



A True Longitudinal Corpus Study of the Lexical Competence of L2 English Major Students: Determining the Effectiveness of the Writing Track

Jiraporn Dhanarattigannon^a, Tirote Thongnuan^b, Pong-ampai Kongcharoen^{c*}

^a jiraporndh2000@yahoo.com, Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand

^b tirote.t@ku.th, Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand

^c pongampai.k@ku.th, Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand

*Corresponding author, pongampai.k@ku.th

APA Citation:

Dhanarattigannon, J., Thongnuan, T., & Kongcharoen, P. (2025). A true longitudinal corpus study of the lexical competence of L2 English major students: Determining the effectiveness of the writing track *LEARN Journal: Language Education and Acquisition Research Network*, 18(1), 352-375.
<https://doi.org/10.70730/NGQY4466>

Received 16/05/2024	ABSTRACT
Received in revised form 11/09/2024	This true longitudinal study explored the lexical development of L2 university English major students in Thailand using two learner corpora. This study analyzed the natural learners' English language writing assignments from two writing courses and tracked their lexical development over a relatively long period (three years with five batches of students). The data included four academic writing assignments for the first learner corpus and a research project for the second corpus. The two corpora comprised 619 pieces of writing with 462,842 tokens in total. The results from the VocabProfile program show that the students used academic words at a high level in their research project, revealing that they tended to develop their lexical competence after being exposed to more advanced
Accepted 20/09/2024	

	<p>English reading sources. The findings also suggest the effectiveness of the course management of the B.A. Program in English, which helps students improve their lexical competence to achieve the requirements for high-quality journals.</p> <p>Keywords: Learner corpora, Longitudinal study, Lexical competence, L2 writing course effectiveness</p>
--	--

Introduction

Recently, the results of the Program for International Student Assessment (PISA) in 2022 raised concerns for the administration of Thai education to promote Thai students' proficiency in the three main aspects assessed by PISA: Reading, Arithmetic, and Science. They reveal that the average scores of Thai students in these three areas are lower than those in 2018. Additionally, the results of PISA from 2000 to 2022 show that Thai students' proficiency in arithmetic and reading (in English) was likely to decrease. PISA is used to evaluate the quality of the educational system of the countries in providing their youth with the fundamental skills necessary to live in a changing world. Based on this, the current results of Thai students lead to questions concerning the quality of the Thai educational system. Therefore, the Commission on Higher Education Standards (CHES, 2024) announced a policy on January 23, 2024 to raise the standard of English language instruction at the undergraduate level, which universities would use as a guideline for improving their English language instruction standards to ensure students' mastery of the English language.

As a crucial part of the educational system, English teachers should be aware of this policy and reflect on the current situation of the curricula where English courses are offered to undergraduate students, particularly English majors, because they are expected to use English more effectively than non-English majors. One of the key elements in helping ESL/EFL learners develop English proficiency is lexical competence or vocabulary. Many studies (Abduh & Rosmaladewi, 2017; Cobb, 1997; Nation, 2006;) on lexical-frequency lists (corpus) suggest that allowing learners to gain exposure to necessary frequency-based word lists can help them expand their English vocabulary. Further, the mastery of these word frequencies will also enable them to reach a higher level of English proficiency. Abduh and Rosmaladewi (2017) provide various reasons why vocabulary is important for learning and developing learners' English proficiency, one of which is that useful academic words are essential for non-native English learners to "read and publish articles in English" (p. 283).

Previous research has delved into the vocabulary size that L2 learners, especially those at the tertiary level, should master. For example, Hu and Nation (2000) suggest that a language learner must know about 98-99% of the lexical-frequency list in written and spoken discourse to understand English texts. Nurweni and Read (1999) studied the vocabulary sizes of freshmen university students in Indonesia and found that they mastered about 1,226 words, which was far below 4,000 – 5,000 words, the threshold for high school completion. Alfalte (2016) conducted a study called *Investigating the Growth of Vocabulary Size and Depth of Word Knowledge in Iraqi Foreign Language Learners of English*, which found that students gained about 800 – 1,000 words every year, and their vocabulary grew moderately over the course of the study. In Thailand, the core curriculum prescribes that students graduating from high school or becoming first-year university students should know at least 3,600-3,750 words (Ministry of Education, 2008 as cited in Mungkonwong & Wudthayagorn, 2017). Mungkonwong and Wudthayagorn (2017) explored the vocabulary size of Thai freshmen concerning years of English study. They concluded that Thai freshmen possessed about 4,200 word families, and their vocabulary size somehow correlated to the number of years they studied English. These studies reveal the significance of the frequency-based word lists that can assist L2 learners in achieving English proficiency. In other words, they confirm that vocabulary size is beneficial for L2 learners in that it enhances their communication skills and reading proficiency. It also helps develop other language skills, namely listening, speaking, and writing. That is why a particular number of frequency-based lists are required at different levels of study according to the core curriculum.

