LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 19, No: 1, January – June 2026

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Vocabulary Size of Thai Graduate Students in Different Disciplines and Their Opinions of Its Influences on the Use of AI in English Language Learning

Sumanee Pinweha[a], Sutthirak Sapsirin[b,*]

[a] sumanee.p@chula.ac.th, Chulalongkorn University Language Institute, Chulalongkorn University, Thailand
[b] sutthirak.s@chula.ac.th, Chulalongkorn University Language Institute, Chulalongkorn University, Thailand
[*] Sutthirak Sapsirin, sutthirak.s@chula.ac.th

APA Citation:
Pinweha, S., & Sapsirin, S. (2026). Vocabulary size of Thai graduate students in different disciplines and their opinions of its influences on the use of AI in English language learning. *LEARN Journal: Language Education and Acquisition Research Network, 19*(1), 343-375. https://doi.org/10.70730/CQMZ6223

Received
17/07/2025

Received in revised form
04/11/2025

Accepted
22/12/2025

ABSTRACT

This study investigates the vocabulary size of Thai graduate students across science and non-science disciplines, alongside their opinions on the effects of vocabulary size on their use of AI tools in their English language learning. A total of 217 students from a public Thai university completed the Updated Vocabulary Levels Test (Webb et al., 2017) and engaged in semi-structured interviews. Quantitative data were analyzed using descriptive statistics, revealing that most students had a low vocabulary size. The Mann-Whitney U test showed significantly higher performance among science students at the 1,000-, 4,000-, and 5,000-word levels. Qualitative data from interviews showed that vocabulary size may not influence most types of AI-assisted language learning activities, their trust in it, or their reliance on AI for language learning, except for the choice of prompt language. These findings offer implications for English vocabulary instruction and highlight the

importance of integrating AI-assisted language learning tools with lexical development in higher education contexts.
Keywords: vocabulary, vocabulary test, AI tools, UVLT, graduate students

Introduction

Vocabulary is essential for comprehension and communication, particularly for non-native speakers. Numerous studies have demonstrated a strong correlation between language skills and vocabulary knowledge (Laufer, 1998; Schmitt et al., 2011). This means that a lack of vocabulary knowledge may negatively affect students' ability to use the language, including their reading, listening, writing, and speaking skills. Thus, understanding students' vocabulary size can benefit both teachers and students.

For Thai graduate students, the academic challenges they face may stem from a limited vocabulary. Given the growing evidence that vocabulary size correlates strongly with academic achievement across various disciplines, understanding the vocabulary sizes of Thai students is vital (Mungkonwong & Wudthayagorn, 2017). Previous research has examined the vocabulary size of specific groups (e.g., Mungkonwong & Wudthayagorn, 2017; Schmitt et al., 2011), the relationship between vocabulary size and other factors, as well as comparisons based on gender (e.g., Sukying, 2023). For example, Mungkonwong and Wudthayagorn (2017) used Nation and Beglar's (2007) Vocabulary Size Test (VST) to identify overall vocabulary size of first-year students from four public and three private universities in Thailand. However, few studies have compared the vocabulary sizes of Thai graduate students across different fields of study or identified vocabulary knowledge of individual frequency levels. To bridge this gap, this study aims to examine the vocabulary size of Thai graduate students and compare it between students enrolled in science and those enrolled in non-science fields. This can help teachers better understand students' vocabulary levels and can also be utilized to create effective teaching materials, methods, curricula, and policies aimed at improving vocabulary instruction and English language competence for students of diverse fields.

As vocabulary knowledge is essential for academic success, many students have turned to Artificial Intelligence (AI) technology, particularly since the launch of ChatGPT in 2022, to overcome challenges related to vocabulary and English proficiency (Alhaisoni & Alhaysony, 2017; Chang et al., 2021; Karataş et al., 2024; Meniado, 2023). AI has been shown to support learning, provide corrective feedback, improve accuracy in grammar and vocabulary, reduce language learning anxiety, and foster motivation. However, limited research has been conducted on the use of AI tools,

particularly generative AI such as ChatGPT, in English language learning in higher education (Liu, 2024). In addition, few studies have investigated the factors that may shape learners' use of AI for English language learning. One factor worth investigating is vocabulary size as it serves as a strong indicator of overall language proficiency (Laufer, 1998; Schmitt et al., 2011). Language proficiency, in turn, is an individual factor that can shape learners' engagement with feedback (Zhang, 2020) and their interaction with learning technologies. Previous studies have shown that learners with different English proficiency levels may use and perceive AI tools differently (Guo et al., 2022; Koltovskaia, 2020). However, these studies were conducted prior to the emergence of generative AI and therefore provide limited evidence regarding leaners' engagement with this new technology. Given that vocabulary size reflects learners' language abilities, examining its influence can yield insights into how proficiency-related factors shape the use of AI tools, including generative AI, for English language learning. Therefore, another objective of the present study is to explore this issue through Nation's (2007) four strands framework, in examining AI-assisted language learning activities, and through Zhang and Hyland's (2018) model of student engagement with AI-generated feedback.

Considering the existing gap in studies regarding Thai graduate students' vocabulary size, and their use and perceptions of AI in English language learning, this study aimed to (1) explore the vocabulary size of Thai graduate students, (2) investigate differences between the vocabulary size of Thai graduate students in science and non-science disciplines, and (3) explore their perceptions of how it impacts their use of AI tools in English language learning. The research questions for this study are as follows:

1. What is the vocabulary size of Thai graduate students in science and non-science disciplines?
2. Is there a difference between the vocabulary size of Thai graduate students with science and non-science disciplines?
3. What are the opinions of Thai graduate students in science and non-science disciplines about the effects of their vocabulary size on their use of AI tools in their English language learning?

By examining students' vocabulary size and AI use, this study provides a dual contribution to theory and practice. First, findings about their present lexical knowledge will facilitate the design and implementation of appropriate learning activities by teachers and curriculum designers to more accurately address students' vocabulary needs. Additionally, the results on how vocabulary size influences students' use of AI for English learning will enhance understanding of the potential link between lexical knowledge and technology use in language education. These findings can be seen as a

theoretical contribution by explicitly highlighting learners' engagement with Artificial Intelligence as a language learning tool—thus expanding Nation's (2007) four strands framework to include technology-enhanced learning contexts. Furthermore, these results will clarify how an individual difference—specifically, vocabulary size—affects learner engagement with AI-generated feedback, supporting Zhang and Hyland's (2018) model of learner engagement with feedback. Although findings related to AI use are tentative because of the exploratory nature of the nature of the research question and the study's dependence on interview data, such data can clearly inform teachers and curriculum designers in similar EFL graduate programs about the types of AI-supported learning activities their students most probably engage in, as well as their attitudes toward AI technologies for English learning in the context of widespread AI use. Considered alongside information about students' lexical knowledge, this data is instrumental in guiding pedagogy and policy decisions on how to effectively incorporate AI tools into the design and delivery of vocabulary-learning activities meant to enhance English vocabulary development within EFL higher education settings.

Literature Review

Academic success in higher education has been found to be closely related to language competence, with vocabulary knowledge being the most significant and easily measurable component (Laufer, 1998; Nation & Waring, 1997). However, as discussed in this review, considerably less is known about the vocabulary size of specific populations in high-stakes contexts, such as Thai graduate students. At the same time, the discipline of language learning has recently been invaded by tools based on Artificial Intelligence (AI) (Meniado, 2023; Teng, 2024). According to new research, gains facilitated by these tools seem not to trickle down equally and are heavily dependent on the proficiency levels already acquired by individual learners (Koltovskaia, 2020; Lee et al., 2024). This review integrates these two foci toward an examination of the interaction between vocabulary proficiency and the use of AI tools to learn English among Thai graduate students.

