Event Coordinator at Open Applied Linguistics. She has won the Best Student Work-in-progress Presentation Award at the BAAL Vocabulary SIG Annual Conference 2024.

Thi Ngoc Yen Dang is an Associate Professor in Language Education at the University of Leeds, UK. Her research interests include vocabulary studies and corpus linguistics. Her articles have been published in various journals such as *Applied Linguistics, Language Learning, Language Teaching, Studies in Second Language Acquisition*, *TESOL Quarterly, Language Teaching Research,* and *System.*