RESEARCH ARTICLE VOLUME 19, ISSUE 1, 51 – 70

The New English Teacher
ISSN 2985-0959 (Online)

Lexical Collocation Analysis in a Corpus of Research Articles in Physical Education and Sport Science

Piriya Thaksanan
Faculty of Liberal Arts and Management Science
Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus
59/6 Moo 1, Vor Por Ror Or 366 Road
Chiang Khruea, Sakon Nakhon, Thailand, 47000
Tel: +6609307367509
Email: peereeya.t@ku.th

Thiratchapon Kamsa-ard
The Office of General Education
Udon Thani Rajabhat University
64, Thahan Road, Thambon Makkaeng,
Muang district, Udon Thani, Thailand, 41000
Tel: +660829208737
Email: Thiratchapon@gmail.com

Abstract: Understanding how words naturally come together in language is critical for developing fluency and natural expression when studying a second or foreign language. However, many language learners struggle with collocations because they have not been taught how words normally combine. The purpose of this study was to look at keywords and lexical collocations in research articles on physical education and sport science that have been published in reputable journals. A corpus called COPES was created, consisting of 737,901 running word tokens taken from 144 high-quality scientific articles in the field of physical education and sport science. These articles were examined with AntConc version 4.2.4, a freeware corpus analysis toolkit for concordancing and text analysis. From this analysis, the first 200 keywords were identified. It was observed that the majority of these words functioned as nouns (64 %), followed by verbs (20.5%), adjectives (14%), and adverbs (1.5%). These keywords were then used as nodes to uncover their collocations. The study found that the most prevalent collocations were noun + noun, adjective + noun, noun + verb, and verb + noun. The study also delved into the pedagogical implications and provided recommendations for future research endeavors.

Keywords: corpus analysis, keyword list, lexical collocations, physical education, sport science

Received: June 3, 2024 Revised: September 25, 2024 Accepted: November 13, 2024

Introduction

Apart from having sufficient knowledge of academic word lists in a specific discipline, knowing how to use those academic words correctly is even more important. Although most language learners know many words, many still exhibit a deficiency in collocation knowledge due to a lack of collocation instruction (Hashemi, Azizinezhad, & Dravishi, S., 2012; Khonamri, Ahmadi, Pavlikova, & Petrikovicova, 2020; Liontas, Bangun, & Li, 2023). Collocations are important and have particular characteristics, making their use essential for enhancing learners' language fluency and for achieving native-like proficiency (Sun & Park, 2023). The importance of collocation education thus

cannot be underestimated. Collocations are essential for expanding L2 vocabulary and improving communication skills. As a result, collocation instruction should be prioritized in the classroom, particularly among non-native speakers. Collocational knowledge is necessary for developing native-like fluency and idiomaticity in second and foreign language learning (Ellis & Ogden, 2017). Consequently, a lack of collocation comprehension leads to errors among EFL learners, lowering their fluency. Students' inattention to the occurrence of collocations often results in their inability to use collocations effectively (Boers, Demecheleer, Coxhead, & Webb, 2014). In acquiring collocations, language use is highly structured, and phraseological patterns and collocations are very important. Psycholinguistic research shows that our language processing is sensitive to the statistical regularities of these patterns, such as in verb-argument constructions, and learning mechanisms enable people to become aware of these patterns (Ellis & Ogden, 2017).

Corpora are valuable tools in corpus linguistics for determining the regularity, frequency, and distribution of formulaic patterns in language. In this field, significant emphasis has been placed on identifying co-occurring patterns of linguistic elements and on defining these formulaic units, as reported in language corpora (Gablasova, Brezina, & McEnery, 2017); Hamed, 2021; Jaafar, 2022). For decades, discipline-specific academic word lists (DSAWL) have been researched across various disciplines (e.g., Alasmary, 2022; Chanasattru & Tangkiengsirisin, 2016; Eguchi & Kyle, 2023; Yotimart, 2021; Yulfi, Seli, & Ariska, 2019). Studies compiling DSAWLs have found that medical disciplines have the most studied word lists (Saeedi, Khany, & Tazik, 2023). The relationship between specific subfields are also studied. For instance, within these medical academic word lists (AWLs), the top three most studied areas were pharmacology, medicine, and nursing. The least explored areas related to the medical discipline were health and life science, physical education and sports science, and health information management (Saeedi et al., 2023).

Although the number of corpus studies is increasing, some disciplines still require further exploration. Saeedi et al. (2023) conducted a systematic review and found that many disciplines have yet to be investigated. Furthermore, almost all previous studies focused on word lists on a single-word basis, with relatively few studies examining the collocation lists of academic word lists. Based on these findings, this study aimed to fill this research gap by conducting a corpus-based investigation of the DSAWL in physical education and sports science. This study also explored the most frequent words and their collocates. The results can be beneficial for undergraduate and graduate students in understanding collocations within their academic disciplines. Additionally, teachers can use the collocation list to design and deliver instructions more effectively. Therefore, this study sought to explore the following research questions:

1. What are the main keywords identified in a corpus of physical education and sports science research articles?

2. What are the common lexical collocations associated with the keywords from the corpus of physical education and sports science research articles?

Literature Review

Vocabulary is the most important component of language learning. It is fundamental to language proficiency, forming the foundation for effective speaking, listening, reading, and writing skills in learners. A lack of vocabulary knowledge is a primary factor that prevents language learners from effectively communicating in English (Richards & Renandya, 2002). The vocabulary knowledge of language learners is a key dimension in assessing their language proficiency and ability. As Coxhead (2021, p. 3) mentioned, “Learners with a large vocabulary in English usually have high levels of proficiency.” Numerous research studies have examined the relationship between learners' vocabulary knowledge and academic performance in this way (e.g., Heeren et al., 2021; Masrai &

Milton, 2017; Okkinga, van Gelderen, van Schooten, van Steensel, & Sleegers, 2023; Szabo, Stickler, & Adinolfi, 2021).

Nation (2013) divides words into four categories based on frequency levels: high-frequency, mid-frequency, low-frequency, and specialized vocabulary. High-frequency words, which comprise around 2,000 word families and include function words (such as "'a,' "'some,' 'two,' 'because,' and 'to'), are critical because they account for a significant proportion of running words in spoken and written texts and appear in a wide range of language uses. Mid-frequency words consist of 7,000 word families that cover the majority of text types. Low-frequency words, often characterized by their infrequent appearance in texts, are predominantly proper nouns. Moreover, specialized vocabularies extend high-frequency words for specific uses by systematically restricting the range of topics or language uses. Furthermore, technical vocabulary is a type of specialized vocabulary. Some vocabulary lists are created by performing frequency counts on a particular corpus (Coxhead, Rahmat, & Lu Yang, 2020; Ma & Qian, 2020; Pinchbeck, Brown, Mclean, & Kramer, 2022), while others are compiled by specialists in the field who collect what they believe to be important terms (Hanks, Egbert, & Hashimoto, 2024; Özer & Akbaş, 2024; Tongpoon-Patanasorn; 2018).

