



Article A Corpus-Based Word Classification Method for Detecting Difficulty Level of English Proficiency Tests

Liang-Ching Chen ^{1,2}, Kuei-Hu Chang ^{3,*}, Shu-Ching Yang ² and Shin-Chi Chen ¹

- ¹ Department of Foreign Languages, R.O.C. Military Academy, Kaohsiung 830, Taiwan
- ² Institute of Education, National Sun Yat-sen University, Kaohsiung 804, Taiwan
- ³ Department of Management Sciences, R.O.C. Military Academy, Kaohsiung 830, Taiwan
- Correspondence: evenken2002@yahoo.com.tw

Abstract: Many education systems globally adopt an English proficiency test (EPT) as an effective mechanism to evaluate English as a Foreign Language (EFL) speakers' comprehension levels. Similarly, Taiwan's military academy also developed the Military Online English Proficiency Test (MOEPT) to assess EFL cadets' English comprehension levels. However, the difficulty level of MOEPT has not been detected to help facilitate future updates of its test banks and improve EFL pedagogy and learning. Moreover, it is almost impossible to carry out any investigation effectively using previous corpus-based approaches. Hence, based on the lexical threshold theory, this research adopts a corpusbased approach to detect the difficulty level of MOEPT. The function word list and Taiwan College Entrance Examination Center (TCEEC) word list (which includes Common European Framework of Reference for Language (CEFR) A2 and B1 level word lists) are adopted as the word classification criteria to classify the lexical items. The results show that the difficulty level of MOEPT is mainly the English for General Purposes (EGP) type of CEFR A2 level (lexical coverage = 74.46%). The findings presented in this paper offer implications for the academy management or faculty to regulate the difficulty and contents of MOEPT in the future, to effectively develop suitable EFL curriculums and learning materials, and to conduct remedial teaching for cadets who cannot pass MOEPT. By doing so, it is expected the overall English comprehension level of EFL cadets is expected to improve.

Keywords: English proficiency test (EPT); lexical threshold; corpus-based approach; Military Online English Proficiency Test (MOEPT); Taiwan College Entrance Examination Center (TCEEC); Common European Framework of Reference for Language (CEFR)

1. Introduction

Many education systems globally have adopted the English Proficiency Test (EPT) as an effective mechanism to evaluate English as a Foreign Language (EFL) speakers' comprehension levels [1,2]. It is also used to determine the success or failure of pedagogies and English learning performances [3,4]. Hence, any exploration of EPT can help stimulate EFL practitioners and learners' further interest [5,6].

With the advancement of Information and Communications Technology (ICT), EPT has gradually developed towards digitalization. For example, internationally well-known English certification tests, such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS), are implemented online [7]. Because the Online English Proficiency Test (OEPT) naturally offers many benefits, including random selection of test items, easy-to-update-and-expand test banks, and the fact that the evaluation process of EFL learners' academic performance is more flexible and ubiquitous (e.g., [8,9]), many educational institutes have developed OEPT for facilitating EFL students' English learning performance (e.g., [10,11]). Since OEPT is used to test EFL students' English proficiency, in order to effectively evaluate their learning performance, the literature has brought forth various derived issues, including whether the test is objective, its reliability and validity, its difficulty tendency, and so on (e.g., [9,12]).



Citation: Chen, L.-C.; Chang, K.-H.; Yang, S.-C.; Chen, S.-C. A Corpus-Based Word Classification Method for Detecting Difficulty Level of English Proficiency Tests. *Appl. Sci.* 2023, *13*, 1699. https://doi.org/ 10.3390/app13031699

Academic Editor: Valentino Santucci

Received: 15 December 2022 Revised: 17 January 2023 Accepted: 26 January 2023 Published: 29 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). When it comes to tests, most people in education think of the item response theory (IRT), along with test response patterns and theories, test reliability and validity, test construction, item banking, computer adaptive test (CAT), and so on [13–15]. IRT embraces two basic concepts: (1) test-takers' performances in test items can be predicted or explained by a single (or a group of) factor(s), which are called "latent traits" or "abilities"; and (2) test-takers' performances and their latent traits can be represented by a continuous increasing mathematical function, which is called the item characteristic curve (ICC) [16]. IRT has achieved remarkable results in the development of test banks and CAT systems (e.g., [13,14,17]). Its advantages include the ability to effectively define the reliability and validity of test items, classify the difficulty of test items, and conduct an integrated analysis of students' ability and test questions [18].

IRT tools or CAT systems rely on statistical experts or in-service teachers who have the relevant statistical expertise (e.g., [13–15,17,18]), which may not be easily implemented by Teaching English for Speakers of Other Languages (TESOL) faculty. Moreover, establishing the CAT system of OEPT requires considerable resources and funds, which may not be accepted by most schools. Thus, seeking a low-cost and efficient approach to explore the composition of EPT or even OEPT and defining its lexical difficulty tendency would undoubtedly be a boon for English language learning and teaching.

Language ability is multidimensional and includes three different concepts: (1) standards describe language performance in specific contexts at different proficiency levels; (2) language competence includes lexical and grammatical competence; and (3) language skills comprise listening, reading, speaking, and writing [12]. EPT is used as the standard to assess whether EFL learners can understand English language knowledge (i.e., lexicon and grammar) when those pop up in real-world English contexts [19,20]. In addition, EPT is fundamentally composed of a bunch of lexical items (i.e., vocabularies).

From the perspective of corpus linguistics, the difficulty level and classification of vocabularies have been processed by many linguists. Furthermore, many word lists have been developed for corpus-based analysis to date. For example, West [21] compiled the general service list (GSL), Coxhead [22] compiled the academic word list (AWL), Browne et al. [23] compiled the new general service list (NGSL) as an expanded version of GSL, Chen et al. [24] compiled the function word list as a standard to optimize the word list results of the corpus software. Domestically, the Taiwan College Entrance Examination Center (TCEEC) also compiled a word list that includes two different levels: Common European Framework of Reference for Language (CEFR) A2 and B1 word lists, which serve as the lexical threshold for the English curriculum of K12 education in Taiwan [25].

Vocabulary is an essential element for composing meaningful linguistic patterns of English [11,15,26]. Thus, Laufer [27] noted that the lexical threshold refers to an EFL reader's vocabulary knowledge that is required for understanding a genuine text. Vocabulary knowledge is also considered as a good predictor of English proficiency. Based on a corpus-based approach, Nation [28] found 3000 high-frequency word families that covered 89–95% of lexical items and 5000 high-frequency word families covered 92–98% of lexical items in written texts. Laufer [27] explained that once EFL learners acquire 6000–8000 vocabularies, they will reach 98% lexical coverage—namely, they may understand 98% of the words in an English text.

Based on the prior studies, it can be inferred that high lexical coverage brings a better comprehension level in reading English contexts with the same for listening comprehension and speaking ability [20,26,29–31]. Lexical coverage determines EFL learners' comprehension level of English contexts [32], while EPT is used to test EFL learners' comprehension level. Thus, this research hypothesizes that the lexical difficulty tendency of OEPT can be explored by a corpus-based approach that conducts vocabulary classification and composition analysis as well as vocabulary coverage calculation.

