INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Vol 14, Issue 10, (2024) E-ISSN: 2222-6990

Academic Word List and Cefr Levels: Profiling Academic Vocabulary in a Technical University Learner Corpus

Noorli Khamis
Pusat Pembelajaran Bahasa (PPB), Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Coressponding Author Email: noorli@utem.edu.my

Nurhidayah Abdul Bareed
Pusat Pembelajaran Bahasa (PPB), Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

Ina Suryani
Centre of Excellence for Autonomous Unmanned Aerial System, Business and Communication Faculty Universiti Malaysia Perlis, Malaysia

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v14-i10/23102 DOI:10.6007/IJARBSS/v14-i10/23102

Published Date: 05 October 2024

Abstract

This study investigates the vocabulary profile in argumentative essays written by students from different disciplines at a technical university in Malaysia. Using Coxhead's (2000) Academic Word List (AWL), the study profiles the students' academic vocabulary across disciplines. Given the national requirement to align the English course with the Common European Framework of Reference (CEFR), this study also analyses the CEFR levels of vocabulary in the essays to determine whether the students are meeting the expected language proficiency benchmark. The study aims to bridge the gap between learners' current vocabulary use and the vocabulary required to achieve academic writing proficiency, especially in relation to both general academic and discipline-specific language demands. From the vocabulary profiles, this study provides insights that can inform more effective instructional approaches for teaching and learning of academic writing and help instructors tailor their strategies to support students' progress toward academic proficiency.

Keywords: Learner Corpus, Academic Writing, Argumentative Essays, AWL, CEFR.

Introduction

Academic writing courses at universities are primarily designed to provide learners with the means to express complex ideas, engage in critical discourse, and make meaningful contributions to their academic fields. These are done using the academic language, or

register, which encompasses specific grammatical structures and vocabulary. It is believed that learners with varying levels of language proficiency can benefit from a clear and structured introduction to academic writing (Ma, 2023; Thaksanan & Chaturongakul, 2023). For second language learners, mastering academic writing means to be able to acquire and effectively use these academic structures and vocabulary in their work (Lahuerta, 2023; Muhammad and Muhammad, 2023). This is especially crucial for learners of English for Specific Purposes (ESP), who need to be able to communicate effectively within their discourse community.

There have been many investigations conducted on the use of academic language. Academic vocabulary, one of the central dimensions in academic texts, receives great attention from many researchers (Charpentier Jiménez, 2023; Chung, Wan, & Fung, 2024; Nuoya, 2024; Therova, 2023). This is because academic vocabulary is not random; the words do not only indicate one's language mastery but also writing quality (Maamuujav, 2021). The words learners choose and how they employ them serve as indicators of their linguistic competence - the higher the vocabulary range, the more competent the learners are in expressing their ideas in the language (Shuxratovna, 2024). Moreover, a limited grasp of academic vocabulary is frequently linked to lower levels of academic performance (Gardner & Davies, 2014). Therefore, considering learners of different specialisations, it is essential to understand the vocabulary needs in academic writing of these learners as an attempt to assist them to be proficient academic writers, particularly in their disciplines.

Background Of The Study

CEFR-Informed Academic Writing Syllabus

In Universiti Teknikal Malaysia Melaka (UTeM), a technical university in Malaysia, Academic Writing is offered as one of the compulsory subjects in all the undergraduate programs. As the result of the national English language education reform 2015-2025 introduced by the Ministry of Education of Malaysia in 2015, all English subjects in all universities are required to be aligned with the Common European Framework of Reference (CEFR) (Ministry of Education, 2015). CEFR is a common framework for English language teaching, learning, and assessment. In the CEFR, English language proficiency is described according to six levels, which ascend from A1 to C2, with A1 representing Basic Language User and C2 Proficient Language User (Council of Europe, 2024). Table 1 shows the descriptions of the CEFR levels.

Table 1
The CEFR Global Scale (Council of Europe, 2024)

Basic User	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has.
	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters.

		Can describe in simple terms aspects of their background, immediate environment and matters in areas of immediate need.
Independent User	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
Proficient User	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

As such, the subject, Academic Writing, has been designed to be CEFR-informed, and the highest expected performance of the students is pitched at the B2 level, i.e., Independent Language User. There are three types of writing introduced in this subject, namely Product Description, Argumentative Essay, and Text Review. All writing skills and language components, mainly words and structures, are devised by adapting the CEFR.

For this study, the focus is given to the argumentative essay writing. This is due to the fact that argumentative thinking and writing skills are important in academic settings, and university learners need to acquire the ability to discuss, evaluate, and negotiate for meaningful communication persuasively in most academic contexts (Zhang, Lopez Wui, Nam, Eunjung Relyea, & Wong, 2023). In addition, argumentative discussions involve several critical skills, including taking a stand, providing supports, justifying claims, and making counterarguments and refutations (Crowell & Kuhn, 2014). As such, this text type offers dynamic language use that serves varied intentions.

Therefore, this study attempts to investigate the vocabulary profile in the

argumentative essays of learners from different university disciplines. The study mainly examines the learners' academic vocabulary in the essays by employing the Coxhead's (2000) Academic Word List (AWL) for the profiling. In addition, since the English course is aligned with the CEFR as a language proficiency benchmark, this study also seeks to identify the CEFR word levels used by learners in their writing. By doing so, it aims to address the gap between learners' current vocabulary and the vocabulary needed to achieve academic writing proficiency. The findings from the vocabulary profiling are intended to inform instructional strategies for more effective teaching and learning in the academic writing course. Hence, the research questions for this study are:

a) What is the distribution of academic vocabulary in the learner corpus according to the AWL and CEFR levels?
b) Are there differences in the distribution of AWL and CEFR-level words between the different disciplines in the learner corpus?
c) How does this information relate to the quality of learners' academic texts?