Academic Word List (AWL) and Discipline-Specific Academic Word List (DSAWL)

Therova (2020) classifies an Academic Word List (AWL) into two categories: 1) a word list that compiles academic vocabulary from various academic disciplines (e.g., English for General Academic Purposes [EGAP]) and 2) a word list that contains academic vocabulary specific to particular academic disciplines (discipline-specific word lists). EGAP words are found across academic areas (Coxhead, 2020), meaning that learners will encounter these words in their studies regardless of their discipline. Much research has focused on implementing AWLs to guide students' learning of academic vocabulary (Durrant, 2016; Csomay & Prades, 2018; Lawrence, Knoph, McIlraith, Kulesz, & Francis, 2022).

However, some scholars argue that the AWL approach might not be fully applicable since the lists created by Coxhead (2000) cover around 10 percent of vocabulary across all disciplines, but their relevance varies greatly among different fields. Up to a quarter of AWL words are irrelevant to at least one discipline (Hyland & Tse, 2007). Additionally, using word families to determine word frequencies does not accurately reflect actual word usage (Gardner & Davies, 2014). Furthermore, the AWL does not include part-of-speech information, making it difficult for users to discern which words can be used across multiple grammatical categories or to determine the frequency of specific grammatical categories (Green & Lambert, 2018). Therefore, a discipline-specific AWL is necessary to compile subject-specific vocabulary, preparing learners for study in that domain or testing vocabulary knowledge in a specific area.

Discipline-specific Academic Word Lists (DSAWL), also known as discipline-based lexical repertoires (Hyland & Tse, 2007) or field-specific academic word lists (Martinez, Beck, & Panza, 2009), focus on academic words from a single discipline (Chang, 2023). These lists contain words that are semantically and/or grammatically linked to a given discipline (Saeedi et al., 2023). According to Nation (2016), each academic subject has a unique vocabulary closely related to its content. This vocabulary is more technical than broad, is directly linked to the subject matter, and is less likely to be known by those unfamiliar with the subject.

Recently, an increasing number of corpus-based studies have investigated DSAWL across diverse disciplines. For example, Coxhead and Demecheleer (2018) studied the technical vocabulary of plumbing, Hsu (2018) investigated the most frequent BNC/COCA mid- and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks, and Gilmore and Millar (2018) conducted a corpus analysis of the language in civil engineering research articles. Valipouri

and Nassaji (2013) conducted a corpus-based study to compile a list of academic vocabulary commonly found in chemistry research articles to assist EFL chemistry students and to assess how this list aligns with the distribution of high-frequency words in Coxhead's (2000) AWL and West's (1965) General Service List (GSL) within the research articles. The researchers created a four-million-word corpus from 1,185 chemistry research articles, revealing 1,400-word families frequently found in the Chemistry AWL target corpus. The study also found that the technical words in the target corpus were not frequently found in the AWL or the GSL, highlighting the importance of creating specialized vocabulary lists tailored to specific fields, based on the genres and texts relevant to students' academic disciplines.

Collocations

A collocation refers to a sequence of words that occur together more frequently than expected by chance (Ellis & Ogden, 2017). These word combinations help us recognize common patterns and demonstrate how words typically behave within specific styles or contexts (Hyland, 2007, p. 168). According to Nation (2013), understanding which words naturally go together is essential for language proficiency and fluency. Learning collocations is crucial because many words consistently follow specific usage patterns. Collocations are highly significant and can capture attention, underscoring their essential role in improving learners' fluency and helping them reach a proficiency level similar to that of native speakers (Durrant & Mathews-Aydinli, 2011; Sun & Park, 2023).

Numerous studies have investigated L2 students' proficiency in using collocations. Men (2018) examined the developmental patterns of verb-plus-noun collocations in Chinese learners across different proficiency levels. The results showed that more proficient learners employed a greater variety of collocations compared to other proficiency groups. Similarly, Garner (2022) explored how verb-noun collocations are employed across different levels of second language (L2) writing proficiency. This study focused on students' ability to produce construction grammar to define verb-noun collocations and their development in L2 writing. The findings indicated that more proficient writers utilized a wider range of verb-noun collocations than lower-proficiency students.

According to Goulart (2019), collocations have restricted commutability, meaning their node words can only co-occur with a limited set of other words. Furthermore, they typically consist of two or three words with specific grammatical components. Table 1 below displays lexical collocations in this manner, as categorized by Benson, Benson, and Ilson, (2010).

Table 1

Examples of Collocation Combinations in English Adapted from Benson et al. (2010)

Combinations	Examples
verb + noun	To face problems, to bend the rules
adjective + noun	Practical activities, warmest regards
noun + verb	Bombs explode, ideas flow, blood circulates
noun + noun	Education system, aisle seat, learning experience
adverb + adjective	Heavily influenced, highly successful
adverb + verb/ verb + adverb	Hurt badly, increase dramatically

Similar to vocabulary acquisition, collocations can be learned under two main conditions: repetition of words and quality of processing of the target words (Nation, 2017). To enhance students' collocation knowledge and maintain their long-term retention, students must have a chance to be exposed to the target words in at least five spacing repetitions (Macis, Sonbul, & Alharbi, 2021;

Peters, 2014; Webb, Newton, & Chang, 2013). Numerous studies have investigated the effects of spacing on vocabulary learning. For instance, Macis et al. (2021) examined how spacing influenced the ability of 55 Arabic students to learn collocations while studying for an undergraduate degree. The study's outcomes revealed that spacing repetitions positively affected the students' collocation knowledge and their long-term retention.

Another crucial factor in vocabulary acquisition is the quality of processing in memory, both incidental and deliberate attention (Nation, 2017). Incidental vocabulary learning or implicit learning relies on how students unconsciously acquire vocabulary from meaning-focused encounters in various contexts (Zhang, 2022). The latest findings on incidental learning, particularly the learning of collocations, have confirmed the positive effects of multimodal input (Pu, Chang, & Wang, 2024; Celik, 2024). While incidental vocabulary learning focuses on unintentional word learning from multimodal input, explicit (deliberate) learning involves direct teaching, where teachers provide deliberate instruction focusing on word form and meaning. Numerous studies have revealed the benefits of explicit instruction on students' vocabulary learning. For instance, Liontas et al. (2023) examined the efficacy of explicit lexical collocation instruction on students' lexical collocation and writing performance via the Corpus of Contemporary American English (COCA). The outcomes of the study demonstrated the positive effects of explicit instruction on the students' competence in learning collocations.