Until now, expectations regarding the effectiveness of the track of English Writing courses in the program, particularly for English-major students, have remained empirically unexplored. The effectiveness of writing courses is typically measured by the test scores and students' evaluation of course content, teaching materials, and teacher performance. This measurement is "summative and subjective in nature", as argued by Crosthwaite (2016, p. 2). To gain reliable results to determine the effectiveness of writing courses in the program, it is necessary to investigate the students' writing development by collecting data systematically and in a natural setting, then analyzing the data based on the academic framework, including the lexical-frequency lists (corpus). In his doctoral study, Coates (2020) suggested the use of a corpus of academic writing to reflect the learners' academic vocabulary and language material and instructions. Therefore, this study used the true longitudinal corpus study to explore the lexical competence of L2 English major students at the tertiary level to reveal

their writing development and determine the effectiveness of the writing track by the Bachelor of Arts Program in English.

Based on the purposes of the study, the two research questions were addressed as follows:

1. How did the students' writing performance develop in terms of lexical competence?
2. How effective was the writing track for the current Bachelor of Arts Program in English in terms of lexical competence?

Literature Review

To facilitate adequate vocabulary learning, four vocabulary learning partners (students, teachers, materials writers, and researchers) need to contribute to the learning process. Vocabulary learning programs need to include both an explicit, intentional learning component and a component based around maximizing exposure and incidental learning. The overriding principle for maximizing vocabulary learning is to increase the amount of engagement learners have with lexical items. All four learning partners need to acknowledge the incremental nature of vocabulary learning, and to develop learning programs which are principled, long-term, and which recognize the richness and scope of the lexical knowledge that needs to be mastered. (Schmitt, 2008, p. 329)

According to Schmitt (2008), students and teachers play important roles in vocabulary learning. To help students expand their vocabulary size, teachers need to provide them with comprehensible input (frequency-based word lists) and assignments in which they can use those words. To develop language learning programs, teachers need to know the students' language repertoire and their learning process via their writing performance. To achieve this, they need to explore the students' vocabulary sizes in their written texts. This longitudinal study will help to learn how they progress after taking courses offered by the program. To achieve this, the study analyzed the vocabulary that they used in their writing assignments in a natural setting to see how the writing courses could assist them in expanding their vocabulary sizes, thus providing a picture of the effectiveness of the program, particularly the writing track.

Learner Corpora and Language Learning

Granger (2008) defined learner corpora as “electronic collections of texts produced by language learners” (p. 259). Learner corpora are beneficial in terms of second language acquisition. For example, they are used for

understanding interlanguage, and for developing pedagogical tools and methods to serve the learners' needs (Granger, 2008).

Learner corpora contribute to language learning and teaching in many aspects, such as classroom methodology, materials design, language testing, and syllabus design (Granger, 2008). The corpora can be used to explore how students use L2 so that the teacher can design the teaching materials and tests to suit their level and help them progress. Besides, the information gained from analyzing learner corpora can be used to design syllabuses and organize language courses appropriately to help students learn and improve their language skills step by step. Although learner corpora are very useful for SLA research as they lead researchers to gain a better understanding of how L2 learners learn the language, learner corpus research has not been popular among SLA researchers, which may be due to "the extreme scarcity of longitudinal learner corpora" (Granger, 2008, p. 266). Another reason why SLA researchers may not be much interested in learner corpora or the use of the CIA approach seems to be related to "the drawbacks of the CIA approach," which assumes that "learners have native speaker norms as a target" (Hunston, 2002, pp. 211-212 as cited in Granger, 2008, p. 270). However, Hunston pointed out the two benefits of the learner corpus approach: "First, the standard is identified and, if felt to be inappropriate, can be changed and replaced by another standard; and second, the standard is realistic: it is *what native/expert speakers do rather than what reference books say they do*" (Hunston, 2002, pp. 211-212 as cited in Granger, 2008, p. 270).

Reppen (2023) conducted a true longitudinal study and she argues:

[r]egarding the first issue, data for longitudinal studies are usually collected in language classrooms or exams, with the writing tasks/topics being tightly controlled. A major disadvantage is that tasks can be unrepresentative of the writing required in disciplinary content courses. The present study analyzes writing development as it occurs 'naturally' in university disciplinary content courses. (Abstract)

Therefore, it collected the writing assignments in a natural setting that was not controlled to reveal the genuine development of English writing by EFL university students.

Vocabulary Size and English Proficiency Development

Schmitt (2008) reviewed the article on L2 vocabulary learning. He divided the research into four main areas, comprising "the scope of the vocabulary learning challenge", "issues in vocabulary acquisition and pedagogy", "intentional learning of vocabulary", and "incidental learning of vocabulary". For this study, the first area was addressed. Schmitt revealed

that, according to Hu and Nation (2000), the study in the first area shows that the percentage of sufficient lexical items in the discourse that a language learner needs is closer to 98-99% for written vocabulary. Many studies (Milton & Hopkins, 2006 as cited in Schmitt, 2008; Nation, 2006) showed that the range of written vocabulary learners should have is around 4500 – 5,000 words families to read authentic texts. However, the study on learner's vocabulary size found that learning such a large amount of vocabulary is difficult for language learners (Schmitt, 2004 as cited in Schmitt, 2008). Further, the vocabulary sizes that the learners possessed in these studies reported by Laufer (2000) as cited in Schmitt (2008), were about 2000 word families smaller than the expected ones. Schmitt (2008) suggested, “[m]ost importantly, students need the willingness to be active learners over a long period of time, for without this, they are unlikely to achieve any substantial vocabulary size, regardless of the quality of instruction” (p. 333).