Literature Review

Types of Vocabulary

Generally, there are four types of vocabulary, which include high-frequency words, academic words, technical words, and low frequency words. *High-frequency words* comprise words that cover almost 80% of all kinds of texts. These words are regarded as the most frequent English words, which provide 'general service' to English learners. *Academic words* include words used in formal academic contexts across various disciplines, but they are uncommon in non-academic texts. They characterise the academic activities, including research procedures, analysis, and evaluation. According to Nation (2001), academic words cover approximately 9% of any academic text. However, the recent computerised approach to quantifying the vocabulary of texts from different disciplines and genres has seen the coverage of academic words in academic texts at about 10-14% (Therova, 2023). *Technical words* are words related to a specific topic in the text but not so common in other texts. The words provide an immediate idea about the topic covered in the text. These words make up 5% of the words in a text. The fourth vocabulary type is known as *low-frequency words*, which cover about 5% of an academic text. They encompass other words than high-frequency, academic, and technical in a particular subject. Figure 1 summarises these vocabulary types.

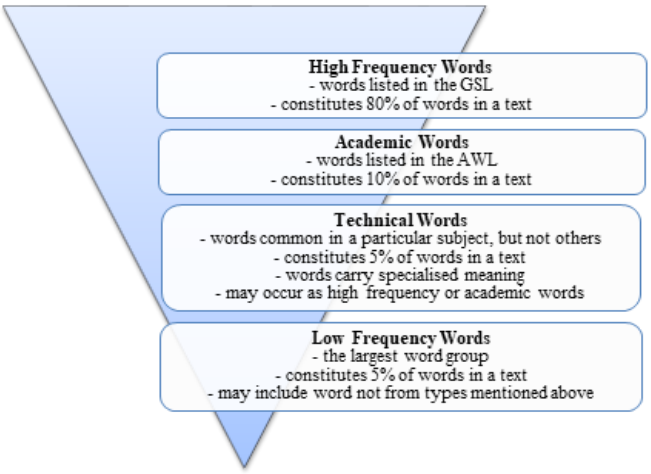


Figure 1. Types of vocabulary (Nation, 2001)

It is generally accepted that a threshold of 95% to 98% of word coverage in a text should be known to the learners to have adequate understanding of the text (Haruna, Tajuddin, & Bashir, 2024; Hu & Nation, 2000; Run & Weina, 2024). Therefore, a combination of high-frequency and academic words is crucial to ensure this threshold is achieved in academic settings. Studies on academic vocabulary suggest that the development of academic vocabulary can improve word knowledge, lexical access, and morphological awareness (Asaad, 2024; McKeown, Crosson, Moore, & Beck, 2018; Zaid, 2024).

In addition, a growing number of studies have proven that there are variations in academic vocabulary across disciplines (Özer and Akbaş, 2024). Because academic vocabulary is what students from all academic fields encounter while reading, they should be able to use the words in their written work. Hence, this underlines the fact that the academic vocabulary needs of learners should be given due attention.

Accordingly, several structured word lists have been developed to assist learners in applying the language for their academic purposes (Drayton & Coxhead, 2023; Lei & Liu, 2016; Sarimah, Noraini, & Ameruddin, 2013; Uba, Irudayasamy, & Hankins, 2023). The use of word lists for vocabulary teaching and learning is becoming popular. According to Folse (2023), word lists are convenient and preferred by many learners. Researchers have also used some word lists to measure vocabulary coverage in texts, including learners' work. This study employs two such word lists to determine the distributions of academic words in the learners' work, thus, understanding their academic vocabulary needs. The first word list, the Academic Word List (AWL) (Coxhead, 2000), is used to measure the academic words in the essays. The second word list, the English Vocabulary Profile (EVP) (Amkham, 2016), functions like a tool to profile the words in the essays according to the CEFR levels.

Academic Word List (AWL)

The Academic Word List (AWL) comprises 570 words that occur frequently in a wide range of academic texts. Developed by Coxhead (2000), this list is intended for students entering a college or university. The AWL was systematically derived from a very large corpus of academic texts, including journals and university textbooks in the areas of arts, commerce, law, and science. The AWL coverage in the four areas was measured. It was observed that the lowest coverage is for science at 9.1%, and the highest coverage is for commerce at 12%. Generally, the AWL is divided into ten sublists in decreasing order of frequency, and each sublist contains 60-word families. Many new word lists adopt the approach used in developing this list (Folse, 2023).

It is noteworthy that the AWL is not subject-specific, which means that the words are not connected to any particular subject. Thus, the list is useful for learners from all disciplines. The AWL has been used in many studies to profile academic vocabulary in texts (Guihua & Luo, 2023; Muhammad & Muhammad, 2023).

English Vocabulary Profile (EVP)

The English Vocabulary Profile (EVP) is more than just a word list. It is a resource that offers information about phrases, idioms, and collocations that are categorised according to each CEFR level. Unlike most word lists, the EVP operates at the level of individual meanings. This means that it provides reliable information about words, such as meanings, word classes, and

examples, as known and used by learners (Capel, 2015). The resource is an output of research using the Cambridge Learner Corpus (CLC), and it is also informed by the Cambridge English Corpus (CEC). The EVP is employed in mainly vocabulary research in EAP and ESP (Matsuzaki & Mark, 2020; Sun, 2017).

Text Inspector (TI)

Both AWL and EVP analyses in this study are retrieved using the online tool, the Text Inspector (TI). This tool has the capability to process a text of a maximum 10,000 words at one time. The available output of this tool includes the statistical information of the texts, the CEFR-labelling of each word, readability index, lexical diversity, and many others (Larrson, 2017; Weblingua Ltd., n.d.).

Methodology

The dataset for this study was collected from a larger corpus, the Academic Writing Learner Corpus (AWLC). The corpus is a compilation of written texts produced by undergraduate students from eight different disciplines who undertook Academic Writing subject in their second year of study at the university. The AWLC consists of 1331 written assignments from three text types, which are product description (PD), argumentative essay (AE), and text review (TR). Table 2 details the composition of the corpus.

It is worth noting that the corpus represents a large sample of undergraduates from different proficiency levels and writing on a wide range of topics according to their disciplines. However, for this study, the analyses were conducted based on the data chosen from the sub-corpus argumentative essay. Ten samples from each discipline were chosen randomly and grouped according to their discipline, with each sample having a different topic and a word count of around 300 words. Overall, 80 samples were collected with a total token of 22,440. Table 3 provides the composition of the sub-corpus extracted for this study.