Related Studies

Some previous studies have explored lexical collocations across a range of registers. For example, Sukman, Triwatwaranon, Munkongdee, and Chumnumnawin, (2022) investigated lexical collocations in business news articles from 2020 to 2021. The news articles were drawn from an online news agency comprising seven hundred articles, totaling 633,895 running word tokens and 23,284 word types. The results of the study revealed that the most important words found in the business news articles pertained to business activities, the coronavirus pandemic, economics and finance, technology and social media, and wars.

Similarly, Suraprajit (2022) conducted a corpus-based study exploring high-frequency words and collocations used in logistic magazines. Raw data were retrieved from two online logistic magazines, including 106,385 running word tokens and 8,007 word types. The findings showed that the top ten most frequent words, nouns, verbs, and adjectives occurred in the logistic magazines. The study's outcomes suggested that the highest number of grammatical collocations found constituted noun + preposition, while the most common form of lexical collocations was noun + verb.

Another corpus-based study by Mandić and Dankić (2020) compiled 262 nursing scientific articles from ten high-quality journals from 2017 to 2018 to determine the most frequent two-word collocations in the target corpus. Consequently, the most frequent two-word collocation was noun + noun, followed by adjective + noun.

A more recent study regarding formulaic language was conducted by Pujiningtyas and Bram (2023). The study attempted to search for lexical collocations in students' reflective writing at the master's level. The data were drawn from 18 reflective writings written by English Education master's program students. The most common lexical collocation pattern found in this study constituted adjective + noun, followed by verb + noun. In contrast, noun + verb, noun + noun, and adverb + adjective were the least frequent collocation types.

In summary, it is well known that the ability to understand and use collocations is a key indicator of EFL learning performance. Previous research has demonstrated that different disciplines use high-frequency terms that are specific to their fields of study. Additionally, the frequency of collocations

varies across academic contexts. While there is an existing study on word lists and collocations in numerous domains, there remains a gap that requires further investigation. As a result, the purpose of this study was to identify high-frequency words in physical education and sport science research. Furthermore, the collocations revealed in this study are useful for improving English language education in this particular field.

Method

Data Collection and Analysis

This corpus-based study investigated keywords and lexical collocations frequently occurring in research articles published in reputable physical education and sports science journals. A total of 144 research articles from ten journals between 2019 and 2023 were selected. In the current study, physical education and sport science were chosen because of their interconnectedness, as evidenced by previous research (Kongcharoen, 2018), along with some current systematic reviews that have also highlighted the relationship between physical education and sport science (Ojeda-Nahuelcura, Carter-Thuillier, López-Pastor, & Fuentes-Nieto, T. 2023; Chiva-Bartoll, Salvador-García, Pérez-Samaniego, & Flórez-García, 2019). Moreover, specialized academic journals frequently publish research that overlaps these two disciplines, reinforcing the relevance of the combination.

Three experts in physical education and sport science, each with extensive experience in higher education, recommended a selection of international academic journals. These experts, who teach both undergraduate and master's level courses and hold doctoral degrees, are assistant professors with prior publications in international journals. In the journal selection process, the experts provided recommendations, ranking the journals from most to least favored. A total of ten journals were recommended. Afterward, access was obtained to each journal's database, and the accessible journals were downloaded for further review. Then, the researchers compiled a list of journals relevant to physical education and sport science. Subsequently, the researchers selected research articles from each journal. Once received, the articles were screened so that research articles written by the same person were not included.

The 144 research articles were reformatted into plain text, and irrelevant parts such as tables, references, and images were removed, retaining only the parts relevant to the analysis. The researcher then created the Corpus of Physical Education and Sport Science (hereafter, the COPES corpus). The software employed was AntConc Version 4.2.4, developed by Anthony (2023). The generated COPES corpus consists of 737,901 running word tokens and 24,888 word types, which was then subjected to analysis with AntConc. The frequency of words was selected following Baker's (2006) guidance (a keyness score of ≥ 100). Function words, such as *the*, *which, although,* or *and,* together with pronouns, proper nouns, function words, abbreviations, and acronyms were screened out, ensuring that only the most frequent content words were obtained (Dang & Webb, 2016).

The keyword lists were chosen based on their statistically significant frequency of occurrence compared to the British National Corpus (BNC), which was used as a reference corpus. In this study, the COPES corpus was determined by considering word families which are regarded as a headword or base word, including all its inflected forms and transparent derivations (Coxhead, 2000). For example, the word family 'perform' contains the following members:

perform (headword)	*performs*	*performed*	*performing*
performance	*performer*	*performative*	

To determine the part of speech or grammatical class distribution of the most frequently used terms in the physical education and sports science articles, the TagAnt program developed by Anthony (2022) was utilized. Additionally, an inter-coder technique was employed to verify the accuracy of the results.

The top 200 keywords found in the analysis were regarded as 'nodes' for further analysis. Once the keywords were determined, the researcher used the nodes to proceed to identify their collocates. The AntConc collocate function was applied to find a list of common collocations within a three-word span to the right of the node (3R). The minimum collocate frequency was set as '10', with the MI value at ≥ 3.

Findings and Discussion

List of Top 200 Keywords Found in Physical Education and Sports Science Research Articles

The first research question sought to investigate the frequency of words used in research articles, specifically in physical education and sport science. Data extracted from 144 research articles were analyzed to identify the most frequent 200 word families occurring as content words. The results are presented in Table 2.

Table 2

List of High Frequency Words Extracted from Physical Education and Sport Science Research Articles

No.	Headword	Freq_Tar	Keyness (Likelihood)	No.	Headword	Freq_Tar	Keyness (Likelihood)
1	physical	2843	3817.88	111	Positive	534	397.262
2	teachers	2157	3717.776	112	Sample	230	396.078
3	students	2631	2874.14	113	Aim	227	390.911
4	training	1949	2493.967	114	Test	575	390.057
5	participants	1274	2194.967	115	practical	221	380.577
6	teaching	1180	2032.928	116	movements	219	377.133
7	sport	1049	1807.131	117	challenges	218	375.411
8	research	2074	1796.297	118	Authors	218	375.411
9	experiences	1041	1793.343	119	Specific	545	372.584
10	study	2042	1737.979	120	Focused	215	370.244
11	learning	1480	1695.044	121	negative	213	366.799
12	sports	963	1658.912	122	developing	210	361.633
13	athletes	898	1546.894	123	Session	209	359.911
14	activity	1301	1483.587	124	Psychological	209	359.911
15	teacher	798	1374.572	125	participate	207	356.466
16	activities	751	1293.586	126	characteristic	203	349.577
17	education	1454	1276.795	127	Tasks	203	349.577
18	coaches	731	1259.125	128	Limited	202	347.855
19	skills	1028	1165.088	129	competition	201	346.133
20	curriculum	645	1110.949	130	Used	1149	342.284
21	fitness	588	1012.746	131	Group	937	342.089
22	body	1208	965.323	132	Healthy	196	337.522
23	findings	547	942.112	133	appropriate	196	337.522
24	content	536	923.161	134	Criteria	196	337.522
25	performance	1021	909.956	135	Setting	195	335.8
26	movement	913	886.457	136	Actions	195	335.8
27	knowledge	995	866.622	137	Explore	194	334.077
28	lessons	492	847.363	138	Prior	194	334.077
29	researchers	492	847.363	139	informed	194	334.077
30	student	484	833.581	140	Task	193	332.355
31	strength	482	830.136	141	Field	436	329.468
32	practices	469	807.742	142	researcher	191	328.911
33	participation	460	792.238	143	processes	189	325.466