In the present study, a Taiwan military academy has established the Military Online English Proficiency Test (MOEPT) for evaluating EFL cadets' English comprehension levels. The reason that English acquisition is also emphasized by the military is because in cadets' future careers, they need to operate many U.S.-made weapon systems [33]. Additionally, frequent military exchanges between Taiwan and the U.S. will rely on officers who have adequate English communication skills. Thus, in addition to using MOEPT to evaluate cadets, more importantly, it has also become a referential indicator for English pedagogical goals.

In the development process of this MOEPT system, the academy hopes MOEPT can be a test system with moderate difficulty level and that meets the needs of English for military purposes. Under the condition that there is not enough professional manpower and sufficient funds to build a CAT system of MOEPT, this study detects the difficulty level of MOEPT by a corpus-based word classification method. In addition, the function word list [24] and TCEEC word list [25] are adopted as standardized measurement tools to classify the lexical items. The results offer important implications for future updates of MOEPT and the developments of English pedagogy and learning strategy.

2. Literature Review

2.1. Lexical Threshold

Vocabulary is an essential element in English language communication and is widely considered to be one of the crucial indicators of successful reading comprehension [27,34]. In other words, when EFL speakers acquire extensive vocabulary knowledge, they will find it easier to perform communication-related tasks (i.e., listening, speaking, reading, writing, or even translating). Conversely, weak vocabulary knowledge will decrease English comprehension [20,26,29–31].

Nation [35] stated the following. (1) Vocabulary coverage indicates the proportion of running words that readers are able to understand in English contexts. Using vocabulary coverage to define the percentage of comprehensive vocabularies in texts is based on the theory that there exists a threshold of language knowledge that draws a line to segregate those with or without sufficient lexical knowledge to achieve reading comprehension. Identifying lexical coverage is a prerequisite to understanding the text because different text genres have different customized vocabularies in composition (i.e., high-frequency words used in certain domains); (2) vocabulary size helps with the necessity of understanding diverse texts and the degree of requirement. Word families, in this case, are the critical consideration that must be taken into account when identifying the vocabulary size. Hence, coverage and size indicate whether EFL readers possess certain English capabilities to handle daily reading and communication; (3) Vocabulary level identification is based on different criteria, such as frequency, range, and dispersion, to categorize tokens in textual data. After confirming the vocabulary level of certain texts, the mechanisms, such as wordlist creation, genre analysis, and customary linguistic usage identification, will be triggered.

As Hsu [36] mentioned, the lexical threshold explores the connection between vocabulary and handling of communication-related tasks, analyzing vocabulary from three aspects: vocabulary coverage, size, and level. Hsu [36] compiled corpora of business textbooks (embracing 7,200,000 running words) and business research papers (embracing 7,620,000 running words) as the basis of linguistic analysis. The study subsequently used the RANGE program [37] and British National Corpus (BNC) to calculate the proportions of lexical coverage and to define the corresponding difficulty levels of the words. The results show that if learners possess 5000–8000 word-level comprehension capability in BNC, then they will be able to handle 98% language coverage to facilitate their understanding in learning English business textbooks and research papers.

Durrant [38] used corpus-based approaches to calculate lexical coverage of the academic vocabulary list (AVL) [39] in the British academic written English (BAWE) corpus. BAWE was considered an outstanding corpus of student writing, created and collected by four British colleges. Lexical coverage calculation is based on Durrant's approach [38], which combines many variations (e.g., level, discipline, text type, and item) to process the corpus in detail. The results showed that there were core lexical items that were frequently used by 90% of the disciplines, indicating that once the students can acquire core lexical items, they would be able to achieve greater success in their academic writing.

To summarize, apart from the cognition of semantic rules, vocabulary is undoubtedly the most basic and important cognition unit for EFL and native speakers [20,26,27,29–31,34,35]. Hence, the lexical threshold illustrates that a certain degree of English vocabulary comprehension is required to reach the cognitive needs of English in specific domains and certain education levels. Lexical threshold identification can be calculated by the proportions of word types and running words, and it can be based on certain criteria to identify lexical coverage, size, and difficulty level. Once lexical coverage, size, and difficulty level are successfully verified, the results can be adopted to facilitate pedagogical applications, and language acquisition (e.g., [35,36,38]).

2.2. TCEEC Word List

TCEEC word list is an important referential basis for the college entrance exam of the English subject and the English curriculum developments of junior and senior high schools in Taiwan. It was released in 2002 and embraced junior high school level (JHSL) and senior high school level (SHSL) word lists, which respectively represent CEFR A2 and B1 difficulty levels [25]. The word list contains 6480 American-style words and mainly includes words' lexemes without their word family. According to *Cobuild English Dictionary*'s statistics, the TCEEC word list includes the 2000 most primary words that occur at over 75% frequency in Anglophone contexts.

The word list is divided into function words and content words, and each group is separated into a different part of speech (POS). This can effectively help understand English words' linguistic behaviors in grammar and pronunciation better [25]. The TCEEC word list took two years, and its development was based on more than 35 kinds of foreign and domestic English books and the most powerful and useful English word lists. A word that can be included in the word list had to meet the following four criteria: (1) needs to be used with high frequency; (2) needs to occur in Taiwanese or American culture; (3) can be inferred to expand its word family or lemma; and (4) can reflect the life experience of Taiwan's K12 students [25]. The goal to develop the TCEEC word list is to bring benefits to the entrance exams, English curriculums, and editing of English textbooks [25].

3. Methodology

3.1. Set Word Classification Criteria

3.1.1. The Function Word List

The advantage of Chen et al.'s [24] function word list is that it embraces the most frequent function words in any corpus data, but there are only 228 word types of function words in their research. In addition to excluding the function words, the meaningless tokens and letters should also be removed for enhancing analytical efficiency. Hence, this study expands the function word list [24], which eventually covers 256 word types (see Figure 1). The revised function word list is adopted for measuring the target corpus.

3.1.2. CEFR A2 Level Word List

The compilation of the CEFR A2 level word list is based on the JHSL word list declared by TCEEC in 2002 [25]. The JHSL word list serves as a lexical threshold that has been utilized in Taiwan primary and junior high schools for defining the difficulty, progress, and scope in English curricula. Given that the JHSL word list in Taiwan's English education system is generally considered as a CEFR A2 level word list to define a Taiwanese's English proficiency level, the JHSL word list is thus set as the CEFR A2 criterion in this study.

Approximately 2000 word types exist on the original JHSL word list, but the word list covers only the words' lexemes (i.e., the base form of verbs and the singular nouns). To accurately determine the difficulty level of a vocabulary, the researcher expanded the words' word family (e.g., *accept, accepts, accepting, accepted*) and made turned the JHSL word list into a CEFR A2 level word list with 3811 word types (see Figure 2). By doing so, *actor*

and *actors*, for example, can be both considered at the CEFR A2 level, rather than just *actor* being defined at the CEFR A2 level, while *actors* becomes an off-list word.



Figure 1. Word cloud of the function words.



Figure 2. Word cloud of the CEFR A2 level word list.

3.1.3. CEFR B1 Level Word List

The compilation of the CEFR B1 level word list is based on the SHSL word list also declared by TCEEC [25]. The SHSL word list serves as the lexical threshold that has been utilized in Taiwan senior high schools for defining the difficulty, progress, and scope in English curricula. As the SHSL word list in Taiwan's English education system is generally considered a CEFR B1 level word list to define a Taiwanese's English proficiency level, the SHSL word list is thus set as the CEFR B1 criterion in this study.