Table 2
The composition of AWLC

No.	Faculty	Programme	No. of files			TOTAL
			PD	AE	TR	
1	Electrical Engineering	BEKM	34	34	35	103
2	Mechanical Engineering	BMCG	22	21	23	66
		BMCK	31	34	34	99
3	Electrical & Electronic Engineering Technology	BEET	19	24	24	67
		BEEE	29	35	35	99
4	Mechanical & Manufacturing Engineering Technology	BMMF	23	23	22	68
		BMMS	28	26	26	80
5	Information & Communication Technology	BITZ	80	50	50	180
		BITM	66	64	61	191
6	Manufacturing Engineering	BMFG	72	73	74	219
7	Technology Management & Technopreneurship	BTMI	24	24	24	72
8	Electrical & Electronic Engineering	BENG	29	29	29	87
	TOTAL		**457**	**437**	**437**	**1331**

Table 3
The composition of sub-corpus for this study

Faculty	Prog	No of Texts	Tokens
Electrical Engineering	BEKM	10	2916
Mechanical & Manufacturing Engineering Technology	BMMF	10	2668
Mechanical Engineering	BMCG	10	2455
Technology Management & Technopreneurship	BTMI	10	2665
Electrical & Electronic Engineering Technology	BEET	10	2773
Information & Communication Technology	BITZ	10	3028
Manufacturing Engineering	BMCG	10	3032
Electrical & Electronic Engineering	BENG	10	2903
Total		**80**	**22440**

The data was cleaned prior to analysis with the Text Inspector. Any forms of image and diagram were removed before the data was converted from the original submission, which was in a PDF format, to a text format. The text files were inspected to make sure there were no missing details from the original source. The dataset was then uploaded into the automated online tool, textinspector.com, in batches according to discipline. Once the data was processed, the scorecard and the results for EVP and AWL were exported into Excel spreadsheets. The CEFR levels, EVP type count, EVP wordlist, AWL token count, and AWL wordlist were extracted for further analyses.

Results and Discussion

Table 4 shows the distributions of AWL words across all programs. Generally, the total token count across the programs is almost similar, with BITZ having the highest number of tokens (2972), and BMMF the lowest (2663). The number of words used in all the programs is fairly even. The students were required to write a maximum of 300 words for the task. Interestingly, the results reveal significant variation across disciplines, with a notable gap between the highest and lowest percentages of AWL coverage. It ranges from 6.34% (BMMF) to 14.39% (BENG). A low AWL coverage in the essays by BMMF students indicates the essays contain a relatively few academic words, while a higher coverage in the BENG essays indicates the texts might be written in more complex and formal style, employing a greater use of academic words. This finding aligns with the earlier discussion by Özer and Akbaş (2024). They also highlight that the AWL fails to cover specific vocabulary needed for particular disciplines. While the generic AWL provides a broad academic vocabulary, it may not fully support the specialised vocabulary required in specific fields. Therefore, these results suggest that the essays by the BMMF students may contain more field-specific words compared to essays by the BENG students. The variation in AWL coverage across these disciplines reveals important differences in lexical demands and writing styles.

Table 4
Coverage of AWL

	BMCG	BENG	BMMF	BTMI	BEET	BEKM	BITZ	BMFG
Token count	2791	2843	2663	2665	2709	2862	2972	2956
AWL All Tokens count	358	401	163	228	293	274	356	309
AWL 1 token count	57	95	50	46	64	59	85	83
AWL 2 token count	52	72	25	58	37	39	99	65
AWL 3 token count	64	55	16	30	50	49	52	31
AWL 4 token count	50	53	5	29	36	19	27	26
AWL 5 token count	18	24	12	23	38	23	37	37
AWL 6 token count	28	32	17	11	17	8	9	25
AWL 7 token count	31	35	4	19	24	29	16	16
AWL 8 token count	41	15	26	5	13	30	13	15
AWL 9 token count	14	11	6	6	12	17	16	9
AWL 10 token count	3	9	2	1	2	1	2	2
Unlisted token count	2295	2347	2339	2295	2318	2371	2495	2507
AWL All Tokens %	**13.21**	**14.39**	**6.34**	**8.83**	**11.06**	**10.02**	**12.29**	**10.75**

Despite the variations, a closer look at the detailed breakdown of AWL sublists reveals that the higher AWL sub lists (AWL 1 to AWL 3) were used by the students across the disciplines. This indicates that there was still a strong reliance on core academic words in their essays. The low usage of the lower AWL sublists (AWL 9 and AWL 10) by students of most disciplines indicates less frequent, more specialised academic words were less employed too. Another worth-noticed observation from Table 4 is the number of unlisted tokens (words not listed in AWL). The number of unlisted tokens is the highest in BITZ (2495) and BMFG (2507) essays. This high number of unlisted AWL words could indicate that students from these disciplines used more technical or field-specific words that are not covered in general academic vocabulary. Therefore, this finding supports the idea that the AWL is limited to certain disciplines, as highlighted earlier. These disciplines rely more on field-specific, non-AWL words. These findings resonate with other studies that attempted to construct field-specific AWL, such as computer science academic vocabulary list (Roesler, 2021), veterinary medicine academic word list (Özer & Akbaş, 2024), Physics Research Articles Word List (PRAWL) (Vuković-Stamatović, 2024), and Philosophical Academic Word List (Phawl) (Khani, & Javadian, 2024).

Thus far, the low and high AWL coverage across all programs in sub-corpus suggests that the students of some disciplines at the university used more field-specific and less academic words, and some used formal and highly academic words. This calls for a closer look at the set of words in the essays of these different disciplines. The analysis of the words is discussed in relation to the CEFR levels. Table 5 displays the distribution of CEFR levels by programs.