The Role of Vocabulary Size in Academic Performance

Vocabulary knowledge is a core component of language mastery, permeating all receptive and productive skills, and academic achievement for both native and non-native speakers. The acquisition of vocabulary among

native English speakers advances rapidly during early childhood development; by the age of three, children typically possess approximately 1,000 words, increasing to around 4,000 to 5,000 words by the time they reach the age of five. This growth continues into adulthood, with vocabularies exceeding 20,000 words (Nation & Waring, 1997).

For second language learners, certain lexical targets are essential to perform competently on academic tasks. Basic knowledge of high-frequency word families is necessary for basic comprehension—estimated at about 3,000 (Nation & Waring, 1997), while 5,000 words are needed to understand unsimplified texts (novels for teenagers) (Hirsh & Nation, 1992). However, knowledge of between 6,000 and 9,000 word families as well as academic and technical words is necessary for second language learners with academic purposes (Nation, 2022).

There is convincing evidence to support a positive linear relationship between vocabulary size and academic success. Studies have been conducted in very diverse contexts with varied foci, but the fundamental finding has been similar: vocabulary size serves as a direct predictor of academic success. As Laufer (1998) states, extensive vocabulary is required for both understanding and production of advanced academic texts.

Recent research within specific contexts has generally supported this principle. For instance, in Thai EFL contexts, the work of Mungkonwong and Wudthayagorn (2017) has demonstrated that vocabulary size is directly correlated with academic success in English-medium instruction among Thai students. An even more direct connection has been observed in the quality of writing. Students who possess a large vocabulary have been found to produce writing with "wide-ranging and more sophisticated" language to articulate complex ideas (Charnchairerk, 2022, p. 858).

Although numerous studies highlight the significance of vocabulary, they also reveal a substantial gap in existing research. According to Mungkonwong and Wudthayagorn (2017), Thai university undergraduates possess a vocabulary size between 4,200 and 5,900 words. However, there is a notable lack of research concerning the vocabulary size of Thai graduate students. This deficiency is particularly concerning, given that this group is required to meet the most rigorous linguistic demands in thesis writing, research publication, and academic discourse at advanced levels. Addressing this gap necessitates the development of a reliable assessment instrument. The methodologies for vocabulary assessment are discussed as follows.

Understanding Vocabulary Assessment Methods: The Updated Vocabulary Levels Test (UVLT)

Various instruments facilitate the evaluation of students' lexical knowledge in linguistics and language learning. They share the same primary purpose of identifying vocabulary knowledge, but besides this, they differ significantly. For example, Nation and Beglar's (2007) VST elicits a wide range of general lexical knowledge. In contrast, more specific profiling across hierarchical levels is available with the Updated Vocabulary Levels Test (UVLT) developed by Webb et al. (2017). Since it assesses knowledge within specific frequency bands: the 1,000, 2,000, 3,000, 4,000, and 5,000-word levels, information from the UVLT is practically and pedagogically valuable because it enables researchers to identify more accurately which lexical thresholds cause problems for learners.

The UVLT, derived from the VLT initially developed by Nation (1983), uses more recent corpora, specifically the British National Corpus and Corpus of Contemporary American English, and employs a matching format (Webb et al., 2017). Its validity is conceptualized within Messick's (1989, 1995) construct validation framework—thus providing a robust basis across content, substantive, structural, and external aspects for making inferences regarding test scores. The test demonstrates high diagnostic capability, effectively differentiating between various strata of proficiency: person separation estimates of 4.72 and above, along with item strata statistics exceeding 6.85, indicate the test's excellent performance in identifying ability levels within a population. The item difficulty has been well-calibrated so that there is clear progression from the 1,000 to the 5,000-word level, thereby covering a broad spectrum of knowledge (see Smith Jr., 2004, p. 106).

The major limitations of the UVLT should be recognized, even though its psychometric properties have been rigorously validated. As recommended by Webb et al. (2017), "validation of the UVLT is, in fact, an ongoing process" (p. 56), and should be further tested in various contexts. Also, since this test measures receptive vocabulary—not productive control (i.e., speaking or writing ability to actively and appropriately use words)—various aspects of lexical knowledge, such as collocations, word parts, polysemy, and word use at different frequency levels need to be explored (e.g. Webb, 2013). Additionally, the interpretation of results remains challenging, as the vocabulary levels indicated by UVLT test-takers do not directly correspond to their actual vocabulary size. Despite these limitations in profiling diagnostic receptive knowledge at key frequency levels, the UVLT remains a suitable instrument within the context of this research.

As the significance of research concerning the vocabulary size of graduate students takes on greater importance, the educational environment is being reshaped by AI-powered learning tools, with growing research underscoring their potential benefits. The next part of this literature review

will discuss the integration of such AI tools in English language learning, as well as how learners of different proficiency levels engage with these tools.

Use of AI Tools in English Language Learning

To improve vocabulary and other language skills, students have increasingly used a wide range of AI tools. These tools include Automatic Evaluation Systems (AES), which assess students' writing and provide feedback; machine translation (MT) tools like Google Translate; paraphrasing and summarizing tools such as Quillbot; and generative AI like ChatGPT. Research has shown that these tools can facilitate language learning by supporting written production, providing corrective feedback, and expanding lexical knowledge (Alhaisoni & Alhaysony, 2017; Chang et al., 2021; Kurniati & Fithriani, 2022; Meniado, 2023; Teng, 2024).

Among generative AI tools, ChatGPT has attracted attention for its potential to transform English language education as it is able to interpret prompts and generate contextually appropriate responses (Liu, 2024; Meniado, 2023; Solak, 2024). Drawing on Nation's (2007) four strands framework, Meniado (2023) highlighted ChatGPT's contribution to all dimensions of language learning. Firstly, it can generate meaningful input such as individualized learning materials. Secondly, it can help students produce meaning-focused language by providing support and guidance. It also aids students in correcting language errors. Finally, it can offer guided practice, and interactive and context-specific language exercises to improve fluency. Empirical studies further show that ChatGPT can boost learner motivation and engagement (Karataş et al., 2024), and deliver personalized and adaptive learning experiences (Solak, 2024). Also, it can provide clear, immediate feedback that learners find useful for writing (Teng, 2024). Overall, these studies highlight ChatGPT's ability to integrate linguistic, cognitive, and affective dimensions of language learning.

While AI tools offer significant advantages in language learning, they also present certain challenges. Several studies identify students' overreliance on machine translation or generative AI as a problem (Alhaisoni & Alhaysony, 2017; Teng, 2024). This may discourage students from engaging with English texts and reduce opportunities to infer new word meanings independently. It can also inhibit learners' critical thinking, problem-solving, and creativity. In addition, there are issues of accuracy and reliability in AI-generated responses (Chang et al., 2021; Karataş et al., 2024; Meniado, 2023; Teng, 2024). AI feedback has been found to be incomplete, misleading, or overly simplistic. Finally, there are concerns about plagiarism and academic misconduct (Karataş et al., 2024; Meniado, 2023; Teng, 2024). This is due to

AI's ability to generate content without proper source attribution and to complete user writing tasks.

In summary, the literature suggests that AI can both enhance and hinder language learning. However, relatively few studies have explored how EFL learners in higher education integrate these tools into their English language learning. Understanding this issue is crucial for incorporating AI effectively into vocabulary learning.