No.	Headword	Freq_Tar	Keyness (Likelihood)	No.	Headword	Freq_Tar	Keyness (Likelihood)
34	assessment	912	788.601	144	multiple	189	325.466
35	health	1368	743.237	145	reflection	188	323.744
36	develop	407	700.942	146	exercise	482	322.437
37	interviews	399	687.161	147	statistical	186	320.3
38	influence	396	681.994	148	effective	185	318.578
39	school	1283	678.97	149	perspectives	184	316.855
40	motor	394	678.549	150	competitive	184	316.855
41	practice	816	659.739	151	Team	592	315.607
42	gender	383	659.601	152	relevant	183	315.133
43	conducted	367	632.041	153	instance	183	315.133
44	data	1200	629.277	154	secondary	183	315.133
45	factors	365	628.596	155	considering	182	313.411
46	analysis	905	611.277	156	Phase	181	311.689
47	educational	352	606.205	157	Create	181	311.689
48	pedagogical	352	606.205	158	Assess	181	311.689
49	social	1215	602.837	159	demonstrated	181	311.689
50	development	914	590.35	160	inclusion	179	308.245
51	perspective	334	575.201	161	responses	179	308.245
52	approach	731	555.514	162	elements	179	308.245
53	outcomes	320	551.087	163	Level	758	306.658
54	games	320	551.087	164	interaction	178	306.522
55	exercises	317	545.92	165	framework	178	306.522
56	perceived	316	544.197	166	methods	426	306.299
57	importance	314	540.753	167	technique	177	304.8
58	critical	313	539.03	168	situations	176	303.078
59	professional	311	535.586	169	Reflect	175	301.356
60	motivation	310	533.863	170	highlighted	174	299.634
61	values	309	532.141	171	obtained	174	299.634
62	literature	307	528.696	172	Theory	174	299.634
63	process	839	526.423	173	technical	174	299.634
64	sessions	305	525.251	174	concepts	173	297.912
65	variables	302	520.084	175	differences	462	292.873
66	related	628	503.756	176	indicated	169	291.023
67	aspects	290	499.416	177	Initial	169	291.023
68	goals	287	494.249	178	Identify	169	291.023
69	factor	286	492.526	179	Using	787	287.731
70	method	283	487.359	180	components	166	285.856
71	opportunities	283	487.359	181	providing	166	285.856
72	performed	279	480.47	182	meaning	164	282.412
73	classes	277	477.025	183	Applied	163	280.69
74	results	754	475.84	184	collection	163	280.69
75	experience	719	470.465	185	promote	162	278.968
76	relation	273	470.136	186	Affect	161	277.246
77	male	273	470.136	187	Ability	399	276.671
78	themes	272	468.414	188	confidence	159	273.801
79	relationships	271	466.691	189	interview	401	273.288
80	mental	270	464.969	190	Program	158	272.079
81	approaches	269	463.247	191	Ideas	157	270.357
82	improve	269	463.247	192	questions	477	267.8
83	author	267	459.802	193	discussed	154	265.191
84	based	980	456.284	194	capacity	153	263.468
85	focus	558	455.187	195	techniques	153	263.468
86	purpose	262	451.191	196	Understanding	467	262.984
87	strategies	261	449.468	197	Achieve	152	261.746
88	different	1054	448.706	198	Noted	152	261.746
89	skill	260	447.746	199	Defined	151	260.024
90	learn	257	442.579	200	influenced	148	254.858

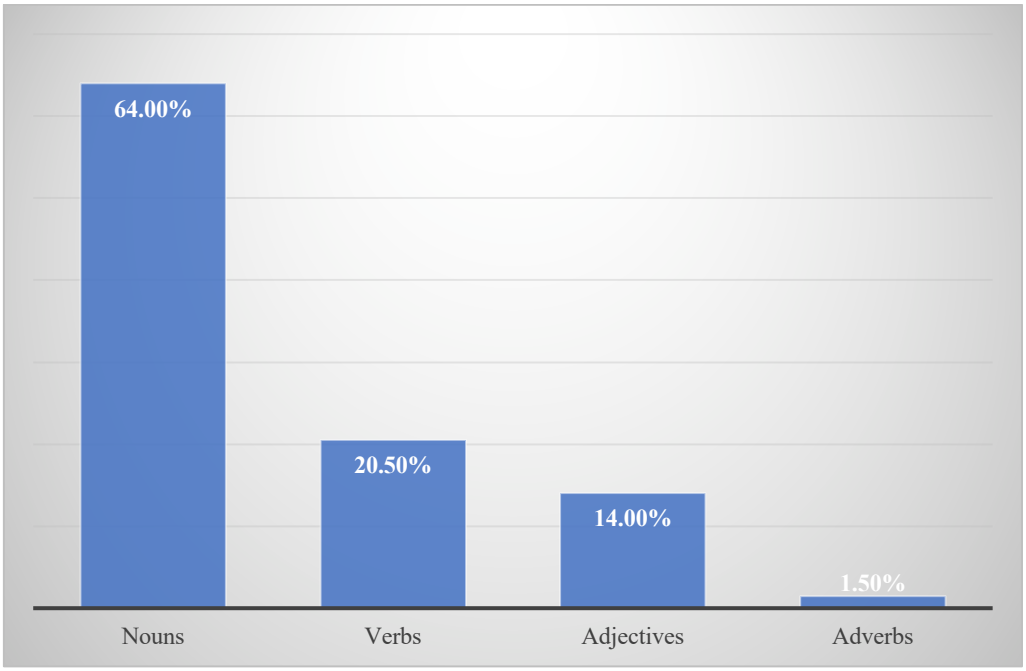
Table 2 displays 200 noteworthy terms found in research articles published in reputable physical education and sport science journals. These findings represent the most frequently occurring words in research articles related to these fields. Of the two hundred most frequent words, 'physical' is the

most frequently occurring, appearing 2,843 times. The second most frequent word in the analysis is 'teachers,' occurring 2,157 times. The terms 'students' and 'training' are the third and fourth most frequent words, respectively, in the reputable journals.