Table 5
Distribution of CEFR levels by programs

	B2+	C1	C1+	C2	C2+	D1	D1+
BEKM			2	5	1	2	
BMMF	1		5	2	2		
BMCG				8	2		
BTMI		1	3	6			
BEET			2	3	5		
BITZ				5	4	1	
BMFG			1	6	3		
BENG			2	3	2	1	2
TOTAL	1	1	15	38	19	4	2
%	**1.25**	**1.25**	**18.75**	**47.5**	**23.75**	**5**	**2.5**

Generally, across all programs, the lowest CEFR level is B2+ and the highest is D1+. The results indicate the students' writing abilities are at varyingly higher CEFR proficiency levels, with a majority falling within the C1+ and C2+ ranges (90%), with a peak at C2 (47.5%). This suggests that all the students were able to display advanced CEFR levels, exceeding the level set for the course, i.e., B2. Only one student (BMMF) achieved B2+, while six students (BITZ and BENG) managed to achieve the D1 and D1+ levels, which represent very high levels of language proficiency, typically displayed by advanced learners who have a native-like command of the language.

This performance may be attributed to the vocabulary used by these students, who generally wrote on topics related to their disciplines, which warrants further analysis of the CEFR vocabulary profiles. Table 6 details the distribution of CEFR word lists in the sub-corpus by programs. The CEFR word list was extracted from the EVP, which is also provided in the TI software.

Table 6 demonstrates the interaction between AWL and CEFR word lists in the overall language performance of the students across the programs. BENG students, who had the highest AWL coverage (as shown in Table 4), also possessed the highest number of higher CEFR word levels, totaling 111 words (C1 and C2 type counts). As indicated in Table 5, BENG students achieved higher CEFR levels, with three students reaching D1 and D1+. Conversely, BMMF students, who had the lowest AWL coverage in the sub-corpus, displayed a low number of higher CEFR word levels, totaling 52 words (C1 and C2 type counts). This trend is similar for BTMI students, who had the lowest number of higher CEFR levels (21 for C1 and C2 type counts), and the second lowest AWL coverage (8.83%) as shown in Table 4.

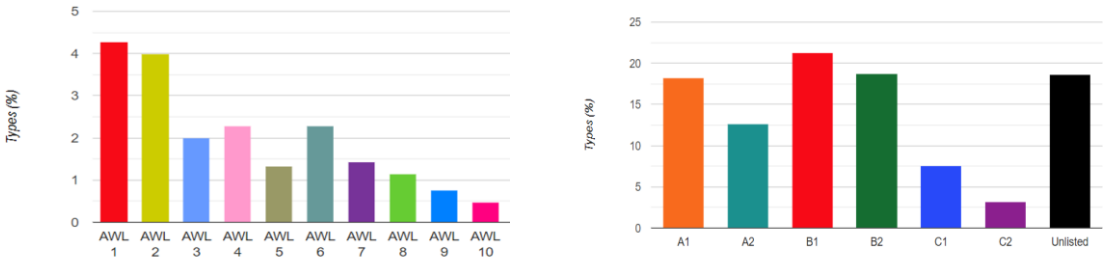
Programs demonstrating higher CEFR levels (C2 and above), such as BITZ, BEKM, and BMCG, also tend to show a higher coverage of AWL token usage, with respective percentages of 12.9%, 10.02%, and 13.21% (as shown in Table 4).

Table 6
Distribution of CEFR word lists by programs

	BENG	BTMI	BMCG	BEET	BMMF	BITZ	BMFG	BEKM
A1 type count	190	234	203	211	207	218	171	218
A2 type count	131	153	140	152	137	148	125	148
B1 type count	221	181	184	189	182	202	197	202
B2 type count	195	154	154	182	118	188	150	188
C1 type count	78	28	58	57	32	53	57	53
C2 type count	33	21	20	21	20	17	22	17
Unlisted type count	216	76	154	202	123	225	262	225

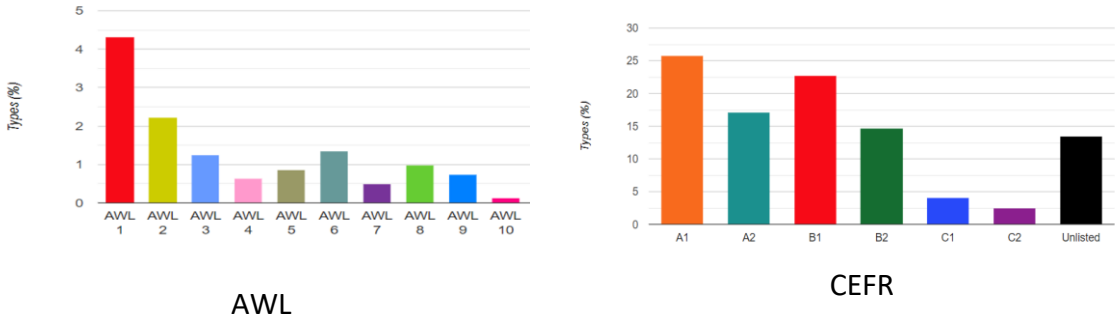
Tables 4 to 6 suggest that students of certain programs (e.g., BENG, BITZ, BEKM) demonstrate a more robust use of both AWL and higher CEFR level words, indicating a higher level of academic language proficiency. In contrast, programs like BMMF and BTMI, which exhibit lower usage of AWL and CEFR higher-level words, may indicate less exposure to academic language exposure or lower proficiency. This finding implies that a stronger grasp of academic vocabulary enhances overall language proficiency. Students who are proficient in using AWL words tend to attain higher CEFR levels (B2+ and above). This relationship supports the notion that mastery of academic vocabulary serves as an indicator of advanced language proficiency (Gardner & Davies, 2014).

To further illustrate the relationship, Figures 2 and 3 compare the distributions of AWL and CEFR word lists for BENG (the highest AWL coverage) and BMMF (the lowest AWL coverage). The data reveal that BENG, with the highest AWL coverage, also exhibits a high coverage of higher CEFR word levels, ranging from B2 to C2. In contrast, BMMF, with lower AWL coverage, shows a higher coverage of lower CEFR word levels, specifically from A1 to B1. Notably, BENG has more unlisted words for CEFR than BMMF, suggesting that BENG students may employ more subject-specific vocabulary. Unlisted words include proper names of persons, places, businesses, technical or internet/media-related terms, which are not included in the CEFR word lists.