Learners' Proficiency Levels and Engagement with AI Tools

Several studies have examined how learners of different English proficiency levels use AI tools in language learning (e.g., Koltovskaia, 2020; Lee et al., 2024). Findings indicate that proficiency level influences both the degree of reliance on AI tools and the ability to critically engage with their feedback. Lower proficiency learners tend to depend heavily on machine translation tools, which may limit opportunities for independent writing practice (Lee et al., 2024). They also tend to accept Grammarly's feedback uncritically, possibly due to limited linguistic awareness (Koltovskaia, 2020; Lee et al., 2024). In contrast, advanced learners are more selective in their use of AI support—reading English texts directly, evaluating Grammarly's suggestions critically, and using AI tools to refine vocabulary, grammar, and overall writing quality. Accordingly, Lee et al. (2024) recommend that beginner learners use AI tools cautiously to avoid overreliance, while advanced learners may benefit more fully from AI-assisted feedback and writing analysis.

As discussed, AI tools have been increasingly used among English language learners and many studies have examined how learners at different proficiency levels—both low and high—engage with AI tools such as machine translation and Grammarly. However, relatively few have explored learners' engagement with generative AI tools, which have recently been developed, representing a newer and more interactive form of technology-assisted language learning. Understanding how learners interact with these tools is particularly important because their use can influence vocabulary development, writing skills, and overall language proficiency. In addition, while previous studies have focused on general language proficiency, this study uses vocabulary size as an indicator of learners' proficiency. Vocabulary knowledge is a strong predictor of overall language ability (Laufer, 1998; Schmitt et al., 2011) and plays a central role in learners' capacity to comprehend and produce language (Nation, 2022). Using vocabulary size provides a measurable and comparable way to examine differences in how learners engage with AI tools, allowing for a more precise analysis of learner–tool interactions.

Theoretical Framework

To examine Research Question 3, the study drew on two complementary frameworks. The first, Nation's (2007) four strands framework, was applied to analyze and describe the influence of vocabulary size on participants' AI-related language learning activities. The framework categorizes language learning activities into meaning-focused input, meaning-focused output, language-focused learning, and fluency development. The meaning-focused input strand involves learning through listening and reading, such as extensive reading or listening to stories. The meaning-focused output strand refers to productive use of language through activities like conversations or diary writing. The language-focused learning strand emphasizes deliberate learning of linguistic features and learning strategies, such as studying vocabulary, grammar, and dictionary use. Finally, the fluency development strand focuses on improving speed and automaticity in understanding or producing familiar content across the four language skills, through activities such as skimming, scanning, and timed writing exercises.

The second framework, Zhang and Hyland's (2018) model of student engagement with feedback on L2 writing, was used to explore how learners interact with and respond to AI feedback. Although originally developed for L2 writing, this model is relevant to AI-assisted language learning because AI tools often provide various forms of feedback (Kurniati & Fithriani, 2022; Liu, 2024; Meniado, 2023; Solak, 2024). The model proposes that learner and contextual factors—such as language proficiency, learning strategies, and feedback source—influence students' engagement with feedback (Zhang, 2020; Zhang & Hyland, 2018). Engagement is conceptualized as comprising three dimensions: behavioral (revision activities and time spent on revision), affective (emotions and attitudes toward feedback), and cognitive (use of revision strategies). Together, these frameworks provide a comprehensive lens for exploring how learners with different vocabulary sizes engage with AI tools in English language learning.

Research Methodology

Research Design

This study employs both quantitative and qualitative methodologies to examine the vocabulary size of Thai graduate students and their perceptions of how vocabulary knowledge influenced their use of AI in English language learning (Creswell, 2012). The UVLT provided vocabulary size data for science and non-science students, serving to address the first two

research questions. Descriptive statistics (mean and SD) and the Mann-Whitney U test were used for quantitative data analysis. Perceptions of vocabulary size's impact on AI usage were provided qualitatively in semi-structured interviews. Table 1 presents the research questions, the types of data collected for each question, and the corresponding data analysis methods.

Table 1

Summary of Research Questions, Data, and Data Analysis

Research Question	Data	Data Analysis
1. What is the vocabulary size of Thai graduate students in science and non-science disciplines?	UVLT scores	Descriptive analysis
2. Is there a difference between the vocabulary size of Thai graduate students in science and non-science disciplines?	UVLT scores	Mann-Whitney U test
3. What are the opinions of Thai graduate students in science and non-science disciplines about the effects of their vocabulary size on their use of AI tools in their English language learning?	Interview data	Content analysis

Ethical Approval

Ethical approval from the university's Research Ethics Review Committee for Research Involving Human Subjects was obtained prior to conducting the research.

Research Context, Sampling, and Participants

This research was conducted at a large university in Thailand, offering graduate programs across the disciplines of science, arts, and social sciences, both in Thai and English. The requirement of English proficiency for graduation was strictly enforced, with criteria that may include IELTS benchmarks or the successful completion of an English course.

Participants were recruited by purposive sampling of volunteers since this is the best way of deliberately choosing people with particular characteristics that are of interest to the research questions (Cresswell & Plano Clark, 2018). The primary purpose of this study was to make a specific comparison between two different academic groups, that is, graduate students in science and non-science disciplines, in terms of their English vocabulary

levels. Thus, purposive sampling can assist in recruiting an adequate and relatively balanced number of respondents from both categories.

Criteria for recruitment were based on three factors: 1) language background (Thai), 2) program level (Master's or Doctoral in Thai programs), and 3) academic discipline (science or non-science). There were 217 graduate students who participated in the UVLT, consisting of 109 science students and 108 non-science students. Their demographic information is presented in Table 2.

Table 2

Demographic Information of UVLT Test Takers

Discipline	Gender		Educational Level	
	Male	Female	Master's	Ph.D.
Science	36.7%	63.3%	76.1%	23.9%
Non-Science	39.8%	60.2%	93.5%	6.5%

Purposive sampling was used for the selection of interview participants. The criteria for recruitment were as follows: 1) UVLT scores at any of the following levels: 1000, 2000, 4000, and 5000; and 2) willingness to partake in the interview. Participants were categorized into two groups based on vocabulary size, a classification adapted from Webb et al. (2017) with learners having a vocabulary of 2,000 words or less considered low-proficiency. Therefore, students scoring at the 1,000- and 2,000-word levels may be categorized as part of the low-vocabulary or low-proficiency group (L), whereas those at the 4,000- and 5,000-word levels can be considered to fall into the high-vocabulary or high-proficiency group (H). Students at the 3,000-word level were excluded to maximize the distinction between the two groups. In total, 21 participants— ten from science and eleven from non-science disciplines—were selected for interviews. For analytical clarity and subsequent reporting, participants were coded (e.g., LSci 1 = low-vocabulary science student 1; HNonSci 1 = high-vocabulary non-science student 1). The number of interviewees at each score level is indicated in Table 3.

Table 3

Number of Interviewees at Each Score Level

Vocabulary Level	Science	Non-Science
1,000	2	2
2,000	3	4
4,000	2	3
5,000	3	2
Total	10	11

Research Instruments

The UVLT (Webb et al., 2017) was administered using a paper-based test, while responses to the test items were submitted via Google Forms to facilitate participant convenience and enable efficient scoring (Figure 1). The content validity of the translated instructions and demographic questions was confirmed through expert review (IOC = 1.00).

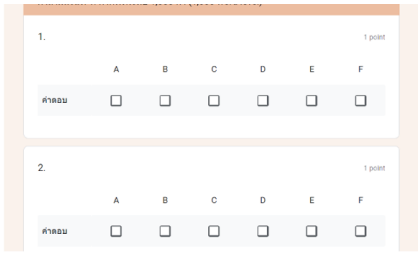
The UVLT was piloted with 30 students, and reliability was estimated using Cronbach's alpha. The reliability values for each vocabulary level ranged from 0.82 to 0.94, indicating good to excellent internal consistency. In the main study, reliability values for each vocabulary level ranged from 0.84 to 0.89, confirming the test's good to excellent reliability.