Approximately 4480 word types exist on the original SHSL word list, and just like the JHSL word list, SHSL covers only the words' lexemes. Thus, the reason to expand SHSL words' word family (e.g., *bully, bullies, bullying, bullied*) is the same as with the principle of the CEFR A2 level word list compilation. After the SHSL word list is expanded as the CEFR B1 level word list, it embraces 5981 word types (see Figure 3).



Figure 3. Word cloud of the CEFR B1 level word list.

3.2. Data Processing and Analysis

This study uses AntConc 3.5.9 [40], a popular corpus software widely used in many corpus-based studies, to conduct data processing and analysis. The procedure can be divided into three major steps: (1) remove the overlapping parts between word classification criteria; (2) explore the composition of the off-list words; and (3) calculate word types proportion and lexical coverage of each word classification criteria. Detailed descriptions are given as follows.

(1) Remove the overlapping parts between word classification criteria

After expanding and compiling the function word list [24] and CEFR A2 and B1 level word lists, the researcher discovered overlapping words between each word list. If the overlapping words cannot be effectively removed, then these words will be classified repeatedly and cause errors and inaccuracy in detecting the difficulty level of the target corpus. Before the initiation of the process, it was assumed that if an overlapping word exists in both two different difficulty level word lists, then the word should be classified into the lower (i.e., easier) difficulty level word list. For example, if "the" exists on the function word list and CEFR A2 level word list simultaneously, then "the" should belong to the function word list.

Under the above condition, the researcher input the CEFR A2 level word list into AntConc 3.5.9 and then input the function word list and set it as "a stoplist" to remove the function words that occurred in the CEFR A2 level word list. Through this process, the CEFR A2 level word list removed 118 overlapping words and decreased its word types to 3692. Similarly, the researcher input the CEFR B1 level word list into the corpus software and then input the word list that integrated the function words and CEFR A2 level word list and set it as "a stoplist" to remove the function words and CEFR A2 level word list emoved 116 overlapping words and decreased its word types to 5865. The purpose of this series of filtering processes is to precisely distribute the words to proper difficulty level word lists to avoid bias in the analytical results.

(2) Explore the composition of the off-list words

After the function words, CEFR A2 level words and CEFR B1 level words are excluded; the remaining words are classified as off-list words. The function words can be recognized as grammar words, and CEFR A2 and B1 level words can be recognized as English for general purpose (EGP) words (i.e., everyday English) with high frequent usage in Anglophone countries [25]. The off-list words should be fully investigated, because these words may cover EGP-oriented words with a higher difficulty level or words that are for military purposes. Because MOEPT is designed for testing cadets' English competency in military contexts, the researcher hypothesizes that the off-list words should cover many military-purpose words (e.g., terminologies, technical words, and acronyms).

In order to address the off-list words, the researcher followed prior studies (e.g., [33,41]) and the words' meanings and usages to conduct the vocabulary taxonomy qualitatively. Three raters (including the researcher) conducted classification tasks based on each off-list word's literal meanings and functions in the test items. After the words' classifications among the off-list words are confirmed, each classification's statistical data can also be calculated.

(3) Calculate word types proportion and lexical coverage of each word classification criterion

In order to detect the difficulty level, the calculation of the proportion of different word classification criteria have to be implemented from two aspects: word types proportion and lexical coverage. Word types proportion represents the lexical composition of the target corpus; in addition, lexical coverage, as Nation [35] mentioned, indicates the proportion of tokens (i.e., running words). If a certain word has a high lexical coverage in a context, a reader will have higher likelihood to encounter that word in the future. Accordingly, identifying lexical coverage of different word classification criteria will make us understand

the overall difficulty level of the target corpus. Furthermore, identifying lexical coverage is also a prerequisite to effectively analyze the text because different text genres have different customized vocabularies in composition. The calculation of word types proportion and lexical coverage can be carried out by Equation (1).

Definition 1 ([35]). If \mathcal{L}_i represents a word classification criterion's word types proportion or lexical coverage, then, \mathcal{T}_i is a word classification criterion's number of word types or tokens on the target corpus, and \mathcal{T}_{σ} is the total number of word types or tokens of the target corpus.

$$\mathcal{L}_i = \frac{\mathcal{T}_i}{\mathcal{T}_\sigma} \times 100\% \tag{1}$$

Taking the function word list as an example, the researcher input the target corpus, then input the function word list and set it as "specific words", and then re-generated the word list on AntConc 3.5.9, which shows the function words' number of word types and tokens that exist in the target corpus. Next, the researcher used Equation (1) to compute for obtaining the word types proportion and lexical coverage values.

4. Results

4.1. Overview of the MOEPT at Taiwan Military Academy

MOEPT is an online version of EPT developed by the Taiwan military academy located in Kaohsiung to evaluate cadets' English proficiency levels (see Figure 4). Its question types and forms are based on the English comprehension level (ECL) test issued by the U.S. Defense Language Institute English Language Center (DLIELC). The ECL test is an instrument used to test international participants' English language reading and listening proficiency in certain U.S.-sponsored military exercises. It is also adopted as criteria for recruiting EFL military personnel or determining their eligibility for commissioning, attending training courses, or being assigned certain jobs (https://www.dlielc.edu/testing/ ecl_test.php, accessed on 1 November 2022).



Figure 4. MOEPT operating interface on the Internet.

Because MOEPT has not yet reached the mature developed stage and the test bank has still continued to be expanded and updated, the reliability and validity of MOEPT have not been tested yet. Hence, in this study, it is expected that the evaluation and analysis of lexical difficulty tendency can effectively be implemented by the corpus-based approach. The composition of question types currently includes 1337 listening comprehension questions, 748 grammar/vocabulary questions, and 97 short reading comprehension passages, which are similar to the U.S. DLIELC-issued ECL test.

Textual data of MOEPT were adopted as the target corpus. Overall, MOEPT included 13 simulation tests and 2182 test items. The compiled target corpus included 5465 word types and 79,526 tokens, and its lexical diverse indicator type/token ratio (TTR) was 0.07.

4.2. Composition of the Off-List Words

When the target corpus removed word types and tokens of the function words, CEFR A2 level words, and CEFR B1 level words, there remained 734 word types and 1833 tokens that did not belong to the aforementioned three-word classification criteria and were defined as the off-list words. The word types of the off-list words account for 13.4% of word types of the target corpus, and the tokens of the off-list words account for 2.3% of tokens of the target corpus.

In order to explore the composition of the off-list words, based on prior studies [e.g., 33,41], the researcher implemented vocabulary taxonomy. Based on each off-list word's literal meanings and functions in the test items, three raters conducted classification tasks. Because the off-list words (N = 734) were not in large quantity, during the process of classification, the raters, through cross-check and discussion methods, distinguished the vocabulary into five major categories: (1) other EGP words (including nouns, verbs, adjectives, and adverbs), (2) Names, (3) Countries/Locations, (4) Military, and (5) Medical.

The definitions, functions, and example sentences were retrieved from MOEPT database (i.e., the target corpus) through the interface of Key Word in Context Index (KWIC) on AntConc 3.5.9 [40]. This information is described in (1) to (5) as follows.