Alfatle (2016) studied “the growth of vocabulary size and depth of word knowledge in Iraqi foreign language learners of English” and investigated the factors affecting their learning and English vocabulary growth. He found significant vocabulary growth, with students gaining around 800 – 1,000 words each year and their growth “accelerated moderately over the course of study” (Abstract). This suggests that years of study highly affect students' growth of vocabulary size and their depth of word knowledge. He also found that using descriptive statistics, the learners' daily language practice, and their mother's education are variables affecting their vocabulary size and depth of word knowledge. Gender is not relevant to the students' lexical outcomes.

Meanwhile, Abduh and Rosmaladewi (2017) examined students' vocabulary levels using the online Lextutor to analyze 13 academic essays written by students of English at an Indonesian university. Their findings indicated that the students in the study used a different range of vocabulary in their academic writing, including basic and academic words, and terminologies. They concluded and suggested:

- Participants in this research will find it difficult to read and understand academic texts, because they had mostly mastered a low level of vocabulary.
- It is inferred that students were at the intermediate vocabulary level of competence, because they were able to use basic words, terminologies and academic words in their academic essays.
- Students who wish to increase their vocabulary levels can start to engage with exercises and word classification in on-line sources: for example: Lextutor. (Abduh & Rosmaladewi, 2017, p. 285)

However, both Alfalte (2016) and Abduh and Rosmaladewi (2017) did not use the AWL and COCA/BNC to investigate the students' academic word families. Alfalte used the *Self-Rated Vocabulary Test* (SRVT) to measure vocabulary size and the *Word Associates Test* (WAT) to measure the depth of word knowledge. This may imply that the growth of vocabulary size might include general and/or academic words. Although Abduh and Rosmaladewi studied both general and academic words, they did not study the learners' vocabulary size gained when they learned English at university. Meanwhile, these two studies tend to suggest that the growth of vocabulary size of L2 learners takes time, and it is the teachers who should help promote their engagement and improvement in lexical knowledge. Therefore, a longitudinal study of learner corpus in academic writing should be conducted to gain tangible insight into this issue.

Word Lists to Identify Students' Vocabulary Size

Academic Word List (AWL)

Academic word list (AWL) was first developed by Coxhead (2000), consisting of 570 word families most frequently found in four academic disciplines. The list was extracted from a 3.5 million word corpus including Arts, Science, Law, and Commerce. Each field comprises seven subject areas of about 875,000 running words each. The process of extracting the AWL starts from screening out the 2,000 most frequent general words or General Service List (West, 1953) to the criteria of frequency and dispersion. The AWL consists of 570 word families after the final process. The AWL has been employed extensively among English teachers and students to facilitate English vocabulary learning. In this research, the AWL was used as one criterion to identify students' lexical ability.

British National Corpus and Corpus of Contemporary American English List (BNC/COCA List)

BNC/COCA list was created by Nation (2016). It consists of lists categorized by frequency level within each of the 1000 word families. There are 25 levels in BNC/COCA list starting from the most frequently used words in the first 1000 to the 25th. The high-frequency lists, like the 1st 1000 and 2nd 1000, include words commonly used in situations like foreign travel, studying in English, and online activities. The BNC/COCA lists were created using two different methods. The first two lists were based on a corpus of 10 million words, with six million coming from spoken British and American English. The remaining four million words were derived from written British

and American English. By using this approach, the initial two lists remained unaffected by the written content incorporated into the later lists. From the 3rd 1000 list onwards, rankings in the BNC (British National Corpus) and COCA (Corpus of Contemporary American English) were used, excluding words from the first two thousand lists. These vocabulary lists were created to help learners of English, particularly those studying it as a foreign language. Henceforth, this study used the BNC/COCA list to signify students' vocabulary level.

Studies on Vocabulary at the College Level in Thailand

Mungkonwong and Wudthayagorn (2017) studied the vocabulary sizes of 484 Thai freshmen from four public universities and three private universities in Thailand. They found that the Thai freshmen had about 4,200 word families. Besides, they explored the correlation between the students' vocabulary size and their years of study of English using the Pearson correlation coefficient. The results revealed a negligible relationship between years of study and students' vocabulary sizes.