Figure 1

The UVLT Sample Test Item and an Answer Sheet on Google Forms

ตัวเลือก A-F สำหรับโจทย์ข้อ 1-3	A. game	B. island	C. mouth	D. movie	E. song	F. yard
1. land with water all around it						
2. part of your body used for eating and talking						
3. piece of music						

ตัวอย่างจาก **Google Forms (Answer Sheet)**



The semi-structured interview included 12 validated items, adapted from previous research (e.g., Karataş et al., 2024). Each item was translated into Thai and validated by three experts who were experienced EFL lecturers/researchers at the tertiary level. The IOC values were 0.67 for 3 items and 1.00 for 9 items, indicating acceptable content validity, and minor revisions were made based on expert feedback. Interviews, which were piloted with five students, were conducted via Zoom and recorded for transcription and analysis.

Data Collection and Analysis

The UVLT was administered within classroom settings at the university. Quantitative data were analyzed using SPSS version 29 (IBM, 2022). Descriptive statistics were employed to examine vocabulary size, while the Mann-Whitney U test was utilized to compare different disciplines.

Following ethical approval, participants were recruited for interviews via purposive sampling, and interviews were scheduled at their convenience. All interviews were conducted online in Thai by the two researchers, each lasting approximately 45 minutes.

Prior to the interviews, an in-depth explanation of the interview protocols was provided to the participants. This included details on the purpose of the study, assurances of voluntary participation, and guarantees concerning anonymity and confidentiality. This process ensured that consent was well-informed; it also clarified that nothing in their responses could impinge on their academic standing. The participants primarily participated from home using either a computer or a smartphone. While they were allowed to keep their video off for personal comfort, all participants used video during self-introduction for identification purposes.

Reflexivity and Researcher Positionality

The research team included two senior EFL lecturers with over twenty years of teaching experience at the tertiary level. Their extensive background knowledge in vocabulary instruction, combined with a developing interest in AI, may influence their perspectives, and this should be noted.

A reflexivity journal was therefore maintained throughout the process of research, to try and "bracket" these assumptions (for example, "I suspect that students with low vocabulary levels may be more reliant on AI," or "Students might hide the whole truth to avoid being judged"). This journal can inform a critical assessment of how the researchers' positionality and

assumptions may have influenced data collection (for example, phrasing of probes) and its subsequent analysis, thus helping in the confirmability of the study.

Data Analysis and Trustworthiness

All interviews were audio-recorded and transcribed verbatim in Thai utilizing Zoom's transcription feature. To guarantee data integrity, both researchers meticulously reviewed the transcripts against the original recordings for accuracy. To ensure data accuracy, the researchers individually went through the transcripts and checked them against the recordings for any errors.

The analysis of these transcripts was conducted using qualitative content analysis (Dörnyei, 2007). A hybrid coding approach was employed in this research. Hybrid codes are those that result from both deductive and inductive methods of analysis. An initial scheme was formulated based on Nation's (2007) four strands framework, as applied by Meniado (2023), and was subsequently refined following a pilot interview. This framework was further developed by incorporating additional inductive codes that emerged directly from the data during analysis.

To enhance the credibility of the study, two techniques were essential. First, in-process member validation was performed during the interviews. Key points were restated by the interviewer (e.g., "If I am understanding you correctly, you feel that...") and clarifying questions were asked (e.g., "Could you please elaborate on that?") to ensure that the perspectives of the participants were accurately captured.

Secondly, to ensure trustworthiness, a codebook was produced made in an iterative process. The interview protocol and the initial coding scheme were piloted with one pilot-study interview. Both researchers independently coded this transcript, then met to discuss any differences between codes, hence refining the definitions of the codes.

This iterative process was then carried out using a second interview to develop inter-coder reliability (ICR) (Campbell et al., 2013), which resulted in 76.79% agreement, with the rest of the differences settled through discussions. The coding was applied individually to the entire data set of interview transcripts using finalized schemes.

Findings

RQ 1: What Is the Vocabulary Size of Thai Graduate Students in Science and Non-Science Disciplines?

Descriptive statistics were applied to determine vocabulary size among science and non-science graduate students. Table 4 presents the descriptive statistics of the scores. The number and percentage in the last column indicate students whose highest successfully passed level was the one shown. For instance, students at the 1,000 level passed only the 1,000-word test but not any higher levels, whereas those at the 5,000 level passed all lower levels and successfully reached the 5,000-word level. This classification therefore reflects each student's highest level of vocabulary mastery.

As noted by Webb et al. (2017), cut-off scores to indicate mastery should be 29/30 for the first three levels—that is, from 1,000 to 3,000—and at 24/30 for the last two levels—that is, from 4,000 to 5,000. Based on the criteria, Table 4 shows that, on average, neither group of students achieve mastery at any vocabulary level. The mean scores of science students and non-science students were 28.20 and 27.36, respectively, at the 1,000 level; 25.52 and 24.85 at the 2,000 level; 21.40 and 21.04 at the 3,000 level; 22.06 and 19.94 at the 4,000 level; and 19.61 and 17.29 at the 5,000 level.

Nevertheless, several students achieved passing scores. As illustrated in the final column, the passing rates varied from 0% to 41.28%. Eleven science students and sixteen non-science students, totaling twenty-seven students, successfully passed within the levels of 3,000 and 5,000; this constituted 12.44% of all students.

Additionally, 36 science students and 51 non-science students (40.09%) were found to have not mastered the 1,000-word level. From the data, it is clear that the number of students at the combined 3,000 and 4,000 levels was less than that at the 5,000 level, which had a total of 21 students. The exact number shows that only 6 students were at the combined 3,000 and 4,000 levels, 21 students passed at the 5,000 level, of whom 15 were non-science students.

Table 4

Descriptive Statistics of Science and Non-Science Students' UVLT scores

Test Level	Discipline	N	Mean	SD	Min.	Max.	Number and Percentage of students achieving mastery
1000	Science	109	28.20	3.04	13	30	45 (41.28%)
	Non-science	108	27.36	4.17	7	30	26 (24.07%)

2000	Science	109	25.52	4.19	11	30	17 (15.60%)
	Non-science	108	24.85	4.90	8	30	15 (13.89%)
3000	Science	109	21.40	6.02	5	30	3 (2.75%)
	Non-science	108	21.04	7.02	4	30	0 (0%)
4000	Science	109	22.06	5.05	9	30	2 (1.83%)
	Non-science	108	19.94	6.88	2	30	1 (0.93%)
5000	Science	109	19.61	5.44	4	30	6 (5.50%)
	Non-science	108	17.29	7.48	3	30	15 (13.89%)

RQ 2: Is There a Difference Between the Vocabulary Size of Thai Graduate Students in Science and Non-Science Disciplines?

A Comparison of the Vocabulary Size of Thai Graduate Students in Science and Non-Science Disciplines

Table 5 presents the results of the assumption testing, which informed the selection of an appropriate statistical test to compare the science and non-science groups. Normality, checked with the Kolmogorov-Smirnov Test, and homogeneity of variance, evaluated using Levene's Test, were examined across all five vocabulary levels.