 Other EGP words: When an EGP word belongs neither to CEFR A2 level nor to B1 level word lists, it is regarded as a high-intermediate- or advanced-level noun (1-1), verb (1-2), adjective (1-3), adverb (1-4), proper noun (1-5), and contemporary gadget (1-6).

1-1 How do you reduce carbon emissions? [retrieved from TEST-03].

1-2 Most researchers **attributed** the soaring price to the greater use of American corn for making biofuel. [retrieved from TEST-06].

1-3 His report is incomprehensible. [retrieved from TEST-08].

1-4 They met coincidently. [retrieved from TEST-11].

1-5 In an **El Niño** phenomenon, the surface water in the eastern Pacific Ocean, near the coast of Peru, gets unusually warm. [retrieved from TEST-1].

1-6 Does the iPhone 5 carry any guarantee? [retrieved from TEST-6].

(2) Names: The function of a name is to make the test questions have situational effects so that test-takers can understand the conditions and timing to properly use vocabularies (2-1, 2-2). Additionally, some test questions also adopted celebrities' biographies or short introductions as contents of short reading comprehension passages (2-3).

2-1 Why did you break up with John? [retrieved from TEST-6].

2-2 **James** doesn't live with his family. He lives alone in Taipei. [retrieved from TEST-10].

2-3 **Margaret Thatcher**, also known as the Iron Lady, lived from 1925 until 2013. [retrieved from TEST-5].

(3) Countries/Locations: The categories of countries/locations and names have a similar function. Countries/locations are used to increase the situational effects in test questions, so that test-takers can understand the geographical cultures (3-1), national characteristics (3-2), and usage of geographical locations (3-3).

3-1 Margaret is deep into Chinese calligraphy. [retrieved from TEST-7].

3-2 A lot of people in Africa are suffering from famine. [retrieved from TEST-10].

3-3 I do not live in **Kaohsiung** so I have to commute from **Pingtung** to **Kaohsiung** every day. It takes me about 40 min. [retrieved from TEST-2].

(4) Military: As mentioned earlier, in order to cultivate English communication abilities among cadets to enable them to handle global affairs during their future careers, military-domain English vocabulary is an important issue in English curricula in the Taiwan military academy. Hence, the functions of the military are included in test questions to enable cadets to acquire U.S. military English usages, such as military ranks and branches (1-1), units (1-1, 1-2), and weapon systems nomenclatures (1-3).

4-1 Caption Murphy is in charge of the infantry company. [retrieved from TEST-9].4-2 Who is your new platoon leader? [retrieved from TEST-3].

4-3 Which do you think is one of the most widely used anti-tank guided missiles? **Javelin**. [retrieved from TEST-9].

(5) Medical: Proper medical nouns are included in MOEPT to make test questions resemble real-world usages (5-1, 5-2) and to increase the diversity of MOEPT questions and the degree of difficulty by embedding medical professional information into short reading comprehension passages (5-3).

5-1 Nancy works in an ICU in a nearby hospital. [retrieved from TEST-12].

5-2 Those who fall victim to **Alzheimer**'s disease often cannot find their way home when they go out. [retrieved from TEST-9].

5-3 **Ebola** is a viral disease of which the initial symptoms can include a sudden fever, intense fatigue, muscle pain, and a sore throat, according to the World Health Organization (WHO). [retrieved from TEST-7].

- 4.3. Word Types Proportion and Lexical Coverage of Each Word Classification Criteria
- (1) The function words

After limiting the word list range by setting the function words as the "specific word" on AntConc 3.5.9 (see Figure 5), it generated 237 word types and 49,190 tokens of the function words that exist on the target corpus (see Figure 6). After calculating their proportions on the target corpus, it was discovered that the word types of function words account for approximately 4% of word types of the target corpus, while their lexical coverage is 62%, indicating the function words are indispensable elements to form meaningful sentences, and the more function words that are extracted, the more simple sentences there are on the target corpus (e.g., [24,42]).

| ategory | Word List Preferences | | | | | | | | |
|--|---|--|--|--|--|--|--|--|--|
| oncordance oncordance Plot lusters/N-Grams | Display Options Rank Frequency Word Lemma Word Form(s) | | | | | | | | |
| ollocates | Other Options | | | | | | | | |
| Vord List | Treat all data as lowercase | | | | | | | | |
| eyword List | Treat case in sort | | | | | | | | |
| | Lemma List | | | | | | | | |
| | Loaded Load | | | | | | | | |
| | Target Corpus | | | | | | | | |
| | O Use raw file(s) ○ Use word list | | | | | | | | |
| | Load Word List | | | | | | | | |
| | | | | | | | | | |
| | Word List Ranger | | | | | | | | |
| | Use all words Use specific words below Use a stoplist below | | | | | | | | |
| | Add Word Add | | | | | | | | |
| | Add Words From File Function wordlist.txt Open | | | | | | | | |
| | your Clear | | | | | | | | |
| | yours | | | | | | | | |
| | 7 | | | | | | | | |
| | | | | | | | | | |

Figure 5. Set the function word list as "specific words".

| ile Global Settings | Tool Prefe | rences He | elp | | | | | |
|------------------------------------|------------|---------------------|------------|--------------------------|--------------|-------------------|----------|----------|
| Corpus Files | Conce | ordance_Co | pocordan | ce Plot File View Cluste | rs/N-Grams C | ollocates Word Li | st Keywo | ord List |
| 0001-Formal.txt | Word | Types: 23 | ch Hits: 0 | | | | | |
| 0002-Formal.txt | Rank | Ereq | Word_ | | | | | Lemr |
| 0002-Hidden.txt | 1 | 5069 | a | | | | | |
| 0003-Formal.txt | 2 | 3066 | the | | | | | |
| 0003-Hidden.txt | 3 | 2746 | b | | | | | |
| 0004-Formal.txt | 4 | 2735 | d | | | | | |
| 0005-Formal.txt | 5 | 2710 | C | 1 | | | | |
| 0005-Hidden.txt | 6 | 2101 | to | 1 | | | | |
| 0006-Formal.txt | 7 | 1721 | : | | | | | |
| 0007-Formal txt | 6 | 1721 | 1 | 1 | | | | |
| 0007-Hidden.txt 0008-Formal.txt | 8 | 1/11 | IS | | | | | |
| | 9 | 1191 | you | | | | | |
| 0008-Hidden.txt | 10 | 976 | it | | | | | |
| J009-Formal.txt | 11 | 971 | of | | | | | |
| 0010-Formal.txt | 12 | 861 | in | | | | | |
| 0011-Formal.txt | 13 | 784 | he | | | | | |
| 0012-Formal.txt | 14 | 656 | s | | | | | |
| 0013-Formal.txt | 15 | 644 | are | | | | | |
| | 16 | 582 | m | | | | | |
| | 17 | 555 | + | | | | | |
| | 17 | 555 | | | | | | |
| | 4 | P 4 P | 4 | | | | Þ | 4 |
| | Search | Term 🔽 🛛 | Words [|) Case 🗌 Regex | | Hit Location | | |
| | | | | | Advanced | Search Only | 0 | |
| | | tart | Ston | Sort | | Lemma List | Loaded | |
| Total No. | | tart | stop | Jon | | Word List | Loaded | |
| 22 Files Processed | Sort b | y Inver | t Order | | | | | |
| neo riocesseu | Carth | . Free | | 7 | | | | |

Figure 6. The function words that exist on the target corpus.