To help Thai students improve their English, Wiriyakarun (2018) studied Thai EFL learners' knowledge of academic English vocabulary because limited vocabulary is one of the main problems for Thai students in learning English. Further, knowing the existing level of their vocabulary will assist teachers in finding effective ways to enhance their English ability. Wiriyakarun constructed a new Academic Vocabulary test based on Coxhead's AWL. The quantitative results showed that students' knowledge of receptive and productive academic English vocabulary is not quite different. In terms of the relationship between these two academic English vocabularies, the results revealed a moderately positive relationship. Moreover, she found a moderately significant relationship between the students' academic vocabulary knowledge and their achievement. She agreed with Mungkonwong and Wudthayagorn (2017) in that students' vocabulary sizes will expand when they gain more experience learning English.

Methodology

Setting

To design a learner corpus, some variables must be considered. They are categorized into two main variables, including learner variables, meaning the learners' characteristics i.e., age, gender, and language proficiency, and task variables, which are related to the language situation such as task types (Atkins et al., 1992; Ellis, 1994; Granger, 2008). To compile a learner corpus,

these variables need to be controlled. For this study, the learner and task variables were controlled in the two writing courses, as described below.

English Writing Course

English Writing is the first writing course for English major students. Its primary goal is to enhance their writing skills and provide them with a solid foundation in academic writing through a process-based approach. The syllabus focuses on the writing process and incorporates activities such as drafting, peer response, teacher-student conferences, and self-assessment. In this course, students are given four writing assignments. They engage in various activities to improve their writing performance and learn how to write, revise, and evaluate their work. This approach aims to help students become independent writers. For each assignment, students begin by writing a first draft, which serves as a form of free writing to allow them to freely express their ideas on the chosen topic. Then, they share their drafts with peers and engage in peer response sessions focusing on content. Based on the feedback, students revise their drafts and use the revised version for the teacher-student conference. During the conference, students discuss their writing with the teacher, who provides guidance to clarify and strengthen their work. Based on the Zone of Proximal Development (ZPD) (Vygotsky, 1978) framework, students' abilities are greatly developed. After the conference, students make further revisions and evaluate their writing process and product using a self-assessment form. These activities foster autonomous learning and help students develop skills to seek out information for their writing. Although some variables such as learner autonomy and lexical exposure outside of the classroom are difficult to control, the course is designed to enhance learning through these activities. Upon completion of the course, students are expected to become more autonomous learners and gain greater exposure to English texts while researching for their writing assignments.

Introduction to Research Writing in English Course

This advanced writing course is the final writing course for students who have completed the previous two writing courses. The primary objective of this course is for students to practice writing a research article. With their prior experience in academic writing, students are assigned to read research articles and learn how to write their own. As students are novices in conducting research, the course is structured to include group work, with students collaborating in pairs or small groups of 2-3 to conduct a research study and write an article on their chosen research topic. This approach

provides students with increased exposure to higher-level academic English texts and helps them become familiar with specialized academic terms. Throughout the course, students apply their writing skills to produce a research article while simultaneously learning how to conduct a research study. To further motivate students to write quality research articles, they are encouraged to submit their articles for presentation and publication at national and international conferences.

The data collected in this study were the writing assignments from five batches of English major students at a public university in Thailand. Two data sets were collected from the two writing courses of two different periods over a relatively long period (three years).

The first set of data comprised the authentic writing assignments written by students. The students were sophomores who enrolled in their first English Writing course. The writing assignments comprised Listing, Sequence, Comparison-contrast, Cause-effect, and Problem-solution organizational patterns.

The second set of data consisted of research papers written by English-major students at the same public university in Thailand. These students were seniors enrolled in the English Research Writing course, which is the last writing course.

This study is considered a true longitudinal research design as evidenced by the following details. First, the two data sets were collected from the same batches of students over a relatively long period (three years). Second, the first set of data in each batch was collected from approximately 50-70% of the students, while the second set of data was collected from all students. The students in the second set of data were assigned to conduct their research papers in groups. Therefore, the assignments of all students needed to be collected to gain a sufficient number of texts for statistical analysis as well as to balance the corpus sizes. Besides, these students were homogenous in terms of study experience, i.e. they studied the same core courses and were given the same assignments for all three years.

Data collection

The data for this study consisted of two sets of learner corpora for English writing assignments. The first learner corpus consisted of four types of writing assignments for the English Writing course each academic year. They were collected from five batches of five academic years between 2017 and 2021. The data for each academic year are shown below:

Table 1*First Learner Corpus*

Year	No. of students	Type of assignments	No. of assignments	No. of words
2017	38	4 : Listing, Sequence, Comparison and contrast, Cause-effect	152	50,496
2018	25	4 : Listing, Sequence, Comparison and contrast, Cause-effect	100	21,470
2019	25	4 : Listing, Sequence, Comparison and contrast, Cause-effect	100	30,827
2020	22	4 : Listing, Sequence, Comparison and contrast, Cause-effect	88	27,938
2021	35	4 : Listing, Comparison and contrast, Cause-effect, Problem-solution	140	59,760
Total	145		552	190,506

Table 1 presents the first learner corpus, which contains a total of 552 assignments of 190,506 words in total. Because of the revision of the curriculum in the academic year 2020, the writing task for the Sequence pattern was removed and substituted by Problem-solution.