AWL CEFR

Figure 2. Distribution of AWL and CEFR word lists for BENG



AWL CEFR

Figure 3. Distribution of AWL and CEFR word lists for BMMF

To provide a clearer interpretation of the results, Figures 4 and 5 compare essays from the BENG (CEFR D1+) and BMMF (CEFR B2+) programs, highlighting the differences in vocabulary usage and writing proficiency. Figures 6 and 7 provide a detailed breakdown of AWL word usage, highlighting the differences in academic vocabulary distribution between the two essays. Both essays discuss technical topics. Despite grammatical issues, the BENG essay demonstrates a clear understanding of the required 5-paragraph structure taught in the writing course, including three body paragraphs, with the third body paragraph providing a counterargument. The student was able to have more coverage of AWL words (Figure 6) throughout the essay. The BENG essay also demonstrates that the student managed to follow academic conventions, including citation of sources. The inclusion of proper names and years (numbers), which are unlisted words in the CEFR (EVP), may account for the higher number of unlisted words in the BENG essays. Therefore, this essay displays high academic quality, with formal and sophisticated vocabulary (language use).

Can AI Replace Humans in the Workplace

Artificial intelligence (AI) is revolutionizing various industries, spanning manufacturing, art, finance, and management. Considered the most significant general-purpose technology of our time, AI has introduced groundbreaking products, accelerated processes, and yielded superior outcomes (Brynjolfsson & McAfee, 2017). Contrary to the belief that human skills are indispensable, the present level of AI possesses the capacity to supplant humans in workplaces.

AI excels at data collection, management, and analysis, leveraging machine learning to potentially surpass human performance. This transformative technology reshapes business operations, disrupts management practices, and enhances efficiency and accuracy in tasks like risk assessment, financial analysis, and portfolio management (Agrawal et al., 2017; Lakhani & Iansiti, 2020; von Krogh, 2018), thereby boosting overall productivity.

Furthermore, when combined with robotics and automation, AI automates repetitive, monotonous, and hazardous tasks in sectors such as manufacturing. This integration drives heightened efficiency, precise outcomes, and reduced error rates at the operational level (Davenport & Kirby, 2015; Davenport et al., 2020; Paschen et al., 2020). Therefore reduce operational costs and increase profits.

Critics argue that AI falters in domains where human skills are indispensable, such as tasks requiring emotional intelligence, creativity, complex decision-making based on incomplete information, and interpersonal interactions. However, AI continues to advance, with ongoing efforts to incorporate human-like characteristics. The hospitality sector exemplifies successful AI implementation, where AI-enhanced systems have revolutionized hotel and tourism operations, bolstering efficiency and decision-making through property management systems, revenue management systems, and customer relationship management (Ruel & Njoku, 2021; Mariani et

al., 2018).

Figure 4. Essay by a BENG student (CEFR D1+)

In contrast, the BMMF essay shows simpler and straightforward academic word usage (Figure 5). The essay displays some structural issues, particularly the absence of the third body paragraph, that should be a counterargument for the topic. Despite discussing a technical topic, that is, types of tires, the essay features more general words, which are not listed in the AWL. This justifies the lower AWL coverage in this BMMF essay collection. Figure 7 displays the small coverage of AWL words in the essay, in comparison to the distribution of the AWL words in the BENG essay (Figure 6).

What type of tyre categories we should buy?

Nowadays there are too many types and different kind of tires that we can see at the tyre shop. As a person who drive a car, we always have a problem on thinking what type or tyre categories that we should buy. Even though those tyres are looking the same, we should not take it easy because each type of them really makes a difference.

First, tires that have more durability will last much longer. If a tyre have a softer rubber it will give you better performance but the tyre will tear down quickly (McCluskey Chevrolet, 2021). While the tyre that have harder rubber not tear down quicker than the softer tyre.

Second, better tires quality can improve your handling. Your car will have better grip on the road and does not lost control easily when you drive (Cube Creative Design, 2021). When you have a better handling, you can avoid accident and will be safe on the road. Some people might claim that all those cheap tyres also can be used for many years. However, those cheap tyres are the main factor where we can see there is a lot of car crash and accident while people driving on rainy day (PitStopArabia, 2020).

Figure 5. Essay by a BMMF student (CEFR B2+)