The 200 word families in the COPES corpus were subjected further analysis to determine the grammatical class distribution. This study utilized version 2.0.5 of the TagAnt software program, developed by Anthony (2022) to identify keyword functions. Then, the inter-coder technique was employed to validate the data retrieved from TagAnt, comprising two experienced university lecturers with more than 10 years of teaching experience each. The two raters separately determined the grammatical class of all the 200 target words. The results of this analysis are displayed in Figure 1.

Figure 1

Grammatical Class Distribution of the Most Frequent Words Used in Physical Education and Sport Science Articles



Upon examining the 200 most frequent words in the research articles, it was found that the highest number constituted nouns, at 128 out of 200 words (64.0 percent), while verbs ranked second, with 41 words out of 200 (20.5 percent), followed by adjectives, with 28 words out of 200 (14.0 percent). The outcomes of this analysis also showed that the number of adverbs was relatively few, with only 3 words out of 200 (1.5 percent), as presented in Figure 1.

Previously, some studies found that the most frequent keywords were ranked as nouns, adjectives, and verbs. Likewise, the findings of the current study revealed that the majority of keywords were nouns, which is consistent with many previous studies (e.g., Sukman & Tangkiengsirisin, 2024; Trinant & Kijpoonphol, 2021). In general, nouns represent people, things, names, places, ideas and phenomena that are the key to communication. As a result, of the content words, nouns are the most prominent words. Of note, Sinar et al. (2023) explored the lexical density of 13 theses published in 2021. The frequency of content words was identified, revealing that the most prominent content words were nouns. This finding also aligns with the Academic Keyword List developed by Paquot (2010), which consists of 355 nouns, 233 verbs, 180 adjectives, 87 adverbs, and 75 other words.

In the present study, verbs were ranked second among the content words. This result differs somewhat from the above previous studies (Sukman et al., 2022; Sukman & Tangkiengsirisin, 2024; Trinant & Kijpoonphol, 2021). Based on the findings of these previous studies, adjectives were the second most frequent content words found. In contrast, the outcomes of the current study are similar to some previous studies in that the most frequent content words were nouns, verbs, and adjectives, respectively (Grabowski, 2015; Tarigan & Stevani, 2021). Grabowski (2015) compiled a corpus of pharmaceutical articles to determine frequent keywords used within English pharmaceutical discourse. The results revealed that the keywords found in the corpus expressed recommendations to patients on how to use medicine. Hence, the keywords mostly consisted of action verbs regarding how to take medicines. Likewise, the scope of this study was to explore the frequent content words used in research articles published in physical education and sport science. Consequently, the subject matter relevant to these disciplines is likely to deal with action, education, exercise, and movement. Therefore, the verbs predominantly express actions, such as 'develop,' 'participate,' 'perceive,' or 'perform.' It can be understood that different areas of study have unique lexicons to some extent. The following extracts are derived from the COPES corpus and demonstrate how the word "perform" was used in various research papers in the field of physical education and sports science. As such, the present study supports the notion that different disciplines have unique language features, and so each should be examined systematically (Khany & Saeedi, 2017).

Figure 2

Examples of the Word 'Perform' Found in the COPES Corpus

exercise gave the students "1" point; a failure to	perform	the exercise gave them "0" entered in the protocol. T
the case of the debate of ideas), and finally to	perform	the action experiencing these solutions (Kirk, 2017; M
eir technique under different fatigue states, and	perform	the appropriate action at the appropriate time (skill).
the intention, that is, the conscious decision to	perform	the behavior (Ajzen and Madden, 1986; Scanlan et al.
During all sessions, subjects were stimulated to	perform	the concentric phase of all exercises, as fast as
s of passive rest, the athletes were instructed to	perform	the drill test "two-line drill wide mode" at
of HRmax. Students were encouraged to run or	perform	the drills almost as intensively as possible'. However,
ility functions and once enabled he was able to	perform	the editing operations alongside his fellow co-resear

Lexical Collocations Used in Physical Education and Sports Science Research Articles

Examining in more detail the most frequently encountered keywords in the research articles, the top 50 keywords were selected. These fifty words were treated as nodes to find common collocations using the 'collocate' function in the AntConc software program. Once the keywords from the COPES were determined, the collocation extraction process began, following the established criteria. These criteria included a 3-word span the right of the node (3R), a specified frequency of occurrence, and a check of the mutual information (MI) score. At this stage, there were 50 nodes in total that were used for further identifying their collocations. The findings are summarized in Table 3 below.

Table 3

Lexical Collocations Patterns in COPES according to Combination Types

No.	Combination Types	No. of Collocation Pairs	Percentage
1	Noun + Noun	89	49.7
2	Adjective + Noun	50	27.9
3	Noun + Verb	31	17.3
4	Verb + Noun	5	2.8
5	Adverb + Adjective	4	2.2
	Total	**179**	**100.0**

Based on the data shown in Table 3, keywords are more likely to be associated with noun and noun collocates. It is evident that the 89 most frequent collocations are noun + noun (49.7 percent). Adjective + noun combinations are the second highest combination type, with 50 pairs (27.9 percent). Moreover, noun + verb combinations account for 31 words (17.3 percent), while there are relatively few instances of verb + noun and adverb + adjective pairs. Table 4 below displays 20 examples of the most frequent nodes and their collocations in order.

Table 4

Sample of 20 Keywords Found in the Field of Physical Education and Sport Science and Notable Collocations

No.	Nodes	Collocates	No.	Nodes	Collocates	No.	Nodes	Collocates
1	physical	activity education literacy fitness educators exercise activities culture classes	8	learning	outcomes process intentions objectives opportunities environment processes goals needs tasks	15	findings	Suggest indicate highlight revealed
2	students	learn	9	activity	levels participation engagement	16	content	validity taught
3	training	sessions program intensity loads tasks volume effects correction	10	education	system programmes	17	performance	assessment
4	participants	provided	11	skills	development	18	movement	cultures skills learning experiences patterns activities
5	teaching	strategies practices methods	12	curriculum	documents content	19	knowledge	requirements
6	research	team question process ethics committee design methods suggests	13	fitness	testing tests levels	20	strength	training endurance exercises
7	study	aimed design conducted showed sought	14	body	ideals mass composition weight image			

No.	Nodes	Collocates	No.	Nodes	Collocates	No.	Nodes	Collocates
		highlights			fat size			
		examined			height			
		carried out			language			
		protocol						
		provides						

Table 4 displays the twenty most frequently used keywords and their two-word cluster collocations utilized in the physical education and sport science research articles. According to Table 4, 'physical' was the highest frequency keyword in almost all the research articles. The terms 'physical activity', 'physical education', 'physical literacy', 'physical fitness, and 'physical educators' all refer to the physical components of education and health, such as activities, education programs, fitness levels, health outcomes, and literacy. Although the term 'students' is the second most frequently found word in the COPES, the results reveal that there is only one collocation: learn. In this case, only the word 'learn' obtained an MI value of ≥ 3 (MI = 3.002). Additionally, the term 'training sessions,' training program,' training intensity,' training intensity' and 'training tasks' refer to various types of activity involved in physical education and sport science programs.