Table 2 below presents the second learner corpus, which comprises the final draft of students' research papers. They were collected from five batches of five academic years between 2019 and 2023. The data for each academic year are shown below:

Table 2*Second Learner Corpus*

Year	No. of research papers	No. words
2019	10	42,019
2020	14	61,465
2021	11	47,816
2022	16	61,144
2023	16	59,892
Total	67	272,336

Altogether, there were 67 research papers of 272,336 words in total in the second learner corpus. The learner corpora for this study were authentic because they were course assignments, and the students chose their topics to research and write about.

Table 3

The Two Learner Corpora

First Corpus			Second Corpus		
Year	Assignments	Total words	Year	Papers	Total words
2017	152	50,496	2019	10	42,019
2018	72	21,470	2020	14	61,465
2019	100	30,827	2021	11	47,816
2020	88	27,938	2022	16	61,144
2021	140	59,760	2023	16	59,892
Total	552	190,506	Total	67	272,336

Table 3 shows clear data collected for the two learner corpora. The first learner corpus was slightly smaller than the second one, which may have been because the writing assignments in the first writing course were written based on general topics, and they were only paragraph-size (250 words) and multiple-paragraph (400 words) essays. Despite being individual tasks, the time allowed for compiling each piece of writing was rather short. For the second corpus, the size was larger since they were full research papers of about 3000 – 4500 words in length. The research papers in the second learner corpus were assigned in groups, with two to three students in each group. Therefore, the size of the second corpus was larger when the students' papers were combined.

Reference Corpus

Thailand's three highly recognized Scopus-indexed journals were included in the reference corpus as they are favored by both students and academics who wish to have their papers published. However, the data collected were only from one issue of each journal published in 2023 as the

Vocabfile program had insufficient space. The purpose of comparing the results from the learner corpus with the reference corpus was to predict the degree of development needed for the students to achieve lexical competence.

Table 4

Reference Corpus

Journals	All tokens
Journal 1	39254
Journal 2	52765
Journal 3	29702
Total	121,721

Table 4 presents the total running words of each journal for the reference corpus. This corpus is compiled from three well-known Scopus-indexed journals in Thailand.

Data analysis

To determine the Academic Word List (AWL) and BNC/COCA vocabulary list levels, the VocabProfile program, created by Paul Nation and available for free at <https://www.lextutor.ca/vp/eng/>, was used to analyze the data. The program can screen the AWL with 570 word families and 25 vocabulary base lists, each of which contains the most 1000 words from BNC/COCA.

Results

Due to the space limit in the VocabProfile program, the data had to be categorized into small pieces with a maximum of only 35,000 words each. Therefore, the results from the first learner corpus needed to be classified into paragraph organizational patterns, and the second learner corpus was classified by year of study.

Table 5

Results from Reference Corpus: AWL Results

Journals	Families	Types	Tokens	Percent of Tokens	All tokens
Journal 1	391	832	4329	11.03%	39254

Journal 2	424	970	5814	11.02%	52765
Journal 3	369	737	2763	9.30%	29702

The reference corpus in this study (Table 5) shows that the highest percentage of the AWL used is in journal 1 at 11.03%, followed by journal 2 at almost the same amount of the AWL used at 11.02%, and journal 3 at 9.30%. The AWL in the reference corpus appears to comply with what has been suggested by experts of about 10% coverage in all academic texts (Coxhead, 2000).

Table 6

Results from the Reference Corpus: BNC/COCA Results

Reference corpus	95%	98%
Journal 1	4000	9000
Journal 2	3000	5000
Journal 3	4000	7000

The reference corpus in this study (Table 6) revealed that the first 4,000 words for Journal 1 and Journal 3 were used to gain 95% coverage, whereas the first 3,000 words for Journal 2 were used to gain 95% coverage. To gain 98% coverage, 5,000 words were required for Journal 2, while Journal 3 needed 7,000 words. However, as many as 9,000 words were needed for Journal 1.

Table 7

Results from the First Learner Corpus: AWL Results

2017		Families	Types	Tokens	Percent of Tokens
Listing	AWL Words:	125	158	329	3.03%
Sequence	AWL Words:	116	142	261	2.30%
Compare-contrast	AWL Words:	134	180	455	3.08%
Cause-effect	AWL Words:	150	212	505	3.74%

2018		Families	Types	Tokens	Percent of Tokens
Listing	AWL Words:	55	67	123	3.39%
Sequence	AWL Words:	62	68	92	2.48%
Compare-contrast	AWL Words:	107	139	366	5.12%

Cause-effect	AWL Words:	97	121	222	3.71%
2019	Families	Types	Tokens	Percent of Tokens	
Listing	AWL Words:	96	124	199	3.38%
Sequence	AWL Words:	90	108	173	2.82%
Compare-contrast	AWL Words:	162	208	444	4.77%
Cause-effect	AWL Words:	156	223	532	5.60%
2020	Families	Types	Tokens	Percent of Tokens	
Listing	AWL Words:	89	105	162	3.31%
Sequence	AWL Words:	75	88	125	2.61%
Compare-contrast	AWL Words:	157	218	412	4.46%
Cause-effect	AWL Words:	181	241	468	5.20%
2021	Families	Types	Tokens	Percent of Tokens	
Listing	AWL Words:	201	277	572	3.88%
Compare-contrast	AWL Words:	196	291	649	4.29%
Cause-effect	AWL Words:	212	325	733	4.93%
Problem-solution	AWL Words:	228	327	803	5.35%