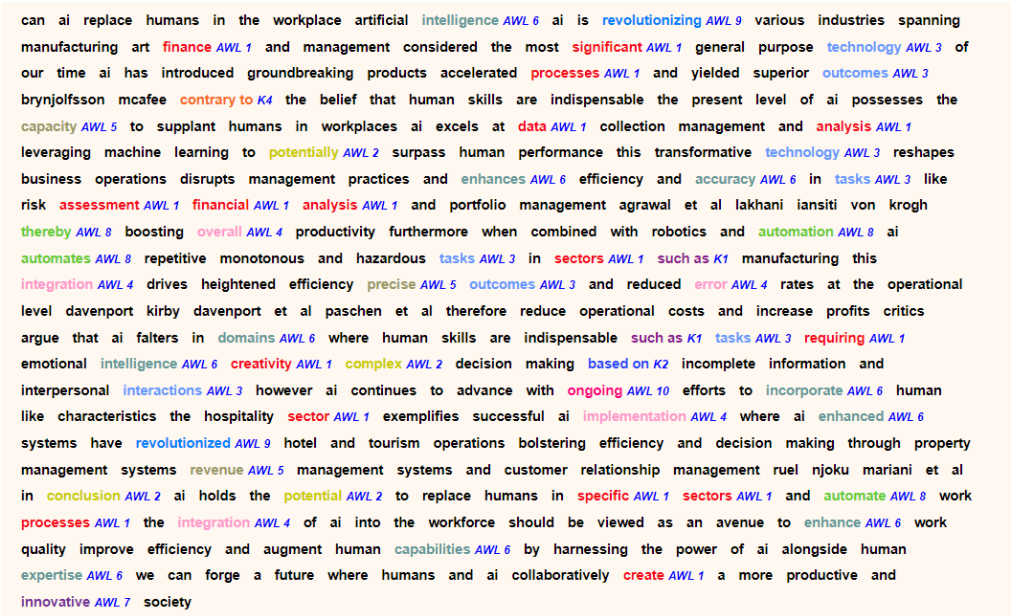


Figure 6. AWL word distribution in BENG student's essay

what type of tyre categories *AWL 2* we should buy nowadays there are *K1* too many *K3* types and different kind of *K3* tires that we can see at the tyre shop as a person who drive a car we always have a problem on thinking what type or tyre categories *AWL 2* that we should buy even though *K2* those tyres are looking the same we should not take it easy because each type of them really makes a difference first tires that have more durability will last much longer if a tyre have a softer rubber it will give you better performance but the tyre will tear down quickly mccluskey chevrolet while the tyre that have harder rubber not tear down quicker than the softer tyre second better tires quality can improve your handling your car will have better grip on the road *K4* and does not lost control easily when you drive cube creative *AWL 1* design *AWL 2* when you have a better handling you can avoid accident and will be safe on the road *K4* some people might claim that all those cheap tyres also can be used for many years however those cheap tyres are the main factor *AWL 1* where we can see there is *K1* a lot *K1* of car crash and accident while people driving on rainy day pitstoparabia in conclusion *AWL 2* tyres really make a difference you can just spend a few *K1* more money to have a better and good quality for your car there is *K1* no doubt *K3* on having a low quality tyres just for saving money

Figure 7. AWL word distribution in BMMF student's essay

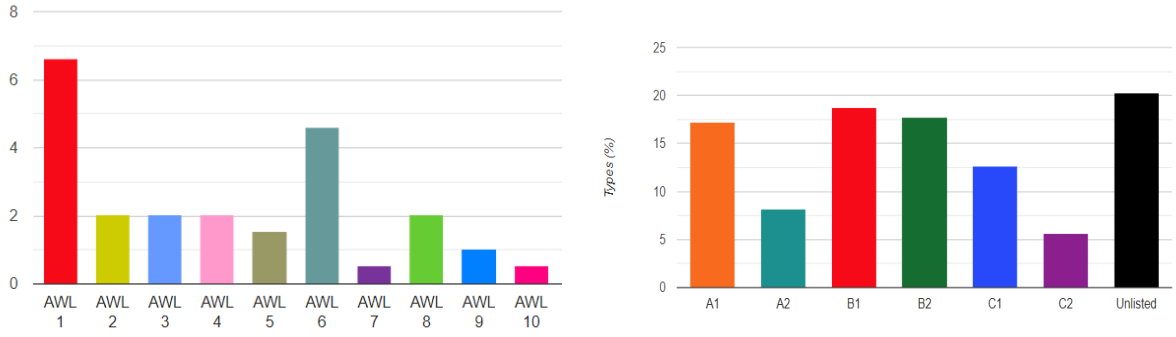


Figure 8. Distribution of AWL and CEFR word lists in BENG student's essay

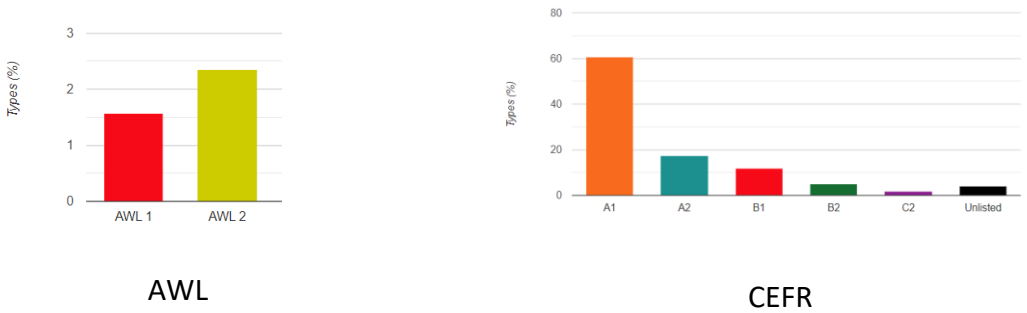


Figure 9. Distribution of AWL and CEFR word lists in BMMF student's essay

Figures 8 and 9 further illustrate the differences in the distributions of AWL and CEFR words between the two essays. With the CEFR achievement of D1+, the BENG essay displays a well-distribution of AWL and CEFR words, while the BMMF essay utilises limited AWL and CEFR words with the CEFR achievement of B2+. Additionally, the bar graphs also indicate that the number of AWL and CEFR words in the BMMF essay is far lower than the BENG essay. This clearly demonstrates that high reliance on both AWL and CEFR words can boost the academic writing performance of the students. This distribution suggests that while the BENG essay displays that the student had a solid grasp of most frequent academic vocabulary, the student

was also incorporating moderately advanced CEFR words, contributing to a more formal and sophisticated style of writing (Figures 4 and 5).

To address the research questions set for the study, it is found that there is a variation in the academic vocabulary (AWL) coverage across the disciplines. The percentage of AWL tokens in each discipline ranges from 6.34% in BMMF to 14.39% in BENG. Despite all students using academic vocabulary in their essays, there is a significant difference in its frequencies across the disciplines. As for the CEFR levels, the distribution across disciplines shows that many students exceed the targeted proficiency level (B2) for the writing course, with a significant proportion of learners in CEFR C2 and D1+ levels. This suggests that learners are developing a proficiency in using academic and subject-specific vocabulary, reflective of advanced language use. In particular, students in disciplines such as BENG, BITZ, and BEKM, display strong command over advanced vocabulary, achieving higher CEFR levels.

The findings suggest that the AWL plays a significant role in helping learners build academic vocabulary, which becomes increasingly important as they move toward higher CEFR levels. In other words, AWL vocabulary complements CEFR words in achieving higher writing performance. Higher AWL coverage, as observed in BENG, often correlates with a more formal and structured writing style, adhering closely to academic conventions. These higher levels, such as C2, D1, and D2, demand more sophisticated language use, particularly in academic contexts. This level of academic accuracy, coupled with the use of subject-specific topics, suggests that students with higher AWL usage are likely achieving higher CEFR levels.

Pedagogical Implication

The findings of this study offer several pedagogical implications that can significantly enhance the teaching and learning activities of the academic writing course. The following recommendations highlight how educators can adapt their instructional approaches to better support student development in academic writing.