The Kolmogorov-Smirnov Test showed that normality was violated by most of the data. At the 1,000-word level, data were not normally distributed for either the science ($D = 0.28$, $p < .001$) or non-science ($D = 0.28$, $p < .001$) groups. This finding was consistent at the 2,000-word level (science $p < .001$; non-science $p < .001$), at 3,000 words (science $p = .03$; non-science $p < .001$), and 4,000 words (science $p < .001$; non-science $p = .04$). The only exception was found at the 5,000-word level where data for both groups was found to be at a normal distribution (science $p = .11$; non-science $p = .20$).

Levene's test analyzed the variance homogeneity. The results of the test indicated that the assumption was true for the first three vocabulary levels, i.e., for 1,000 ($p = .09$), 2,000 ($p = .22$), and 3,000 ($p = .13$). However, this assumption was not met at the two lower frequency vocabulary levels. Levene's test yielded significant results for the 4,000-word level ($F(1,215) = 10.00$, $p < .001$) and for the 5,000-word level ($F(1,215) = 11.43$, $p < .001$). In other words, the variance between the two groups at these particular levels was not equal.

Since the normality assumption was largely violated at four out of five vocabulary levels, it does not seem appropriate to use a parametric test, such as an independent samples *t*-test, for comparison. Moreover, there was a violation of the homogeneity of variance at the 4,000 and 5,000-word levels.

Therefore, to ensure a robust comparison that is not dependent on these assumptions, the nonparametric equivalent, the Mann-Whitney U-Test, was used to compare vocabulary scores between the two groups.

Table 5

Kolmogorov-Smirnov Test and Levene's Test Results

Vocabulary Level	Group	Kolmogorov-Smirnov *D*	*p*	Levene's Test *F*	*df*	*p*
1000	Science	0.28	< .001	2.94	1, 215	.09
	Non-Science	0.28	< .001			
2000	Science	0.15	< .001	1.50	1, 215	.22
	Non-Science	0.15	< .001			
3000	Science	0.09	.03	2.32	1, 215	.13
	Non-Science	0.11	< .001			
4000	Science	0.13	< .001	10.00	1, 215	< .001
	Non-Science	0.09	.04			
5000	Science	0.08	.11	11.43	1, 215	< .001
	Non-Science	0.06	.20			

Note: Significance level for normality and homogeneity of variance was set at $p < .05$.

The results in Table 6 show that the differences between the two groups at the vocabulary levels of 1,000, 4,000, and 5,000 words were statistically significant. The science group performed significantly better than the non-science group at the 1,000-word level (U = 4957.5, $p = .036$), the 4,000-word level (U = 4917, $p = .036$), and also at the 5,000-word level (U = 4741.5, $p = .013$).

An examination of the mean ranks confirms the direction of this difference. In all three significant instances, the science students' mean rank (1,000 = 117.52; 4,000 = 117.89; 5,000 = 119.50) was considerably higher than that of non-science students (1,000 = 100.40; 4,000 = 100.03; 5,000 = 98.40).

There was no statistically significant difference between the two groups at the 2,000-word level (U = 5479, $p = .376$) or at the 3,000-word level (U = 5856.5, $p = .949$). In fact, mean ranks at these mid-frequency levels were virtually identical, further confirming that there was no meaningful difference between the two groups' vocabularies at these particular word levels. It might

be construed as indicating a "bookend" pattern of differential knowledge across the two cohorts.

The lack of variability at the 2,000- and 3,000-word levels is understandable, as this mid-frequency vocabulary constitutes the "common core" of both general, academic and popular texts. Universally shared by all graduate students, regardless of their particular discipline, it includes the fundamental vocabulary necessary for effective functioning within an academic setting.

The most important finding is the significant divergence at the 4,000 and 5,000 levels. This suggests that academic discipline could be an influential factor in vocabulary acquisition.

The difference at the 1,000-word level is also worth noting. This may suggest that students undertaking studies in the field of science had acquired a better basic vocabulary before embarking on, or during, their graduate studies, possibly as a result of slightly higher English proficiency requirements for their field or of more rigorous undergraduate education.

Table 6

Score Differences Between Science and Non-Science Students

Test Level	Mean Rank		Mann-Whitney U-Test	Asymp. Sig. (2-tailed)
	Science	Non-science		
1000	117.52	100.40	4957.5	0.036
2000	112.73	105.23	5479	0.376
3000	109.27	108.73	5856.5	0.949
4000	117.89	100.03	4917	0.036
5000	119.50	98.40	4741.5	0.013

RQ 3: What Are the Opinions of Thai Graduate Students in Science and Non-Science Disciplines about the Effects of their Vocabulary Size on their Use of AI Tools in their English Language Learning?

The interview analysis revealed that 10 students (47.62%) used AI to learn English: 4 low-science, 2 high-science, 1 low-non-science and 3 high-non-science students. The AI tools reported were ChatGPT, Gemini, Perplexity, Grammarly, Quillbot and Google Translate. Students' interviews showed that there were both similarities and differences between the low- and high-vocabulary groups regarding types of AI-assisted English language learning activities, the language used to prompt AI, their trust in AI, and their reliance on these tools.

Types of AI-assisted Language Learning Activities

Students with low and high vocabulary sizes shared some similarities and differences in the types of activities they engaged in with AI to learn English. The similarities included engaging in a variety of activities involving: (1) meaning-focused output (which may be followed by language-focused learning); (2) meaning-focused input for academic purposes; and (3) language-focused learning for academic purposes. The difference, however, was found in the low-level group using language-focused learning for general non-academic purposes.

Meaning-focused output activities

The first similarity was that students in both groups used AI to improve their English productive skills, which involved three strands: meaning-focused output and meaning-focused output together with language-focused learning. For instance, a low-vocabulary-level, non-science student (LNonSci 6) reported that she practiced speaking with ChatGPT, which involved meaning-focused output. She said that talking to AI was like having a conversation with friends. Another student, HSci 4, reported using ChatGPT to practice conversation skills and to prepare for the IELTS test:

> "I ask ChatGPT to have conversations with me about my daily life so I can practice English. I do this almost every day. Also, when I was considering taking the IELTS exam, I asked ChatGPT to act as an examiner and create sample speaking and writing questions for me to practice with. I also asked it to act as a rater and give feedback on the content of my answers and suggest ways to improve."
>
> HSci 4

This excerpt shows that HSci 4 engaged in speaking and writing activities, which are considered meaning-focused output. He also learned from feedback and suggestions, which is language-focused learning. Similarly, LSci 3 stated that she wrote diaries in English every day and had Grammarly check her grammar and explain the mistakes she made. This also involved meaning-focused output and language-focused learning.

Meaning-focused input activities for academic purposes

Another similarity was that students from each group used AI to improve their understanding of academic texts, which can be considered learning through reading, a meaning-focused input activity:

> "To improve my reading, after I read academic texts, I ask ChatGPT to summarize the texts. Then, I read the AI summary to check if my understanding is correct."
>
> HSci 8

> "I asked ChatGPT to translate some paragraphs in academic texts to Thai to check if I understand them correctly."
>
> LSci 1

Language-focused learning activities for academic purposes

In addition, students in both groups used AI for a variety of language-focused learning activities for academic purposes. Some used AI to learn how to improve vocabulary and grammar when writing in academic settings. HNonSci 4 learned how to use academic words from ChatGPT and Grammarly, while HNonSci 7 asked ChatGPT and HNonSci 9 asked ChatGPT and Gemini to suggest some words after they wrote their research assignments. HNonSci 7 stated that:

> "ChatGPT is really helpful for paraphrasing. I normally start with my own writing and then ask ChatGPT to paraphrase. The tool facilitates me by showing different synonyms, antonyms, and sentence patterns. Sometimes, the AI tool reminds me of some words that I have known but never had a chance to use. My writing skills may be improved because I have more exposure to variations of academic words."
>
> HNonSci 7

Similarly, some science students mentioned that they learned how to use particular grammar structures and new vocabulary from AI. For example, every week LSci 2 used Grammarly to check grammar and Quillbot to paraphrase some words after she wrote a summary assignment in her seminar class. She said, "I learned new vocabulary and what grammatical features I should use when writing."