The overall results from the AWL in the first learner corpus (Table 7) indicated that students in this study used approximately 2%-5% of the AWL, which was considered relatively low. Kongcharoen et al. (2024) clarified that the students seemed to use the lowest number of the AWL in a sequence organizational pattern almost every year, approximately 2% lower than the number suggested by experts (Coxhead, 2000). However, the highest number of the AWL was used in a cause-effect organizational pattern as they had to demonstrate their thinking at a more advanced level, resulting in a more elaborate choice of topics

Table 8

Results from the First Learner Corpus: BNC/COCA Results

Academic Year	Listing		Sequence		Compare-contrast		Cause-effect		Problem-solution	
	95%	98%	95%	98%	95%	98%	95%	98%	95%	98%
2017	3000	5000	3000	5000	3000	4000	3000	4000	N/A	N/A

2018	3000	5000	3000	5000	3000	4000	3000	4000	N/A	N/A
2019	3000	4000	3000	4000	3000	5000	3000	5000	N/A	N/A
2020	3000	4000	2000	3000	3000	5000	3000	6000	N/A	N/A
2021	3000	5000	N/A	N/A	3000	4000	3000	5000	3000	5000

The results of the BNC/COCA list from the first learner corpus (Table 8) showed that students in this study used BNC/COCA at various levels depending on the paragraph organizational patterns. The maximum BNC/COCA level used by the students was 6000 in the year 2020.

According to the BNC/COCA results, the students were likely to possess general academic words at the first 3,000-word level (95% coverage). The fact that this was their first English writing course and that they were sophomores limited their exposure to higher levels of English (Kongcharoen et al., 2024).

Here are examples of the AWL used in students' writing.

1. , or you was scolded. So, here is the option that you should do when you feel down.
2. The last step is to promote you blog by sharing on Facebook or Twitter.
3. do it and I hope you and he recover from this heart wound soon.
4. are messed and improper; the interviewer might reject you because of your looks.
5. must recognize that what you should do for sustain your relationship and what you shouldn't

Here are examples of the BNC/COCA's first 2000 words used in students' writing.

1. convenience in searching for any product. You are able to buy or view all products anywhere
2. Following these tips, making notes will never be monotonous and your boring notes will become
3. setting priority of thing you will do for tomorrow during the day to prevent stress interrupting you
4. less frequent than before as I became more aware of places I went and became
5. we move closer to these apes, they can contract our diseases such as COVID-19 and Ebola.
6. old hobby. It will help you make a decision easier by knowing all of these benefits.

Table 9*Results from the Second Learner Corpus: AWL Results*

Year	Families	Types	Tokens	Percent of Tokens	All tokens
2019	355	655	4128	9.82%	42019
2020	372	757	6282	10.71%	61465
2021	347	711	5638	11.79%	47816
2022	416	897	7422	12.14%	61144
2023	447	1018	8055	13.45%	59892

In the second learner corpus (Table 9), the students appeared to use the AWL at a higher level. The highest percentage of the AWL used by students was 13.45% in the year 2023, while the lowest was 9.82% in the year 2019. However, the students in this second learner corpus seemed to use the AWL more than those in the reference corpus when comparing the reference corpus, which was conducted from well-known Scopus-indexed journals in Thailand. This suggested that students used academic words that met the suggested level from the reference corpus.

When compared with the first learner corpus, the students seemed to use the AWL at a much higher level in the second corpus, with a significant increase from about 5% to 13%. This suggested that the B.A. English curriculum in this study navigated in the right direction in terms of developing students' lexical competence.

Table 10*Results from the Second Learner Corpus: BNC/COCA Results*

Year	95%	98%
2019	4000	7000
2020	4000	6000
2021	3000	5000
2022	4000	7000
2023	4000	6000

When considering the BNC/COCA vocabulary used by students in their second corpus, 95% coverage aligns with the reference corpus, which suggests a coverage of 4000 words. However, when considering 98% coverage, the vocabulary used by the students in this study was in the middle of the number suggested by the reference corpus with 5000 to 7000 words, while the reference corpus suggested about 5000 to 9000 words. This suggested that the students in this study developed their vocabulary

competence in order to meet the minimum vocabulary coverage to be able to publish their works in well-known international journals.

Discussions

Development of Students' Writing Performance

Based on the findings in Tables 4 and 6, the results of the AWL from the two data sets reveal the development of students' writing performance.