- Because the students demonstrated performance beyond the targeted B2 CEFR proficiency level for this academic writing course, it would be appropriate to raise the target proficiency level to at least C1. This study proves that it is possible to assist the students to achieve any set target of CEFR proficiency levels. Furthermore, CEFR C1 level has been the language ability set by the Ministry of Education for learners upon leaving universities (Ministry of Education, 2015).

- The study demonstrates that students can be guided to produce advanced academic writing by leveraging topics and knowledge from their areas of specialisation. These topics often include words and concepts that are not only specialised but also sophisticated, aligning with the higher-level descriptors of the CEFR, particularly at advanced levels. This suggests that students' academic writing can reach higher proficiency levels when they engage with subject matter relevant to their fields. Therefore, to assist students achieve their targeted CEFR levels, this study recommends that instructors encourage students to discuss or write about topics related to their specialisations. While facilitating this, instructors should focus on teaching the conventions of academic writing (such as thesis statements, topic sentences, supporting

details, counterarguments, and refutations). Students should have the freedom to select topics within their field, while instructors guide them on the technical aspects of academic writing, such as coherence, cohesion, and mechanics. Given the niche area of this technical university, it is only appropriate to encourage the students to discuss topics of their discipline in preparing them for the workplace. According to Brabazon (2024), effective teaching practices that incorporate writing within disciplinary fields can enhance students' understanding of academic writing conventions.

- While the AWL focusses on academic vocabulary and the CEFR emphasises language proficiency, this study suggests that a balanced approach incorporating both can lead to more effective language learning outcomes. As noted by Redlich and Pattison (2024), combining AWL and CEFR can inform curriculum design, enabling educators to create materials tailored to specific academic contexts and learner needs. Therefore, instructional preparation should prioritise teaching academic structures derived from the AWL and CEFR word lists to support the writing tasks designed for the course. Many studies in ESL, EFL, and EAP settings have shown explicit vocabulary learning is effective in improving vocabulary learning (Tahir, Shah, Shak, Albakri, & Adnan, 2021). Mastery of vocabulary, in consequence, can significantly influence writing skills (Harli, Sartika, & Amelia, 2024; Laufer, 2024).

- This study shows that a learner corpus does more than just highlight language errors students may encounter; it can also reveal their potential competence, particularly within their specialised fields, in relation to CEFR language proficiency. As a result, this technical learner corpus can be further expanded and utilised for studies on both language errors and proficiency among technical students.

- It is crucial for language instructors to have a solid understanding of how the CEFR framework can be adapted and applied in teaching and learning. Combined with knowledge of the AWL, this understanding enables instructors to better support learners in reaching the targeted CEFR levels in their English courses. Input from both frameworks sets clear parameters for teaching strategies, guiding students toward achieving their target competency levels in academic writing. Studies show that using both frameworks can enhance vocabulary acquisition and retention, as they provide structured learning objectives and benchmarks for progress (Li, Hai, & Zhang, 2024).

Conclusion

The relationship between AWL and CEFR vocabulary plays a key role in developing academic writing proficiency. The findings suggest that students who employ both types of vocabulary tend to achieve higher CEFR levels and demonstrate stronger academic writing skills. This indicates that, pedagogically, instructors should integrate both academic and subject-specific vocabulary into their teaching strategies, aiming to cultivate more formal and sophisticated language use. By fostering a vocabulary-rich environment that emphasises AWL, CEFR, and subject-specific words, educators can better tailor their instructional approaches to improve academic writing outcomes across various disciplines.

The limitations of this study are worth noting. First, the analysis focusses solely on vocabulary, without examining grammatical features, which are essential components of

writing proficiency. Since writing quality is influenced by more than just vocabulary, the exclusion of grammar, sentence complexity, and cohesion markers limits the scope of the findings. Additionally, the study is restricted to a relatively small sample size, with only 10 essays analysed per program. This limited dataset may not fully represent each discipline's academic writing practices. To gain a more comprehensive understanding, future research should include qualitative analyses of grammatical structures, as well as larger and more diverse datasets. This would provide a fuller picture of students' writing proficiency.

Acknowledgements
The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for all the supports.

References

Amkham, C. (2016). Introducing the English Vocabulary Profile (EVP) to students: An attempt at enriching students' written language. In S. Borg, L. Patsko, L. Grimes, R. Phuchongcharoen, M. Rabbers, & R. Van Duin (Eds.), Final Report [Report]. https://www.englishprofile.org/images/pdf/2015-16_Amkham_CUP_TRP_final_report.pdf

Asaad, H. Q. M. (2024). The role of morphological awareness in L2 postgraduates' academic writing: is vocabulary knowledge a mediating variable? *Cogent Education, 11*(1), 1-17. https://doi.org/10.1080/2331186X.2024.2327787

Brabazon, T. (2024). Learning to write writing to learn : Anchoring student writing cultures with disciplinary literacies. *Journal of Languages and Translation*, *3*(1), 1-26. https://doi.org/10.70204/jlt.v3i1.304

Capel, A. (2015). The English vocabulary profile. *English Profile in Practice, 5*(1), 9-27.

Charpentier Jiménez, W. (2023). Complexity of vocabulary in the works of university students of ILE. *Journal of Modern Languages,* (37) 1-22. https://doi.org/10.15517/rlm.v0i37.50826

Chung, E., Wan, A., & Fung, D. (2024). Understanding academic vocabulary learning in higher education: Perspectives from first-year undergraduates in Hong Kong. *International Journal of Applied Linguistics*, 1–17. https://doi.org/10.1111/ijal.12576

Council of Europe. (2024). Global scale - Table 1 (CEFR 3.3): Common Reference levels. Common European Framework of Reference for Languages (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. https://www.lextutor.ca/myanmar/2/coxhead_2000.pdf

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15*(2), 363–381. https://doi.org/10.1080/15248372.2012.725187

Drayton, J., & Coxhead, A. (2023). The development, evaluation and application of an aviation radiotelephony specialised technical vocabulary list. *English for Specific Purposes, 69*, 51-66. https://doi.org/10.1016/j.esp.2022.10.001

Folse, K. S. (2023). *Academic Word Lists: What Every Teacher Needs to Know*. University of Michigan Press.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Guihua, Luo. & Jiamiao, Song. (2023). Research on General Word List and Its Application. *Education Research and Development, 2*(5), 98-107. https://doi.org/10.57237/j.edu.2023.05.002