(2) CEFR A2 level words

After limiting the word list range by setting CEFR A2 level words as the "specific words" on AntConc 3.5.9 (see Figure 7), it generated 2557 word types and 22,589 tokens of CEFR A2 level words that existed on the target corpus (see Figure 8). After calculating their proportions on the target corpus, it was discovered that the word types of CEFR A2 level words account for approximately 47% of word types of the target corpus, while their lexical coverage is 28%.

| ategory | Word List Preferences | | | | | | |
|---|---|--|--|--|--|--|--|
| Concordance Concordance Plot Clusters/N-Grams | Display Options Rank @ Frequency @ Word @ Lemma Word Form(s) | | | | | | |
| Collocates | Other Options | | | | | | |
| Vord List | Treat all data as lowercase | | | | | | |
| eyword List | Treat case in sort | | | | | | |
| | Lemma List | | | | | | |
| | Loaded Load | | | | | | |
| | Target Corpus | | | | | | |
| | O Use raw file(s) ○ Use word list | | | | | | |
| | Loaded Load Word List | | | | | | |
| | West 11 to Develop | | | | | | |
| | O Use all words | | | | | | |
| | Add Word | | | | | | |
| | | | | | | | |
| | Add Words From File Refined JHSL wordlist (JH Open | | | | | | |
| | youth Clear | | | | | | |
| | yummy | | | | | | |
| | zero | | | | | | |
| | 200 | | | | | | |

Figure 7. Set CEFR A2 level word list as "specific words".

| File Global Settings | Tool Prefe | rences H | lelp | | | | | | |
|------------------------------------|------------|--|-----------------|-------------------|----------|---------------------|--------------|--|--|
| Corpus Files | Conce | ordance 6 | toncordance Plu | File View Cluster | -Grams C | ollocates Word List | Keyword List | | |
| 0001-Formal.txt | Word | Word Types: 2557 Word Tokens: 22589 Search Hits: 0 | | | | | | | |
| 0002-Formal.txt | Rank | Freq | Word | | -1 | | Lemn | | |
| 0002-Hidden.txt | 1 | 386 | yes | | | | | | |
| 0003-Formal.txt | 2 | 245 | like | | | | | | |
| 0003-Hidden.txt | 3 | 196 | qo | | | | | | |
| 0004-Formal.txt | 4 | 188 | people | | | | | | |
| 0005-Formal.txt | 5 | 187 | man | | | | | | |
| 0005-Hidden.txt | 6 | 182 | time | | | | | | |
| 0006-Formal.txt | 7 | 174 | aoina | | | | | | |
| 0007-Formal.txt | 0 | 172 | going | | | | | | |
| 0007-Hidden.txt | 0 | 1/5 | need | | | | | | |
| 0008-Formal.txt | 9 | 163 | good | | | | | | |
| 0008-Hidden.txt | 10 | 160 | woman | | | | | | |
| 0009-Hidden txt | 11 | 149 | take | | | | | | |
| 0010-Formal.txt 0011-Formal.txt | 12 | 119 | think | | | | | | |
| | 13 | 115 | want | | | | | | |
| 0012-Formal.txt | 14 | 115 | new | | | | | | |
| JU15-Formal.txt | 15 | 108 | one | | | | | | |
| | 16 | 105 | aive | | | | | | |
| | 17 | 103 | let | | | | | | |
| | | | | | | | | | |
| | 4 | P 4 1 | • • | ~- | | | P. 4 | | |
| | Search | i Term 🛃 | Words Case | e 📋 Regex | | Hit Location | 1.1 | | |
| | | | | | Advanced | Search Only 0 | | | |
| Total No | S | tart | Stop | Sort | | Lemma List 🗌 Lo | aded | | |
| 22 | | | | | | Word List | aded | | |
| Files Processed | Sort b | y 🗌 Inve | ert Order | | | | | | |
| | Sort b | y Freq | ~ | | | | | | |

Figure 8. CEFR A2 level words that exist on the target corpus.

(3) CEFR B1 level words

After limiting the word list range by setting CEFR B1 level words as the "specific word" on AntConc 3.5.9 (see Figure 9), it generated 1937 word types and 5914 tokens of CEFR B1 level words that existed on the target corpus (see Figure 10). After calculating their proportions on the target corpus, it was discovered that the word types of CEFR B1 level words account for approximately 35% of word types of the target corpus, while its lexical coverage is 7%.

| Tool Preferences | |
|--|---|
| Category | Word List Preferences |
| Concordance Concordance Plot Clusters/N-Grams Collocates Word List Keyword List | Display Options Rank Ø Frequency Ø Word Ø Lemma Word Form(s) Other Options Treat all data as lowercase Treat case in sort |
| | Lemma List Loaded Load Target Corpus Use raw file(s) Use word list Loaded Load Word List Word List Range Use all words Add Word Add Add Words From File Refined SHSL wordlist (SI Open Zipper zippers zone zoom |

Figure 9. Set CEFR B1 level word list as "specific words".

| File Global Settings | Tool Prefe | rences | Help | | | | | | |
|-----------------------|---|-----------|----------------------------|---------------------|---------------------|--------------|--|--|--|
| Corpus Files | - Conce | ordance - | Concordance Plot file View | -Closters/N-Grams C | ollocates Word List | Keyword List | | | |
| 0001-Formal.txt | Word Types: 1937 Word Tokens: 5914 Search Hits: 0 | | | | | | | | |
| 0002-Formal txt | Rank | Freq | Word | == | | Lemn | | | |
| 0002-Hidden.txt | 1 | 62 | person | 1 | | | | | |
| 0003-Formal.txt | 2 | 62 | better | 1 | | | | | |
| 0003-Hidden.txt | 3 | 48 | according | 1 | | | | | |
| 0004-Formal.txt | 1 | 11 | passago | 1 | | | | | |
| 0004-Hidden.txt | 4 | 41 | passage | 1 | | | | | |
| 0005-Hidden.txt | 5 | 34 | tv | 1 | | | | | |
| 0006-Formal.txt | 6 | 33 | plane | 1 | | | | | |
| 0006-Hidden.txt | 7 | 30 | global | 1 | | | | | |
| 0007-Formal.txt | 8 | 29 | recently | 1 | | | | | |
| 000/-Hidden.txt | 9 | 29 | best | 1 | | | | | |
| 0008-Formal.txt | 10 | 26 | earthquake | | | | | | |
| 0009-Formal.txt | 11 | 26 | iack | 1 | | | | | |
| 0009-Hidden.txt | 10 | 20 | Jack | 1 | | | | | |
| 0010-Formal.txt | 12 | 25 | sergeant | | | | | | |
| 0011-Formal.txt | 13 | 24 | military | | | | | | |
| 0012-Formal.txt | 14 | 23 | products | 1 | | | | | |
| oor of ormal ext | 15 | 22 | parents | | | | | | |
| | 16 | 21 | attitude | | | | | | |
| | 17 | 20 | nhone | 1 | | | | | |
| | | | in | | | | | | |
| | 4 | Þ 4 | P 4 | | | P 4 | | | |
| | Search | n Term 💽 | Words Case Rege | ex | Hit Location | | | | |
| | | | | Advanced | Search Only | 0 | | | |
| | | | Cine Cent | | Lemma List 🔲 🛛 | .oaded | | | |
| Total No. | 5 | tan | stop | | Word List | .oaded | | | |
| 22 Files Processed | Sort b | y 🗆 Inv | ert Order | | | | | | |
| riles Processed | Sort b | y Freq | ~ | | | | | | |
| | | f | | | | | | | |

Figure 10. CEFR B1 level words that exist on the target corpus.