In terms of lexical competence, the students' writing performance develops with the AWL, from approximately 2-5% in the first corpus to 9-13% in the second corpus. Based on BNC/COCA lists, students' writing performance develops from using the first 2000-3000 words to gain 95% coverage and 4000 – 6000 words to gain 98% coverage to using the first 3000 – 4000 words and 5000 – 7000 words to gain 95% and 98% coverage, respectively. Unlike Knoch et al. (2015), who studied *ESL students' writing after three years of study at an English medium university* and found that their writing improved in terms of fluency but not lexical competence due to the lack of extensive writing, the lack of collective feedback on writing accuracy, and the nature of the tasks (generic writing tasks), which to them, enabled the students to develop their writing during the term of study, this study revealed that the two writing courses exposed students to extensive writing with numerous specific writing tasks, and provided room for collective feedback. The results of this study agreed with those by Alfatile (2016), Abduh and Rosmaladewi (2017), and Mungkongwong and Wudthayagorn (2017) in that more years of study resulted in larger vocabulary sizes among students. The results of the study suggest that the students' exposure to authentic reading texts at a more advanced level enabled them to achieve a higher level of lexical competence.

When compared with the reference corpus, the results from the second corpus show that the students in this study used AWL to reach the benchmark and, in some years, the level was even higher. This confirmed that students in this study developed their lexical competence after they completed the program's writing track.

Effectiveness of the Writing Track for the Current Bachelor of Arts Program in English

The results from Tables 7-10 revealed the students' writing development in terms of lexical competence. This reflected the effectiveness of the writing track for the current Bachelor of Arts Program in English. Based on the AWL results of the second learner corpus from Table 9, the

results suggest that the program can enhance students' writing competence, particularly lexical competence (AWL 9 – 13%). This is related to three main factors as follows:

1. The program's writing track,
2. Methods of teaching writing, materials, tasks, and assignments, and
3. Encouraging students to get their research papers published.

Program writing track

The program has a systematic writing track for English-major students. To illustrate, the first writing course emphasizes the writing process to help students generate ideas fluently as well as practice their academic writing. The second writing course encourages them to practice academic writing of advanced genres, including problem-solution and argumentative writing. The third writing course emphasizes the writing skills needed for research writing, including summarizing, analyzing, paraphrasing, writing annotated bibliographies, and reviewing literature. The last course, Research Writing, combines the writing skills learned from the previous courses and focuses on reading and writing research articles. This reflects the step-by-step planning that takes into consideration levels of difficulty and important skills for writing development. Meanwhile, the students are provided with sufficient and appropriate reading texts and writing assignments for each course while taking these courses.

This is strongly supported by the results of interviews with English-major students by members of the Curriculum Revision Committee, who surveyed their satisfaction with the courses offered by the program and used the results as solid evidence to develop a revised version of the English Curriculum (B.E. 2568). Interviewed students mentioned the effective course arrangements based on the level of difficulty in the writing track as one important factor for developing their lexical competence. While the first writing course provides them with a firm background in academic writing, the second writing course offers a practical perspective of writing where they can develop their logical thinking and select appropriate academic vocabulary to reflect the content they create for each organizational pattern. With essential research writing skills acquired from the third course, they prepare themselves academically for the last course, where they can fully apply what they have learned from the previous three courses in the production of high-quality research papers.

Methods of teaching writing, materials, and tasks and assignments

For teaching methods, the process-based approach is used for the first writing course to help students learn their writing process and how to generate their ideas freely. Also, the materials are created by the instructors to provide them with authentic writing essays and how to write the organizational patterns required. Throughout the course, they are engaged in various activities necessary for improving their writing process, such as multi-drafting, peer-response, teacher-student conferences, and self-assessment. These activities enable them to learn and reflect on their writing as well as their peer's writing; they can then use the good points to improve their writing. Through teacher-student conferences, the students learn about their weaknesses and strengths, and they learn to think logically when conversing with the teacher. In the end, they can use this conference as a fruitful model for giving feedback on their writing as well as that of their peers. The last activity is self-assessment, which helps the students evaluate their written essays using the rubric provided by the teachers. These activities enhance their authorship and enable them to become better writers.

Writing assignments in all writing courses were task-based and focused on writing performance. In each course, the students were assigned writing tasks corresponding to the course objectives. For the first writing course, they were assigned to write four essays with different types of organizational patterns. In the Research Writing course, they were asked to conduct research in a step-by-step manner, including writing each part of a research article.

Encouraging students to get their research papers published

Since 2017, the program has been encouraging the students enrolled in the Research Writing course to produce research papers suitable for publication. In 2017, only one paper was published in the form of a proceeding at a national conference. Later, more students' papers were accepted for publication at international conferences. Encouraging the students to have their papers published seems to motivate them to produce high-quality research papers and push them to reach their potential to have their research papers published. As a result, two to six papers were successfully published at international conferences from 2018 to 2022. In 2022, one of the students' papers received the first prize award for the best paper.

Based on interviews with English-major students, they ranked the Research Writing course as one of the most beneficial courses in the program. Apart from promoting their logical thinking, which is essential for writing

research papers, the course motivated them to maximize their academic caliber through paper publication.