The outcomes of the current study in combination with a number of previous studies including those by Ackermann and Chen (2013), Nguyen and Coxhead (2023), and Lei and Liu (2018) highlight the considerable variation in word combination structures across genres and disciplines. Based on the findings of this study, combinations of noun + noun constitute the most frequent in COPES. The outcome of this study is similar to Sukman and Tangkiengsirisin (2024) in that the most frequent collocation found in their English business news study was the noun + noun combination. Likewise, Hong (2023) stated that noun collocations are more prevalent combinations in some academic contexts than others. To sum up, noun + noun collocations are the most prevalent sort of lexical collocations in academic written English (Biber, Johansson, Leech, Conrad, & Finegan, 1999), and these combinations are essential to the development of academic writing.

The second most frequent word combination found in the current study is adjective + noun combinations. Instances like 'pedagogical practices,' 'social interaction,' and 'critical health' exemplify this pattern. The findings of the current study differ from Mandić and Dankić (2020) and Pujiningtyas and Bram (2023). Mandić and Dankić (2020) found that adjective + noun combinations were the dominant type found in their nursing scientific articles corpus. In addition, Pujiningtyas and Bram's (2023) indicated that adjective + noun constituted the highest number of combination types across master's students' writing. The current study also differs dramatically from the studies of Ackermann and Chen (2013) and Lei and Liu (2018), which found that adjective + noun pairings were more common than noun + noun combinations.

In conclusion, lexical collocations vary across genres, registers, and contexts. Different disciplines apparently have their own prevalent combinations. In the present study, as with some previous studies, noun + noun were the most prominent combinations, while adjective + noun collocations were the second most frequently found. Accordingly, the instruction and the testing of collocation knowledge should focus on these most frequent collocation patterns. The following section discusses the pedagogical implications.

Pedagogical Implications

Collocation knowledge is one indicator for language learners' proficiency, particularly for their fluency. Thus, the teaching of collocations should not be understated. Based on the findings of this study, high frequency words should be incorporated into classroom teaching. To provide instruction on vocabulary learning, instructors should primarily focus on the most frequent words together with their word families. Because such high frequency words tend to be seen often across academic texts in physical education and sport science, students should understand those words in order to

facilitate their understanding of academic texts in their study field. When instructors design lessons, headwords along with their families should be incorporated. Taking the word 'participate' as an example, instructors should include the main forms of the word, namely participate (verb), participant (noun, person), and participation (noun). Instructors can provide contexts for the students and ask them to choose the word form that best describes the situation or context.

When the students understand the headwords' meanings, instructors should deliberately present the target words together with their collocations by drawing students' attention to the words that commonly associate with those headwords. Initially, explicit instruction through deliberate activities is considered to be a useful technique for teaching vocabulary and collocates, especially for novice students (Hulstijn, 2003; Webb & Nation, 2013). Traditional activities and techniques can be used in class, such as flashcards, picture books, gap-filling activities, or role-play. When students become familiar with the learning process, instructors can gear the lessons towards more indirect ways of teaching and learning.

Instructors should be careful to activate students' long-term memories. To do so, instructors should consider how many times students have to come across the target words and their collocations or experience repetition of the target words, which can enhance students' long-term memories. As mentioned in the literature review, numerous researchers have suggested that collocation knowledge can be best learned and memorized when students experience at least five spacing repetitions (Macis, Sonbul & Alharbi, 2021). Therefore, instructors should provide students with enough exposure to the target words.

With the help of teaching tools, for instance COCA (Davies, 2008), WordandPhrase (Gardner & Davies, 2014) and AntWordProfiler (Lawrence, 2015), instructors can draw students' attention to the target words and their collocates. By equipping students with useful learning tools, students will be able to learn by themselves. To this end, instructor can organize classes by gradually moving them from being teacher centered to being teacher facilitated. These small steps encourage students towards eventually becoming autonomous learners.

Limitations and Suggestions for Further Studies

Despite the careful selection of the research articles in the COPES corpus and the adherence to the set criteria, there are some limitations. The number of running tokens in this research was 737,901, which may be considered a small corpus size. As such, the study's findings might not cover all the keywords used in physical education and sports science subjects. Additionally, the data gathered in this study covered a five-year period, between 2019 and 2023. Future research should integrate more research articles and extend the publication period to refine the keywords and collocations used. Additionally, it is worth exploring frequent words and collocations across disciplines.

Conclusion

This study examined 737,901 running word tokens from 144 research articles published in reputable journals in the field of physical education and sports science. The aim was to identify frequently used words in these articles and to examine their collocations. First, the researchers built a discipline-specific corpus using the AntConc version. Then, the researchers identified the keywords' grammatical functions using the TagAnt software. Next, the word tokens were analyzed to identify words that appeared frequently in the academic articles. On examining the frequency of the keywords, this study's most common word type was a noun; verbs and adjectives were the second and third most frequent. Subsequently, 50 keywords were used as nodes to investigate their collocations. This study identified the main keywords and their collocations, as presented earlier.

Based on the findings, the outcome of the study can be used to facilitate language learning and teaching in the field of physical education and sports science subjects. For language learning, the results should be applied to help learners in order to increase their knowledge of collocation and to raise their awareness of efficient language use (Mandić & Dankić, 2020; Pujiningtyas & Bram, 2023). For language teaching, lexical collocation instruction is effective for improving the English structure skills of language learners (Hashemi et al., 2012; Khonamri et al., 2020; Liontas et al., 2023). In addition, it should be integrated into language teachers' professional development in terms of designing teaching materials. For instance, Suraprajit, (2022) adapted the AntConc software to identify students' errors in conditional sentences.