(4) Off-list words

After finishing the vocabulary taxonomy, the composition of the off-list words included the following classifications: other EGP words embraced 405 word types (proportion = 7.41%) and 760 tokens (lexical coverage = 1%), Name embraced 186 word types (proportion = 3.4%) and 683 tokens (lexical coverage = 0.9%), Countries/Locations embraced 109 word types (proportion = 1.99%) and 308 tokens (lexical coverage = 0.4%), Military embraced 27 word types (proportion = 0.49%) and 70 tokens (lexical coverage = 0.09%), and Medical embraced 7 word types (proportion = 0.13%) and 12 tokens (lexical coverage = 0.02%) (see Table 1).

Table 1. Lexical composition of the target corpus.

| Word Classification Criteria | Word Types | Word Types Proportions | Tokens | Lexical Coverage |
|------------------------------|---------------|---------------------------|--------|---------------------|
| The function words | 237 | 4.34% | 49,190 | 61.85% |
| CEFR A2 level words | 2557 | 46.79% | 22,589 | 28.4% |
| CEFR B1 level words | 1937 | 35.44% | 5914 | 7.44% |
| Other EGP words | 405 | 7.41% | 760 | 1% |
| Name | 186 | 3.4% | 683 | 0.9% |
| Country/Location | 109 | 1.99% | 308 | 0.4% |
| Military | 27 | 0.49% | 70 | 0.09% |
| Medical | 7 | 0.13% | 12 | 0.02% |
| Total | 5465 | 100% | 79,526 | 100% |

To summarize, after the composition of the target corpus was comprehensively investigated by the proposed method, the results indicated that from the perspective of word types, CEFR A2 level words account for the largest proportion (46.79%) on the target corpus, while Medical had the lowest proportion (0.13%); in addition, from the perspective of lexical coverage, the function words has the largest lexical coverage (61.85%), while, Medical also had the lowest lexical coverage (0.02%) (see Table 1).

4.4. Overall Difficulty Level of the Target Corpus

Because the function words are grammatical words that are used for composing meaningful sentences, such words usually meaningless but have larger lexical coverage, which may misjudge the over difficulty tendency of the target corpus. Thus, to effectively detect overall difficulty level of the target corpus, the researcher firstly excludes the function words from the target corpus, then, analyzes the composition of the target corpus from two perspectives, word types and tokens for deep interpretation.

From the perspective of word types, the word types' difficulty tendency of the target corpus mainly tends toward the CEFR A2 level and CEFR B1 level, because the above two word lists account for about 85.96% of word types of the target corpus (see Table 2). This indicates that the settings of MOEPT still mainly belong to the EGP-type test bank.

| Word Classification Criteria | Word Types | Word Types Proportions | Tokens | Lexical Coverage |
|------------------------------|---------------|---------------------------|--------|---------------------|
| CEFR A2 level words | 2557 | 48.91% | 22,589 | 74.46% |
| CEFR B1 level words | 1937 | 37.05% | 5914 | 19.49% |
| Other EGP words | 405 | 7.75% | 760 | 2.51% |
| Name | 186 | 3.56% | 683 | 2.25% |
| Country/Location | 109 | 2.08% | 308 | 1.02% |
| Military | 27 | 0.52% | 70 | 0.23% |
| Medical | 7 | 0.13% | 12 | 0.04% |
| Total | 5228 | 100% | 30,336 | 100% |

Table 2. Lexical composition of the target corpus (removing the function words).

After comparing the lexical coverage of the word classification criteria, the difficulty tendency of the target corpus mainly tends toward the CEFR A2 level (74.46%), indicating that many test items are composed of CEFR A2 level words; in addition, CEFR B1 level words have the second large lexical coverage (19.49%), while the classifications of the off-list words are infrequently occurred (6.04%) on the target corpus (see Table 2).

Lexical coverage reflects the total frequency of word types in the target corpus. In addition, the frequency also affects the probability of the appearing vocabulary [11,20,26–29,35]. In this case, when test-takers take MOEPT, they may encounter CEFR A2 level words with a higher likelihood. After a series of filtering, computing, and analyzing procedures, this study anatomizes the composition of MOEPT in detail and successfully detects the lexical difficulty level by means of the word classification criteria, thus defining the difficulty level of MOEPT is mainly EGP-type of the CEFR A2 level.

5. Discussion

This study used the corpus-based word classification method to detect the difficulty level of MOEPT at a Taiwanese military academy. The analytical results indicated that MOEPT is mainly EGP-type at the CEFR A2 level. This is completely inconsistent with what the academy expected when the MOEPT system was first built. The military academy belongs to the academic system of universities. Under the policy of the Ministry of Education, undergraduate students in Taiwan are expected to have CEFR B2 level English proficiency, because TCEEC has already set JHSL and SHSL word lists as the lexical thresholds for English curricula in junior and senior high school, respectively. In other words, students who complete the senior high school (K12) are expected to have acquired the TCEEC word list (N = 6480) and be able to learn more academic-oriented or CEFR B2 level words in their future higher education [43]. However, the researcher discovered current MOEPT included only 405 word types (760 tokens) that may be categorized as higher CEFR B2 level words (i.e., other EGP words). Moreover, these words accounted for only 7.75% of word types and 2.51% tokens of the target corpus (see Table 2). This information serves as an important warning call in the future update and development of the MOEPT test bank. The lexical difficulty tendency should not remain at the CEFR A2 level. On the contrary, it should be

improved to the CEFR B2 level. Once the lexical difficulty level is enhanced, the difficulty of MOEPT will also be enhanced, as prior research studies have reported that vocabulary knowledge is considered as a good predictor of EPT [11,15,20,26,29–31].

From the perspective of English for a specific purpose (ESP), the growing emphasis is on including ESP in curricula to improve the English performance of each professional field in EFL countries (e.g., [36,44]). Nguyen [45] mentioned that ESP is a pedagogical approach that provides EFL learners with language skills for specific professional and academic purposes. The cadets will face military exchanges between Taiwan and the U.S. after graduating from the academy. Their English proficiency and communication capabilities will be involved in English for military purposes. As for other ESP cases, military English embraces uncommon terminologies, slang, and lexical bundle usages [33]. The academy's English curricula must also play an important role in teaching military English. However, the results show that there were even fewer military-oriented words than other ESP categories. There were only 27 word types (70 tokens) that can be categorized as military-purpose words. Moreover, these words accounted for only 0.52% of word types and 0.23% of tokens of the target corpus (see Table 2). In order to enhance MOEPT's practicality in future real-word military English contexts, the researcher suggests that military-oriented words should be embedded into MOEPT to increase military-purpose dialogues and military related articles for making cadets better understand real-word situations and make MOEPT more aligned with cadets' future careers.