LSci 1 learned to improve his grammar when writing abstracts by using ChatGPT:

> "I first wrote my abstract in Thai and translated it to English. Then, I used ChatGPT to translate the Thai version one or two sentences at a time. Afterward, I typed my English translation and asked the AI to compare it with ChatGPT's version to check for correctness. When the AI's word choice

> or grammar differed from mine, I asked for an explanation. When I understood, I adopted the AI's suggestions."
>
> LSci 1

Another language-focused learning activity that students in both groups reported was learning about the meaning of technical terms in their subject areas. For example, HSci 4 used ChatGPT and Perplexity to explain these terms while LSci 1 used ChatGPT for this purpose, commenting that "I use ChatGPT as a more advanced talking dictionary."

Language-focused learning activities for general purposes

In addition to similar learning activities in both groups as presented above, there were some activities that were reported only by students with a small vocabulary size. These activities were language-focused activities related to non-academic settings. LSci2 stated that she used Google Translate to learn how to pronounce words while LSci 6 used it to understand the meaning of general vocabulary. LSci 1 also learned how to use correct grammar, vocabulary and appropriate email writing style from AI. He wrote his email in Thai first and had Google Translate translate it to English. Then he asked ChatGPT to explain "whether my vocabulary and grammar were correct and if my sentences were polite and appropriate."

As can be seen, the high- and low-vocabulary size groups reported using AI for similar learning activities. That is, both groups engaged in activities involving meaning-focused output (which may be followed by language-focused learning), meaning-focused input for academic purposes, and language-focused learning for academic purposes. This may indicate that vocabulary size may not affect the types of learning activities students pursue with AI tools to improve their productive skills, understand academic texts, improve vocabulary and grammar for academic writing and learn about technical terms. This may suggest that AI can respond to the various needs of students with different vocabulary knowledge or language proficiency for practicing productive skills and academic skills in English. Another finding shows a difference between the two groups: only the low-vocabulary group applied AI for language-focused activities for non-academic purposes. This may indicate that vocabulary size may influence AI use for improving grammar, vocabulary, pronunciation, and writing for general (non-academic) purposes. AI tools may particularly benefit lower proficiency learners or those with a smaller vocabulary size by supporting non-academic language practice.

Languages Used to Prompt AI

The interviews revealed differences in the prompt languages used by the high- and low- vocabulary size groups when interacting with generative AI, suggesting that vocabulary size may influence the language students use to interact with AI. All students with larger vocabularies were found to use English when prompting AI. HSci 4 described his prompts to ChatGPT as follows:

> "You are an IELTS Writing rater. Evaluate my essay, provide feedback on the content, correct vocabulary and grammar, and suggest how I can improve."
>
> HSci 4

He also commented that: "I believe using English prompts would result in responses that match my objectives". He further explained that his large vocabulary size enabled him to do so.

However, some students with small vocabularies used both English and Thai, while others relied on Thai only. For example, LSci 2 stated that "I use English in my prompts but I switch to Thai when I don't know how to express myself." She explained that she once used a Thai prompt asking ChatGPT to make the vocabulary in her presentation script more formal because she did not know how to write the prompt in English. In contrast, she used an English prompt when asking ChatGPT about a news story on the US economy she had listened to on Spotify: "Explain what happened in the current US economy." Another student, however, used only Thai as he thought "my English is weak" (LSci 1).

Trust in Using AI

According to the interviews, mixed responses regarding trust were observed in both groups. That is, many students in both groups believed that AI gives accurate feedback when it comes to the English language, while others did not completely trust AI responses. For example, HSci 4 reported believing in AI feedback on his IELTS test performance: "I found ChatGPT's feedback to be highly reliable and valuable. I actually really trust what ChatGPT told me." Similarly, HNonSci 4 firmly believes that Grammarly provides correct grammar feedback and LSci 1 trusts AI corrections to his emails and accepts all suggestions.

Interviews also showed that students who trusted AI were likely to accept its feedback. However, one student with a small vocabulary size and another with a large vocabulary size, despite trusting the accuracy of AI feedback, stated they would not follow all suggestions:

> "I would only correct parts I understand. Also, if I believe my writing is correct, I would ignore the suggestion. And I wouldn't follow AI recommendations if they suggest words that seem too sophisticated for my level. I feel that it is not appropriate to use these words. They are too advanced for my ability."
>
> LSci 2

> "I typically write in English and use AI (ChatGPT) to help paraphrase it into English. I will review the wording and sentence structure and may adjust the sentences to fit my level."
>
> HNonSci 7

Other students did not completely trust AI responses, especially for AI summaries of research papers or AI's ability to understand or give feedback on general vocabulary and technical terms. HSci 8 stated that he would check the content accuracy with the original papers after ChatGPT summarized papers. Likewise, LSci 3 believed that AI like ChatGPT does not quite understand academic language. In terms of vocabulary, HNonSci 4 found that "AI sometimes makes errors by using idioms or providing information that doesn't exist." LSci 6 stated that she used Google Translate for quick word definitions, but she would check its correctness with Longdo, a non-AI bilingual online dictionary because "Google Translate may be wrong. I believe Longdo is accurate." The mixed responses regarding trust in AI feedback may indicate that vocabulary size may not significantly influence students' trust in using AI.

Reliance on AI Tools

With regards to reliance on AI, almost all students in both groups strongly believed that despite numerous benefits, AI is not an indispensable tool for students. This indicates that vocabulary size might not significantly influence students' reliance on AI. Students reported that while AI facilitates learning, other resources are available for English language development. For example, HSci 8 stated that although "ChatGPT can help us learn and save a lot of time when studying new topics," without it, he can still study. He compared AI to an iPad:

> "I think ChatGPT is essential but if we don't have it, that's fine. It is similar to an iPad. Before iPads, everyone used paper to do their assignments and submitted them to instructors. The iPad just made things more convenient. But if we don't have an iPad, can we still study? Yes, we can. Without ChatGPT, we can still study, using traditional methods."
>
> HSci 8

Other students seemed to have similar opinions:

"AI is not that necessary. It just helps make life easier."
LSci 2

"Without AI, there are other tools that I can use."
LSci 6

However, there was only one student who disagreed. HSci 4 believed that "AI is quite necessary for practicing English."

Discussion

This study analyzed the vocabulary size of Thai graduate students, compared vocabulary sizes between students from science and non-science disciplines, and explored students' opinions regarding the influence of their lexical knowledge on AI use for language acquisition. The key elements to be discussed next are concerns regarding the restricted lexical proficiency of Thai graduate students, AI applications, as well as the implications and recommendations.