Conclusion and Implications

To conclude, the results of this study suggest that 1) years of study are related to the L2 development of English-major students' lexical competence, 2) the true longitudinal study of students in the same batch can reveal their vocabulary growth, and 3) an appropriate study plan for courses in the writing track, along with sufficient and effective teaching and learning activities and tasks, can enhance students' academic growth, particularly in terms of lexical competence. To make the program successful, courses in the writing track must be well planned. The study plan for the track should be well-organized in terms of the content and teaching approaches, as well as the activities and tasks. To illustrate, the content for each course in the track should be organized from basic to advanced levels. The teaching approach should emphasize students' learning, engagement, and hands-on practice, and assignments should be task-based.

For further studies and implications, the results of this study suggest the following:

1. To gain more insights into students' academic growth, the longitudinal development of grammatical complexity in ESL students' writing should be explored further. The results of future studies will provide a more complete picture of students' growth in terms of writing competence.

2. This longitudinal study of two sets of data collected from the same batches can be used as a model to explore students' development in other study tracks, such as the reading track. The strength of this study is that the two data sets were collected from students who had the same experiences through the teaching and learning of major courses, engaged in the same activities, and completed the same assignments in a relatively long period (three years).

3. Further studies on the development of other skills or cognitive domains in other tracks such as reading, translation, and cultural studies should be explored by the methods used in this study to gain more information on L2 students' academic growth and to reflect the effectiveness of the program.

Acknowledgements

This research was funded by the Faculty of Humanities, Kasetsart University.

About the Authors

Jiraporn Dhanarattigannon, Ph.D.: An assistant professor in the Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand. Her areas of interest include EFL writing (process-based approach), Extensive Reading, self-assessment, online language learning and teaching, and corpus-based research. She can be reached at jiraporndh2000@yahoo.com.

Tirote Thongnuan: A lecturer of English in the Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand. His areas of interest are corpus linguistics and corpus-based research, discourse analysis, English syntax and grammar, and translation studies. He can be reached at tirote.t@ku.th.

Pong-ampai Kongcharoen: An assistant professor of English in the Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Bangkok, Thailand. She also holds a position as EFL-IS (English as a Foreign Language Interest Section) Past Chair at the TESOL International Association. Her research interest lies in corpus linguistics, semantics, discourse analysis, vocabulary learning and teaching, and second language acquisition. She can be reached at pongampai.k@ku.th.

References

Abduh, A., & Rosmaladewi, R. (2017). Taking the Lextutor on-line tool to examine students' vocabulary level in business English students. *World Transactions on Engineering and Technology Education*, 15(3), 283 – 286. <https://www.researchgate.net/publication/320840538>

Alfatle, A. B. M. (2016). Investigating the growth of vocabulary size and depth of word knowledge in Iraqi foreign language learners of English. [Master's Thesis, Missouri State University]. <https://bearworks.missouristate.edu/theses/2230>

Atkins B. T. S., Clear J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing* 7(1), 1-16.

Commission on Higher Education Standards (CHES). (2024). *Notification of the Commission on Higher Education Standards RE: The policy to raise English language standards in higher education institutions B.E. 2567 (2024, 23 January)*. <https://www.ops.go.th/th/all-announcement/circular/item/9660-0204-2-1646-2567>

Coates, A. D. (2020). A corpus analysis of academic writing and how it informs writing instruction on a university pre-sessional course. [Doctoral dissertation, Manchester Metropolitan University]. <https://e-space.mmu.ac.uk/627416/1/Andrew%20Coates%20-%20Final%20Thesis%202.pdf>

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.

Coxhead, A. (2000) A new academic word list. *TESOL Quarterly*, 34(2): 213-238. <https://doi.org/10.2307/3587951>

Crosthwaite, P. (2016). L2 English article use by L1 speakers of article-less languages. *International Journal of Learner Corpus Research*, 2(1), 68-100. <https://doi.org/10.1075/IJLCR.2.1.03CRO>

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press, Oxford, UK.

Granger S. (2008). Learner corpora. In Lüdeling, A. & Kytö, M. (Eds.), *Corpus linguistics. An international handbook. Volume 1* (pp. 259-275). Walter de Gruyter.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 23(1), 403-430.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, 28, 39-52.

Kongcharoen, P., Dhanarattigannon, J., & Thongnuan, T. (2024). Lexical level in English major EFL students' writing: A learner corpus study. *PASAA* (68). 203-228.

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

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-81. DOI:10.1353/cml.2006.0049

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

Nurwени, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, 18(2), 161-175.

Reppen, R. (2023). Exploring individual longitudinal development in a corpus of 'natural' disciplinary writing: what could it mean for teaching? (Abstract). *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2023-0173>

Schmitt, N. (2008). Review article instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
<https://doi.org/10.1177/1362168808089921>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

West, M. (1953). *A General Service List of English Words*. Longman, Green & Co.

Wiriyakarun, P. (2018). Examining Thai EFL learners' knowledge of academic English vocabulary. *The Liberal Arts Journal, Mahidol University*, 1(1), 119-132.