In terms of cadets' academic performances and English curriculum developments, the academy should not consider passing MOEPT as the major pedagogical goal; otherwise, it will fall into the predicament that the test dominates the English curriculum. Moreover, passing MOEPT will not substantially improve cadets' English competency as the overall lexical tendency is toward the CEFR A2 level, while the CEFR A2 level should be targeted after finishing the English curriculum in junior high school for Taiwan's education system. For cadets who cannot pass MOEPT (i.e., their score is less than 60 points), the academy is suggested to relocate them to more basic English classes for remedial teaching, because they may have a problem with the basic construction of vocabulary knowledge. Hence, there is a necessity for implementing English competence-based class grouping (e.g., [46,47]).

6. Conclusions

Prior research studies have supported that a significantly positive correlation exists between vocabulary knowledge and EPT performances [11,15,26,48,49]. EFL learners' vocabulary size is crucial for comprehending English contexts—that is, better vocabulary knowledge brings out higher lexical coverage and a better understanding of English contexts [20,26,29–32]. Based on the lexical threshold theory, this study adopted the corpusbased approach to explore the lexical difficulty tendency of MOEPT, as developed by a Taiwanese military academy. The findings presented have important implications for the future development and updating of MOEPT. Moreover, for English pedagogical implications, the findings are vital indicators for setting vocabulary goals, selecting materials and lexical tasks, designing lexical syllabi, or even monitoring cadets' vocabulary learning progress.

This paper has the following contributions. First, the word classification process has successfully separated the word types and tokens of the target corpus into function words, CEFR A2 level words, CEFR B1 level words, and off-list words in a machined-based way. Furthermore, off-list words have also been investigated to further be divided into sub-classifications, including other EGP words, Name, Country/Location, Military, and Medical by the qualitative vocabulary taxonomy that was conducted by the three raters, which made the composition of the target corpus be more clearly anatomized. Second, the corpus-based word classification method used for the detection of difficulty level can be replicated and used to explore any EPT corpus data (e.g., entrance exams, TOEFL, and IELTS). Third, the TCEEC word list, including CEFR A2 and B1 level word lists, is well-known within Taiwan's education system [43]. Thus, this difficulty level analysis will allow Taiwanese teachers and students to immediately understand the trend of EPT that

they want to analyze. To summarize, although MOEPT may not be a representative EPT sample, the proposed corpus-based word classification method is still capable of unveiling the mystery of the difficulty level of any EPT and can be utilized widely.

This paper still has some limitations. For example, unlike IRT, this corpus-based analysis did not take EFL learners' English competency into consideration or integrally analyze the correlation between cadets' English competency and MOEPT difficulty tendency. Moreover, the reliability and validity of MOEPT were not explored and discussed in this study. Future researchers can rely on the present study to develop different difficulty level word lists as measurement tools to conduct more extensive corpus analysis and to conduct in-depth evaluations of the reliability and validity of EPT by integrating information on linguistic aspects. Other factors may affect test analysis, including average text length, percentage of complex words, percentage of parts of speech, and readability indices, can also be discussed in the future (e.g., [50,51]). In addition, exploring the relationships between tests, pedagogical design and practice, learning strategy, the washback effect, and an external monitoring mechanism can help ensure that the curriculum does not deviate from the test preparation tendency (e.g., [52–56]).

Author Contributions: Conceptualization, L.-C.C. and K.-H.C.; methodology, L.-C.C.; software, L.-C.C., S.-C.Y. and S.-C.C.; validation, L.-C.C. and K.-H.C.; formal analysis, L.-C.C.; investigation, L.-C.C.; resources, L.-C.C. and K.-H.C.; data curation, L.-C.C., K.-H.C., S.-C.Y. and S.-C.C.; writing—original draft preparation, L.-C.C.; writing—review and editing, L.-C.C., K.-H.C. and S.-C.Y.; supervision, K.-H.C.; funding acquisition, K.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the National Science and Technology Council, Taiwan, for financially supporting this research under Contract No. MOST 110-2410-H-145-001 and MOST 111-2221-E-145-003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Brooks, M.D. What does it mean? EL-identified adolescents' interpretations of testing and course placement. *Tesol Q.* 2022, *56*, 1218–1241. [CrossRef]
- Chan, S.; Taylor, L. Comparing writing proficiency assessments used in professional medical registration: A methodology to inform policy and practice. Assess. Writ. 2020, 46, 100493. [CrossRef]
- 3. MacGregor, D.; Yen, S.J.; Yu, X. Using multistage testing to enhance measurement of an English language proficiency test. *Lang. Assess. Q.* **2022**, *19*, 54–75. [CrossRef]
- 4. Yuksel, D.; Soruc, A.; Altay, M.; Curle, S. A longitudinal study at an English medium instruction university in Turkey: The interplay between English language improvement and academic success. *Appl. Linguist. Rev.* **2021**, *5*, 387–402. [CrossRef]
- Culbertson, G.; Andersen, E.; Christiansen, M.H. Using utterance recall to assess second language proficiency. *Lang. Learn.* 2020, 70, 104–132. [CrossRef]
- Yeom, S.; Jun, H. Young Korean EFL learners' reading and test-taking strategies in a paper and a computer-based reading comprehension tests. *Lang. Assess. Q.* 2020, 17, 282–299. [CrossRef]
- Lestari, S.B. English language proficiency testing in Asia: A new paradigm bridging global and local contexts. *RELC J.* 2021, 53, 757–759. [CrossRef]
- Isbell, D.R.; Kremmel, B. Test review: Current options in at-home language proficiency tests for making high-stakes decisions. Lang. Test. 2020, 37, 600–619. [CrossRef]
- Ockey, G.J. Developments and challenges in the use of computer-based testing for assessing second language ability. *Mod. Lang. J.* 2009, 93, 836–847. [CrossRef]
- 10. Clark, T.; Endres, H. Computer-based diagnostic assessment of high school students' grammar skills with automated feedback–An international trial. *Assess. Educ.* 2021, *28*, 602–632. [CrossRef]
- 11. Fehr, C.N.; Davison, M.L.; Graves, M.F.; Sales, G.C.; Seipel, B.; Sekhran-Sharma, S. The effects of individualized, online vocabulary instruction on picture vocabulary scores: An efficacy study. *Comput. Assist. Lang. Learn.* **2012**, 25, 87–102. [CrossRef]