Limited Lexical Knowledge

Regarding the vocabulary sizes of Thai graduate students in both scientific and non-scientific disciplines, the findings were alarming and raise major concerns. No group, on average, showed mastery at any of the five vocabulary levels. Only 27 out of 217 students (13%) reached high levels at 3,000, 4,000, or 5,000 words. Based on the mean scores, approximately 40% of the Thai graduate students (Master's and Ph.D.) in this study have not mastered even the 1,000-word vocabulary level. This suggests that many graduate students in this sample have lexical knowledge considerably lower than that reported for Thai undergraduates by Mungkonwong and Wudthayagorn (2017), who knew between 4,200 and 5,900 words. Vocabulary knowledge in many students in our study also falls well below the 3,000-word threshold required for basic comprehension (Nation & Waring, 1997) and the 5,000 words needed to understand unsimplified novels aimed at teenagers (Hirsh & Nation, 1992). This phenomenon may reflect a decline in language proficiency after completing the first degree. For graduate students who are working adults, if English does not require extensive use in their professional practice, exposure to usage remains limited. Bardovi-Harlig and Stringer (2010, p. 2) noted that "the length of time without input" can become one of the extralinguistic variables, along with age and motivation,

which may contribute to second language attrition or loss; thus, as students advance to higher degrees, their vocabulary diminishes further. Schmitt (2010) regarded vocabulary attrition as "a natural fact of learning," which should be viewed as the "loss of lexical access." In other words, continuous exposure to English language resources is essential for maintaining lexical knowledge.

Another reason for inadequate lexical knowledge may be low proficiency in English overall. In the English Proficiency Index (EPI) by Education First, Thailand was ranked 106th out of 116 countries in the year 2024—classified under the very low proficiency category (EF Education First, 2024). Prapphal (2003) assessed English proficiency among graduate degree applicants at a leading university in Thailand and found that most students did not meet the language standards required for graduate program admission, thus supporting the need for a revision of the English syllabus. Since this study was conducted in one of the leading and most renowned universities, there are likely to be even more significant concerns regarding inadequate vocabulary and language proficiency among other institutions across the country. It is an inconvenient truth which highlights the critical importance of focused vocabulary training to enhance the success of students with limited vocabulary in graduate education.

Another important concern may be the gap between the two academic disciplines. Science students significantly outperformed non-science students at the 1,000-word, 4,000-word, and 5,000-word levels. These results corroborate those of Srimonkontip and Wiriyakarun (2014), which also reported higher scores for science students than non-science students. This may imply that graduate students in science are better at recognizing and understanding basic vocabulary. Great exposure to English-language resources among science students might be one reason for this advantage.

Graduate-level science studies require huge input and ongoing engagement with materials written in English, even when conducted in Thai. Journal articles, technical manuals, and international textbooks all constitute discipline-specific reading that can accumulate large volumes of material, described as an 'input flood' (Krashen, 1985). Naturally, this facilitates exposure for science students to a broader range of vocabulary, which is a definite advantage. Conversely, non-science students, excluding language arts students whose vocabulary may be beyond the 5,000 level, might have a much lower need for, or access to, English-language resources. This circumstance might indicate that students with limited vocabulary across different academic disciplines have distinct needs and require additional support.

While Research Questions 1 and 2 dealt with the specific levels of vocabulary acquired by Thai graduate students, Research Question 3 probed students' perceptions of the effect of their vocabulary size on their use of AI for English language learning.

Mixed Results for the Influence of Vocabulary Size on AI Use for English Language Learning

Regarding the opinions of Thai graduate students in science and non-science disciplines about the effects of their vocabulary size on their use of AI tools for English language learning, several points merit discussion. First, the findings about the learning activities of the two vocabulary groups suggest that Nation's (2007) four strands framework is useful in categorizing AI-assisted language learning activities. With regard to the effect of vocabulary size, the findings suggest that vocabulary size may not influence types of learning activities for developing productive skills and academic skills, but it may affect learning activities for non-academic language-focused skills. This indicates that AI tools, which help reduce language barriers, can facilitate English learning for both high- and low-proficiency students, whose needs may differ from one another, highlighting the beneficial and inclusive role of AI in language learning. These findings are consistent with prior studies. For example, Lee et al. (2024), found that the use of artificial intelligence grammar and vocabulary applications enhanced general writing skills. Solak (2024) noted that language learners can effortlessly generate exercises in AI both for practice and individual needs. Similarly, Teng (2024) observed that feedback from ChatGPT is not only accurate but also immediate as well as pertinent to the matter at hand. Due to its competence, artificial intelligence fills the gaps left by traditional instruction by providing individualized assistance that specific learners are yearning for. This too is validated by the study conducted by Solak (2024), who found that AI makes possible a personalized learning experience by adjusting proficiency levels, interests, and preferred learning styles.

Additionally, the results seem to suggest that vocabulary size might influence the choice of language students use when communicating with AI, with lower-level students tending to prefer using Thai in their prompts. This finding reflects the rapid development of technologies since newer versions of AI can respond more efficiently in a greater variety of languages (Liu, 2024). Such developments can assist users speaking any native language or possessing any English proficiency.

Regarding trust in AI, mixed responses in both groups were found, suggesting no effect of vocabulary size. That is, some students—regardless of

their vocabulary size—expressed trust in AI's grammar corrections and feedback on their writing and speaking. Others, however, did not, citing AI's occasional inaccuracies in understanding and generating general and academic vocabulary. This suggests that while AI has advanced considerably and can serve as a language consultant for students, especially regarding grammar and productive skills, it still has limitations that need to be addressed. For example, AI tools can struggle with word translation and idiomatic expressions (Alhaisoni & Alhaysony, 2017; Lee et al., 2024). This highlights their limited understanding of human language, which can lead to hallucination issues and incorrect responses.

In addition, the findings regarding trust suggest that some students, irrespective of their vocabulary size, are aware of AI limitations in generating general and academic vocabulary, which indicates no significant influence of vocabulary size. This awareness may stem from factors such as a high level of language proficiency, frequent use of dictionaries to verify word meanings, or familiarity with technical terms in their fields. Another finding is that low- and high-vocabulary students who trust AI feedback tend to accept its suggestions, except for a student in the low- and another in the high-vocabulary group who ignored AI corrections if the original writing was already correct or if the suggestions were not appropriate for their proficiency levels. These findings seem to contradict Lee et al. (2024) and Koltovskaia (2020), which reported that only high-level students critically evaluated AI feedback and decided whether to follow it whereas low-level students were more dependent on AI and more likely to accept its suggestions.

Finally, although all students from both groups found AI helpful, most felt that its presence was not essential. Without AI, they believed they could still learn English through other methods as AI simply made the process more convenient, which indicates that vocabulary size might not have a significant influence on students' reliance on AI. The findings are likely due to the fact that the participants in our study are learners with pre-AI learning experience. Therefore, they are capable of utilizing other traditional or non-AI digital tools to support their language learning. The findings also contradict Karataş et al. (2024) and Lee et al. (2024), which found that students expressed concerns about overdependence on AI tools.

The findings about trust and reliance on AI may be further explained by Zhang & Hyland's (2018) model of student engagement with feedback. Our findings support and elaborate on the role of individual factors in student engagement. In our study, learner factors in addition to vocabulary size may have influenced learners' revision activities and attitudes towards AI feedback. Such factors include past English learning experiences, AI usage experiences and awareness of AI capabilities and limitations.