References

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL)–A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235-247. https://doi.org/10.1016/j.jeap.2013.08.002

Alasmary, A. A. (2022). Comparing lexical bundles across three advanced mathematical text types: A corpus-based genre-focused investigation. *SAGE Open, 12*(3), 1–16. https://doi.org/10.1177/21582440221113824

Anthony, L. (2022). TagAnt (Version 2.0.5) [Computer Software]. Tokyo: Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo: Waseda University. https://www.laurenceanthony.net/software

Baker, P. (2006). *Glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Benson, M., Benson, E. and Ilson, F. R. (2010). *The BBI dictionary of English word combinations* Amsterdam: John Benjamins Publishing Company.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, *18*(1), 54-74. https://doi.org/10.1177/1362168813505389

Celik, B. (2024). The effects of input flood and input enhancement techniques in teaching collocations to EFL learners. *Forum for Linguistic Studies*, *6*(1). https://doi.org/10.59400/fls.v6i1.2088

Chanasattru, S., & Tangkiengsirisin, S. (2016). Developing of a high frequency word list in social sciences. *Journal of Studies in the English Language*, *11*. https://so04.tci-thaijo.org/index.php/jsel/article/view/73309

Chang, L. (2023). A corpus-based mechanical engineering academic word list. *International Journal of TESOL Studies*, *5*(3), 126. https://doi.org/10.58304/ijts.20230310

Chiva-Bartoll, O., Salvador-García, C., Pérez-Samaniego, V., & Flórez-García, M. (2019). University service-learning in physical education and sport sciences: A systematic review. *Journal of Hospitality, Leisure, Sport & Tourism Education, 25*, 100195. https://doi.org/10.1016/j.jhlste.2019.100195

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

Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed.*), Handbook of vocabulary studies* (pp. 97–110). Oxfordshire: Routledge.

Coxhead, A. (2021). Vocabulary in English in tertiary contexts: Connecting research and learning. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 1–14.

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, *51*, 84-97. https://doi.org/10.1016/j.esp.2018.03.006

Coxhead, A., Rahmat, A., & Lu Yang, W. (2020). Vocabulary for academic purposes: What do engineering students need to know? *English for Specific Purposes, 58*, 65-77. https://doi.org/10.1016/j.esp.2019.12.001

Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes*, *33*, 100-118. https://doi.org/10.1016/j.jeap.2018.02.003

Dang, T. N. Y., & Webb, S. (2016). Making an essential word list for beginners. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and teaching* (pp. 153–167). Amsterdam: John Benjamins Publishing Company.

Davies, M. (2008). *The Corpus of contemporary American English (COCA)*. Retrieved from https://www.english-corpora.org/coca/.

Durrant, P. (2016). To what extent is the academic vocabulary list relevant to university student writing? *English for Specific Purposes*, *43*, 49-61. https://doi.org/10.1016/j.esp.2016.01.004

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, *30*(1), 58-72. http://doi.org/10.1016/j.esp.2010.05.002

Eguchi, M., & Kyle, K. (2023). L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. *Journal of Second Language Writing*, *60*. https://doi.org/10.1016/j.jslw.2023.100975

Ellis, N. C. & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science, 9* (3), 604-620. https://doi.org/10.1111/tops.12256

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning, 67*(S1), 155–179. https://doi.org/10.1111/lang.12225

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

Garner, J. (2022). The cross-sectional development of verb–noun collocations as constructions in L2 writing. *International Review of Applied Linguistics in Language Teaching*, *60*(3), 909-935.https://doi.org/10.1515/iral-2019-0169

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes, 51*, 1–17. https://doi.org/10.1016/j.esp.2018.02.002

Goulart, L. (2019). The use of collocations across proficiency levels: A literature review. *Brazilian English Language Teaching Journal, 10*(2), 1–15. https://doi.org/10.15448/2178-3640.2019.2.34129

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38*, 23–33.https://doi.org/10.1016/j.esp.2014.10.004

Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes, 35*, 105–115. https://doi.org/10.1016/j.jeap.2018.07.004

Hamed, D. (2021). Keywords and collocations in US presidential discourse since 1993: A corpus-assisted analysis. *Journal of Humanities and Applied Social Sciences, 3*(2), 137–158. https://doi.org/10.1108/JHASS-01-2020-0019

Hanks, E., Egbert, J., & Hashimoto, B. (2024). The contracts word list: Integral vocabulary for reading and writing English contracts. *English for Specific Purposes*, *75*, 37–48. https://doi.org/10.1016/j.esp.2024.03.002

Hashemi, M., Azizinezhad, M., & Dravishi, S. (2012). Collocation: A neglected aspect in teaching and learning EFL. *Procedia - Social and Behavioral Sciences, 31*, 522–525. https://doi.org/10.1016/j.sbspro.2011.12.097

Heeren, J., Speelman, D., & De Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism, 24*(10), 1458–1473. https://doi.org/10.1080/13670050.2019.1709411

Hong, L. A., (2023). A corpus-based analysis of frequently occurring noun collocations in geographical information system (GIS) research articles. *3L: Southeast Asian Journal of English Language Studies*, *29*(3), 308–325. https://doi.org/10.17576/3L-2023-2903-21

Hsu, W. (2018). The most frequent BNC/COCA mid- and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks. *English for Specific Purposes, 51*, 98–110. https://doi.org/10.1016/j.esp.2018.04.001

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.

Hyland, K. (2007). *Second language writing*. Cambridge: Cambridge University Press.

Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly, 41*(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jaafar, E. A. (2022). Collocation networks of selected words in academic writing: A corpus-based study. *Arab World English Journal, 13*(1), 240–255. https://doi.org/10.24093/awej/vol13no1.15

Khany, R., & Saeedi, M. (2017). Material development and English for academic purposes word lists: A reductionist approach. *English Language Teaching and Learning*, *9*(19), 53-72.

Khonamri, F., Ahmadi, F., Pavlikova, M., & Petrikovicova, L. (2020). The effect of awareness raising and explicit collocation instruction on writing fluency of EFL learners. *European Journal of Contemporary Education, 9*(4), 786-806–806. https://doi.org/10.13187/ejced.2020.4.786

Kongcharoen, P. (2018). Basic physical education and sport science English word list for physical education students. *REFLections*, *25*(2), 120–148. https://doi.org/10.61508/refl.v25i2.165396

Lawrence, J. (2015). *AntWordProfiler* [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antwordprofiler/

Lawrence, J. F., Knoph, R., McIlraith, A., Kulesz, P. A., & Francis, D. J. (2022). Reading comprehension and academic vocabulary: Exploring relations of item features and reading proficiency. *Reading Research Quarterly, 57*(2), 669–690. https://doi.org//10.1002/rrq.434

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, *23*(2), 216-243. https://doi.org/10.1075/ijcl.16135.lei

Liontas, J. I., Bangun, I. V., & Li, S. (2023). Lexical collocational instruction in EAP writing via COCA. *Teaching English with Technology, 23*(1), 80–106. https://doi.org/10.56297/BKAM1691/RQCC4475

Ma, H., & Qian, M. (2020). The creation and evaluation of a grammar pattern list for the most frequent academic verbs. *English for Specific Purposes*, *58*, 155–169. https://doi.org/10.1016/j.esp.2020.01.002

Macis, M., Sonbul, S., & Alharbi, R. (2021). The effect of spacing on incidental and deliberate learning of L2 collocations. *System*, *103*, 102608. https://doi.org/10.1016/j.system.2021.102608

Mandić, K., & Dankić, I. (2020). Collocations of high frequency words in nursing research articles and the Academic Collocation List: Similarities and differences. *Explorations in English Language and Linguistics*, *8*(1), 1-13. https://doi.org/10.2020/Explorations.v8n1.1

Martinez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*(3), 183–198. https://doi.org/10.1016/j.esp.2009.04.003

Masrai, A., & Milton, J. (2017). Recognition vocabulary knowledge and intelligence as predictors of academic achievement in EFL context. *TESOL International Journal, 12*(1), 128–142.