Harli, Y. T., Sartika, D., & Rizki Amelia, K. R. (2024). The relationship between the eight-grade students' vocabulary mastery and their writing achievement at smp pgri pedamaran. *Language and Education Journal, 9*(1), 125-138. https://doi.org/10.52237/lej.v9i1.727

Haruna, H., Tajuddin, A. J. A., Bashir, I. (2024). A corpus-based lexical coverage of coursebooks in Nigeria: A case study. *Register Journal*, *17*(1), 164-188. https://doi.org/10.18326/register.v17i1.164-188

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430. http://hdl.handle.net/10125/66973

Khani, R., & Javadian, S. (2024). Development of philosophical academic word list (Phawl). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4914684

Lahuerta, A. C. (2023). Analysis of changes in L2 writing over the time of a short-term academic English programme. *Porta Linguarum Revista Interuniversitaria De Didáctica De Las Lenguas Extranjeras,* (39), 111–127. https://doi.org/10.30827/portalin.vi39.22842

Larsson, S. (2017). *Vocabulary profiles of English language learning textbooks : A lexical analysis of textbooks used in EFL classrooms* (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-64840

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for academic purposes, 22*, 42-53. https://doi.org/10.1016/j.jeap.2016.01.008

Li, L., Hai, X., & Zhang, X. (2024). Creation and application of an academic spoken collocation list. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3339

Ma, L. (2023). Source use in integrated writing among Chinese college ESP learners with different language proficiency. *Praxis International Journal of Social Science and Literature*, *6*(8), 181-194. https://doi.org/10.51879/pijssl/060821

Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, *49*, 100540. https://doi.org/10.1016/j.asw.2021.100540

Matsuzaki, T., & Mark, K. (2020). Investigating the difficulties for university learners of English in Japan of CEFR B1-level phrases. *CEFR Journal - Research and Practice, 3*, 59 – 86. https://doi.org/10.37546/JALTSIG.CEFR3-4

McKeown, M., Crosson, A., Moore, D., & Beck, I. (2018). word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal, 55*(3), 572-616. https://doi.org/10.3102/0002831217744181

Ministry of Education. (2015). *English Language Education Reform in Malaysia: The Roadmap 2015-2015*. English Language Standard and Quality Council.

Muhammad, K. Q., & Muhammad, H. A. (2023). A study of academic vocabulary use by advanced EFL university students. *Journal of Kurdish Academy*, (56), E7-E38.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.

Nuoya, Z. (2024). Comments on linguistic features of academic writing: Based on the article "Testing a model of teaching for anxiety and success for English language teaching.". *Lecture Notes on Language and Literature, 7*(4), 85-189. https://doi.org/10.23977/langl.2024.070427

Özer, M., & Akbaş, E. (2024). Assembling a justified list of academic words in veterinary medicine: The veterinary medicine academic word list (VMAWL). *English for Specific Purposes, 74*, 29-43. https://doi.org/10.1016/j.esp.2023.12.002

Redlich, J., & Pattison, S. (2024). Combining cefr and 4c clil frameworks for principled literature in language teaching. *Nobel: Journal of Literature and Language Teaching*, *15*(1), 1-21. https://doi.org/10.15642/nobel.2024.15.1.1-21

Roesler, D. (2021). When a bug is not a bug: an introduction to the computer science academic vocabulary list. *Journal of English for Academic Purposes, 54*, 101044. https://doi.org/10.1016/j.jeap.2021.101044

Run, W., & Weina, L. (2024). Vocabulary thresholds study for business English major based on the range corpus. *International Journal of Strategic Research in Education Technology and Humanities*, *12*(1), 20-38. https://doi.org/10.48028/iiprds/ijsreth.v12.i1.02

Sarimah, S., Noraini, H., & Amerrudin, A. M. (2013). Exploring fundamental engineering word list for engineering students: A literature review. *Procedia - Social and Behavioral Sciences, 70*, 1275-1281. https://doi.org/10.1016/j.sbspro.2013.01.187

Shuxratovna, K. S., (2024). Importance of linguistic competence and factors influencing on the issue. *Journal of Learning on History and Social Sciences, 1*(5), 13–16. https://doi.org/10.61796/ejlhss.v1i5.530

Sun, D. (2017). A contrastive analysis between English Vocabulary Profile and college English wordlist. *Theory and Practice in Language Studies, 7*(9), 729-736 http://dx.doi.org/10.17507/tpls.0709.04

Tahir, M., Shah, D., Shak, M., Albakri, I., & Adnan, A. (2021). Explicit vocabulary instruction: Effects of vocabulary learning on form two ESL learners. *Studies in English Language and Education*, *8*(3), 1227-1247. https://doi.org/10.24815/siele.v8i3.19539

Thaksanan, P., & Chaturongakul, P. (2023). Enhancing EFL students' performance and genre awareness in academic writing through genre-based instruction. *Arab World English Journal (AWEJ)*, *14*(12), 295-311. https://doi.org/10.2139/ssrn.4497563

Therova, D. (2023). Core academic vocabulary in four genres of novice student writing. *TESL-EJ, 27*(1). https://doi.org/10.55593/ej.27105a8

Uba, S. Y., Irudayasamy, J., & Hankins, C. A. (2023). A new computer science academic word list. *World, 13*(5), 319-330.

Vuković-Stamatović, M. (2024). Creating and validating a corpus-based English academic word list for physics. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*. https://doi.org/10.1075/resla.22041.vuk Weblingua Ltd. (n.d.). *Text Inspector*. https://textinspector.com/who-are-we/

Zaid, T. A. A. (2024). Lexical knowledge and its effects on language proficiency: a synonymous study. *Journal Ink of Literature, 14*(35), 1759-1774. https://doi.org/10.58564/ma.v14i35.1381

Zhang, J., Lopez Wui, M. G., Nam, R., Eunjung Relyea, J., & Wong, S. S. (2022). Improving argumentative writing of sixth-grade adolescents through dialogic inquiry of socioscientific issues. *Journal of Writing Research, 14*(3), 375–419. https://doi.org/10.17239/jowr-2023.14.03.03