Implications and Recommendations

The results for students' vocabulary size show that on average, students' vocabulary knowledge is limited, with only 12.44% of students achieving mastery of 3,000 to 5,000 words. Science students performed significantly better than non-science students at the 1,000, 4,000, and 5,000 levels. Regarding use of AI, about half of the interviewees reported using AI for English language learning. Overall, vocabulary size does not appear to influence most types of AI activities used to support language learning or their degree of trust and reliance on these tools, except for the choice of prompt language and for engagement with language-focused activities for non-academic contexts. Based on the results, the implications and recommendations are as follows. First, there are several diagnostic and practical benefits of vocabulary tests. Cameron (2002) stated that vocabulary level tests help diagnose students' receptive vocabulary knowledge. Schmitt et al. (2011) highlighted that such tests can determine whether learners have mastered specific word bands, which may assist educators in selecting materials, tracking learners' vocabulary development, and identifying suitable vocabulary learning goals. Laufer (1998) suggests that vocabulary can be rapidly acquired passively. However, in more productive contexts, vocabulary instruction may require intentional pedagogical support or output-focused teaching that encourages students to take risks in production tasks. Instructors or program administrators may administer the UVLT before students begin their postgraduate study to evaluate their students' vocabulary levels and develop vocabulary training suitable for students in different fields in higher education. Second, as AI-assisted language learning tools have the potential to support language learners across vocabulary levels, both EFL students and teachers should be required to attend sessions on how to use these tools to enhance vocabulary development. They should learn about the capabilities of AI tools, including strategies to use them effectively and their current limitations. In addition, curriculum developers should include a variety of sample learning activities (Nation, 2007) targeted at each vocabulary level in language class syllabi, and these activities should be modeled in class so that students can apply them out of class to further improve their vocabulary. In addition, teachers should guide students on how to verify AI responses—for example, by recommending reliable resources such as dictionaries, educational websites, and textbooks. Students should also be encouraged to discuss the accuracy of AI content with peers and consult teachers to develop their critical evaluation skills. Finally, policymakers should support the implementation of AI systems and provide non-AI resources that

can be used to verify AI feedback in universities to ensure that they can be accessible and beneficial to all.

Limitations

This study has some limitations. First, the study was conducted with graduate students at one university. Studies with participants from different institutions would be more generalizable to a larger population of graduate students. Data about AI usage was collected through interviews only; no other methods, such as observation of AI usage or analysis of prompts, were used to triangulate and validate the results. Future studies should employ multiple instruments to collect data for better validation of results. Additionally, the UVLT measures vocabulary level knowledge but does not measure depth of vocabulary knowledge. Tests regarding the "depth of knowledge" that might be included in future studies are those that elicit how well students can use their vocabulary (Schmitt et al., 2011, p. 31). Finally, data was collected at one point in time. A longitudinal study would provide better insight into the development of students' engagement with AI over time.

Conclusion

This research examined the vocabulary size and opinions of the effects of vocabulary size on use of AI tools in English language learning among 217 Thai graduate students from a public university, comprising both science and non-science disciplines. Using the UVLT and interviews, the study found that, on average, students did not achieve mastery across the five vocabulary levels. Science students demonstrated significantly better performance than non-science students at the 1,000-, 4,000-, and 5,000-word levels. Interview data provided insights into the effect of vocabulary size on types of AI activities for language learning, their prompt language, trust and dependency on AI tools. These findings underscore a potential need for AI-integrated vocabulary instruction for EFL learners in higher education to enhance their vocabulary development.

Acknowledgements

The authors are grateful for the research grant from the Research Department, Chulalongkorn University Language Institute, Chulalongkorn University. Also, we would like to express our gratitude to the Office of the Research Ethics Review Committee for Research Involving Human Subjects,

Chulalongkorn University, for approving this study (Project Number 670230).

About the Authors

Sumanee Pinweha: A lecturer at Chulalongkorn University Language Institute, Bangkok, Thailand. Her research interests include second-language vocabulary acquisition and AI-assisted language learning.
Sutthirak Sapsirin: A lecturer at Chulalongkorn University Language Institute, Bangkok, Thailand. Her research interests include language assessment, vocabulary, and technology in language learning and testing.

References

Alhaisoni, E., & Alhaysony, M. (2017). An investigation of Saudi EFL university students' attitudes towards the use of Google Translate. *International Journal of English Language Education*, *5*(1), 72–82. http://dx.doi.org/10.5296/ijele.v5i1.10696

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition. *Studies in Second Language Acquisition, 32*(1), 1–45. https://doi.org/10.1017/S0272263109990246

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, *6*(2), 145–173. https://doi.org/10.1191/1362168802lr103oa

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semi structured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research, 42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Chang, T. S., Li, Y., Huang, H. W., & Whitfield, B. (2021). Exploring EFL students' writing performance and their acceptance of AI-based automated writing feedback. In *Proceedings of the 2021 2nd International Conference on Education Development and Studies* (pp. 31–35). https://doi.org/10.1145/3459043.345906

Charnchairerk, C. (2022). Key defining linguistic features in the writing performance of first-year university students across different language proficiency levels. *LEARN Journal: Language Education and Acquisition Research Network*, *15*(2), 858–891. https://so04.tci-thaijo.org/index.php/LEARN/article/view/259954

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

EF Education First. (2024). *EF English proficiency index 2024: A ranking of 116 countries and regions by English skills*. Signum International AG. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2024/ef-epi-2024-english.pdf

Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, *35*(9), 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696. https://doi.org/10.64152/10125/67046

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0) [Computer software]. IBM Corp.

Karataş, F., Abedi, F. Y., Ozek G. F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29:19343–19366. https://doi.org/10.1007/s10639-024-12574-6

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, Article 100450. https://doi.org/10.1016/j.asw.2020.100450

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics*, *7*(3), 437–451. https://dx.doi.org/10.21462/jeltl.v7i3.852

Laufer, B. (1998). The development of passive and active vocabulary: Same or different? *Applied Linguistics*, 19, 255–271. https://doi.org/10.1093/applin/19.2.255

Lee, Y.-J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies, 14*(1), Article e202412. https://doi.org/10.30935/ojcmt/14195

Liu, Y. (2024). Reshaping and transforming of English teaching in higher education in the ChatGPT era: An empirical study based on big data. In *2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C)* (pp. 1302–1311). IEEE. https://doi.org/10.1109/qrs-c63300.2024.00169

Meniado, J. C. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal*, *14*(4), 3–18. https://dx.doi.org/10.24093/awej/vol14no4.1

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Mungkonwong, P., & Wudthayagorn, J. (2017). An investigation of vocabulary size of Thai freshmen and its relationship to years of English study. *LEARN Journal: Language Education and Acquisition Research Network*, *10*(2), 1–9. https://so04.tci-thaijo.org/index.php/LEARN/article/view/111681

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines 5*, 12–25. https://so04.tci-thaijo.org/index.php/LEARN/article/view/282280/189364

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 1–12. https://doi.org/10.2167/illt039.0

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

Nation, I. S. P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. https://jalt-publications.org/i/2007-07_31.7

Nation, I. S. P. & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy, (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.

Prapphal, K. (2003). English proficiency of Thai learners and directions of English teaching and learning in Thailand. *Journal of Studies in the English Language, 1*, 6–12. https://so04.tci-thaijo.org/index.php/jsel/article/view/21840

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*, 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Smith Jr., E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models and applications* (pp. 93–122). JAM Press.

Solak, E. (2024). Revolutionizing language learning: How ChatGPT and AI are changing the way we learn languages. *International Journal of*

Technology in Education, 7(2), 353–372. https://doi.org/10.46328/ijte.732

Srimonkontip, S., & Wiriyakarun, P. (2014), Measuring vocabulary size and vocabulary depth of secondary education students in a Thai-English bilingual school. *Journal of Liberal Arts, Ubon Ratchathani University, 10*(2), 181–209. https://so03.tci-thaijo.org/index.php/jla_ubu/article/view/94542/73929

Sukying, A. (2023). The role of vocabulary size and depth in predicting postgraduate students' second language writing performance. *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 575–603. https://so04.tci-thaijo.org/index.php/LEARN/article/view/263457

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270.

Webb, S. (2013). Depth of vocabulary knowledge. In C. Chappelle (Ed.), *Encyclopedia of Applied Linguistics.* (pp.1656–1663). Wiley-Blackwell.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, *168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90-102. https://doi.org/10.1016/j.asw.2018.02.004