Men, H. (2018). *Vocabulary increase and collocation learning: A corpus-based cross-sectional study of Chinese learners of English*. New York, NY: Springer.

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nation, P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins Publishing Company.

Nation, P. (2017). How vocabulary is learned. *Indonesian JELT*, *12*(1), 1-14. https://doi.org/10.25170/ijelt.v12i1.1458

Nguyen, H., & Coxhead, A. (2023). Investigating academic collocations in science and engineering research writing. *Journal of English for Academic Purposes, 53*, 101045. https://doi.org/10.1016/j.jeap.2023.101045

Ojeda-Nahuelcura, R., Carter-Thuillier, B., López-Pastor, V., & Fuentes-Nieto, T. (2023). Impact of generic or transversal competences on the performance of specialists in physical education and sports sciences: A systematic review. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *32*. https://doi.org/10.1016/j.jhlste.2023.100418

Okkinga, M., van Gelderen, A. J. S., van Schooten, E., van Steensel, R., & Sleegers, P. J. C. (2023). Does vocabulary knowledge matter in the effectiveness of instructing reading strategies? Differential responses from adolescents with low academic achievement on growth in reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 36*(10), 2549–2575. https://doi.org/10.1007/s11145-022-10359-2

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

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. New York, NY: Continuum.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75–94. https://doi.org/10.1177/1362168813505384

Pinchbeck, G. G., Brown, D., Mclean, S., & Kramer, B. (2022). Validating word lists that represent learner knowledge in EFL contexts: The impact of the definition of word and the choice of source corpora. *System*, *106*. https://doi.org/10.1016/j.system.2022.102771

Pu, P., Chang, D. Y. S., & Wang, S. (2024). Incidental learning of collocations through different multimodal input: The role of learners' initial L2 proficiency. *System, 125*. https://doi.org/10.1016/j.system.2024.103416

Pujiningtyas, M. D. A., & , & Bram, B. (2023). Lexical collocation analysis in master's student reflective writings. *Journal of Education and Teaching, 4*(3), 283–293. https://doi.org/10.51454/jet.v4i3.264

Richards, J. C. & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practices*. Cambridge: Cambridge University Press.

Saeedi, M., Khany, R., & Tazik, K. (2023). Research themes and sub-themes in academic wordlist studies between 2000 and 2020: A systematic review. *Journal of Research in Applied Linguistics, 14*(1), 95-111. https://doi.org/10.22055/RALS.2023.18070

Sinar, T. S., Zein, T. T., Ganie, R., Syarfina, T., Mahriyuni, M., Yusuf, M., & Rangkuti, R. (2023). Content words and readability in students' thesis findings. *Journal of Curriculum and Teaching*, *12*(6), 347–355. https://doi.org/10.5430/jct.v12n6p347

Sukman, K., & Tangkiengsirisin, S. (2024). Lexical collocations of keywords used in business news: A corpus-based study. *European Journal of English Language Teaching, 9*(1), 25017136. https://doi.org/10.46827/ejel.v9i1.5202

Sukman, K., Triwatwaranon, W., Munkongdee, T., & Chumnumnawin, N. (2022). A corpus-based study of lexical collocations of keywords found in online business news articles. *European Journal of English Language Teaching, 7*(3), 25017136. https://doi.org/10.46827/ejel.v7i3.4275

Sun, W., & Park, E. (2023). EFL learners' collocation acquisition and learning in corpus-based instruction: A systematic review. *Sustainability, 15*(17). https://doi.org/10.3390/su151713242

Suraprajit, P. (2022). Use, errors, and self-perceptions of Thai EFL learners with conditional sentences. *English Language Teaching*, *15*(8), 23-33.

Szabo, C. Z., Stickler, U., & Adinolfi, L. (2021). Predicting the academic achievement of multilingual students of English through vocabulary testing. *International Journal of Bilingual Education and Bilingualism, 24*(10), 1531–1542. https://doi.org/10.1080/13670050.2020.1814196

Tarigan, K. E., & Stevani, M. (2021). The interpretation of content and function words in Chinese textbooks for Indonesian students. *International Journal of Translation and Interpretation Studies, 1*(1), 10–14. https://doi.org/10.32996/ijtis.2021.1.1.2

Therova, D. (2020). Review of academic word lists. *TESL-EJ*, *24*(1), 1-15.

Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes, 51*, 45–54. https://doi.org/10.1016/j.esp.2018.03.002

Trinant, K., & Kijpoonphol, W. (2021). Lexical collocations in a sample corpus of tourism research articles (SCTRA). *NKRAFA Journal of Humanities and Social Sciences*, *9*, 94–108. https://so04.tci-thaijo.org/index.php/KANNICHA/article/view/249524

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248–263.https://doi.org/10.1016/j.jeap.2013.07.001

Webb, S., & Nation, I. S. P. (2013). Learning vocabulary in another language: Methodological approaches. *TESOL Quarterly, 47*(4), 771-786. https://doi.org/10.1002/tesq.117

Webb, S., Newton, J., & Chang, A. (2013). The effects of repetition on collocation learning. *Language Learning, 63*(2), 91-120. https://doi.org/10.1111/lang.12045

West, M. (1965). *A general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman.

Yotimart, D. (2021). Academic vocabulary in sport tourism news: A corpus-based study. *Journal of Language & Linguistics Studies, 17*(3), 1527–1535. https://doi.org/10.52462/jlls.110

Yulfi, Y., Seli, S., & Ariska, R. (2019). An analysis of English lexical collocation found in English newspaper. *Linguistic, English Education and Art (LEEA) Journal, 2*(2), 100–115. https://doi.org/10.31539/leea.v2i2.349

Zhang, P. (2022). How does repetition affect vocabulary learning through listening to the teacher's explicit instruction? The moderating role of listening proficiency and preexisting vocabulary knowledge. *Language Teaching Research*, 13621688221140521.