- 12. Min, S.C.; Aryadoust, V. A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Stud. Educ. Eval.* **2021**, *68*, 100963. [CrossRef]
- 13. Chang, H.H. Psychometrics behind computerized adaptive testing. Psychometrika 2015, 80, 1–20. [CrossRef] [PubMed]
- 14. Kaya, E.; O'Grady, S.; Kalender, I. IRT-based classification analysis of an English language reading proficiency subtest. *Lang. Test.* **2022**, *39*, 4. [CrossRef]
- 15. Mizumoto, A.; Sasao, Y.; Webb, S.A. Developing and evaluating a computerized adaptive testing version of the word part levels test. *Lang. Test.* **2019**, *36*, 101–123. [CrossRef]
- 16. Guilera, G.; Gomez, J. Item response theory test equating in health sciences education. *Adv. Health Sci. Educ.* **2008**, *13*, 3–10. [CrossRef]
- 17. Settles, B.; LaFlair, G.T.; Hagiwara, M. Machine learning-driven language assessment. *Trans. Assoc. Comput. Linguist.* 2020, *8*, 247–263. [CrossRef]
- 18. He, L.Z.; Min, S.C. Development and validation of a computer adaptive EFL test. Lang. Assess. Q. 2017, 14, 160–176. [CrossRef]
- 19. Brezina, V.; Gablasova, D. Is there a core general vocabulary? Introducing the new general service list. *Appl. Lingusit* **2015**, 36, 1–22. [CrossRef]
- Enayat, M.J.; Derakhshan, A. Vocabulary size and depth as predictors of second language speaking ability. System 2021, 99, 102521. [CrossRef]
- 21. West, M.P. A General Service List of English Words: With Semantic Frequencies and a Supplementary Word-List for the Writing Of Popular Science and Technology; Longman Green: London, UK, 1953.
- 22. Coxhead, A. A new academic word list. Tesol Q. 2000, 34, 213–238. [CrossRef]
- 23. Browne, C.; Culligan, B.; Phillips, J. The New General Service List. [Corpus Data]. Available online: http://www. newgeneralservicelist.org (accessed on 10 November 2022).
- 24. Chen, L.C.; Chang, K.H.; Chung, H.Y. A novel statistic-based corpus machine processing approach to refine a big textual data: An ESP case of COVID-19 news reports. *Appl. Sci.* **2020**, *10*, 5505. [CrossRef]
- TCEEC. The TCEEC Word List [Corpus data]. Available online: https://www.ceec.edu.tw/SourceUse/ce37/ce37.htm (accessed on 10 November 2022).
- 26. Chen, Y.Z. Comparing incidental vocabulary learning from reading-only and reading-while-listening. *System* **2021**, *97*, 102442. [CrossRef]
- Laufer, B. Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *Tesol Q.* 2013, 47, 867–872. [CrossRef]
- 28. Nation, I.S.P. How large a vocabulary is needed for reading and listening? Can. Mod. Lang. Rev. 2006, 63, 59–82. [CrossRef]
- Du, G.H.; Hasim, Z.; Chew, F.P. Contribution of English aural vocabulary size levels to L2 listening comprehension. *IRAL-Int. Rev. Appl. Linguist. Lang. Teach.* 2021, 60, 937–956. [CrossRef]
- 30. Masrai, A. The development and validation of a lemma-based yes/no vocabulary size test. SAGE Open 2022, 12, 21582440221074355. [CrossRef]
- McLean, S.; Stewart, J.; Batty, A.O. Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. Lang. Test. 2020, 37, 389–411. [CrossRef]
- 32. Fan, N. Strategy use in second language vocabulary learning and its relationships with the breadth and depth of vocabulary knowledge: A structural equation modeling study. *Front. Psychol.* **2020**, *11*, 752. [CrossRef]
- Chen, L.C.; Chang, K.H.; Yang, S.C. An integrated corpus-based text mining approach used to process military technical information for facilitating EFL troopers' linguistic comprehension: US anti-tank missile systems as an example. *J. Natl. Sci. Found. Sri Lanka* 2021, 49, 403–417. [CrossRef]
- 34. Qian, D.D. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Lang. Learn.* 2002, *52*, 513–536. [CrossRef]
- 35. Nation, I.S.P. Learning Vocabulary in Another Language; Cambridge University Press: Cambridge, UK, 2001.
- Hsu, W.H. The vocabulary thresholds of business textbooks and business research articles for EFL learners. *Engl. Specif. Purp.* 2011, 30, 247–257. [CrossRef]
- Nation, I.S.P.; Heatley, A. RANGE [Computer software]. Available online: http://www.vuw.ac.nz/lals/staff/paul-nation/nation. aspx (accessed on 15 November 2022).
- 38. Durrant, P. To what extent is the academic vocabulary list relevant to university student writing? *Engl. Specif. Purp.* **2016**, *43*, 49–61. [CrossRef]
- 39. Gardner, D.; Davies, M. A new academic vocabulary list. Appl. Lingusit 2014, 35, 305–327. [CrossRef]
- 40. Anthony, L. AntConc (Version 3.5.9) [Computer Software]. Available online: https://www.laurenceanthony.net/software/ antconc/ (accessed on 1 November 2022).
- Munoz, V.L. The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *Engl. Specif. Purp.* 2015, 39, 26–44. [CrossRef]
- Chen, L.C.; Chang, K.H. A novel corpus-based computing method for handling critical word ranking issues: An example of COVID-19 research articles. *Int. J. Intell. Syst.* 2021, 36, 3190–3216. [CrossRef]
- 43. Shih, C.M. The general English proficiency test. Lang. Assess. Q. 2008, 5, 63–76. [CrossRef]

- 44. Kim, C.; Lee, S.Y.; Park, S.H. Is Korea ready to be a key player in the medical tourism industry? An English education perspective. *Iran J. Public Health* **2020**, *49*, 267–273. [CrossRef]
- 45. Nguyen, V.T. *Towards a New Paradigm for English Language Teaching: English for Specific Purposes in Asia and Beyond;* Routledge: Oxfordshire, UK, 2022.
- 46. Pu, S.; Yan, Y.; Zhang, L. Peers, study effort, and academic performance in college education: Evidence from randomly assigned roommates in a flipped classroom. *Res. High. Educ.* **2020**, *61*, 248–269. [CrossRef]
- 47. Skrabankova, J.; Popelka, S.; Beitlova, M. Students' ability to work with graphs in physics studies related to three typical student groups. J. Balt. Sci. Educ. 2020, 19, 298–316. [CrossRef]
- Cheng, J.Y.; Matthews, J. The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. Lang. Test. 2018, 35, 3–25. [CrossRef]
- 49. Dong, Y.; Tang, Y.; Chow, B.W.Y.; Wang, W.S.; Dong, W.Y. Contribution of vocabulary knowledge to reading comprehension among Chinese students: A meta-analysis. *Front. Psychol.* **2020**, *11*, 525369. [CrossRef] [PubMed]
- 50. Ali, M.M.; Hamid, M.O. Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Lang. Assess. Q.* 2020, *17*, 129–146. [CrossRef]
- 51. Chen, Y.C. Assessing the lexical richness of figurative expressions in Taiwanese EFL learners' writing. *Assess. Writ.* **2020**, *43*, 7–18. [CrossRef]
- 52. Jiang, L.J.; Yu, S.L. Appropriating automated feedback in L2 writing: Experiences of Chinese EFL student writers. *Comput. Assist. Lang. Learn.* 2020, 35, 1329–1353. [CrossRef]
- Nghia, T.L.H. "It is complicated!": Practices and challenges of generic skills assessment in Vietnamese universities. *Educ. Stud.* 2018, 44, 230–246. [CrossRef]
- 54. Santucci, V.; Santarelli, F.; Forti, L.; Spina, S. Automatic Classification of Text Complexity. Appl. Sci. 2020, 10, 7285. [CrossRef]
- 55. Forti, L.; Bolli, G.G.; Santarelli, F.; Santucci, V.; Spina, S. MALT-IT2: A New Resource to Measure Text Difficulty in Light of CEFR Levels for Italian L2 Learning. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2016; The European Language Resources Association (ELRA): Paris, France, 2020; pp. 7204–7211.
- Santucci, V.; Bartoccini, U.; Mengoni, P.; Zanda, F. A Computational Measure for the Semantic Readability of Segmented Texts. In International Conference on Computational Science and Its Applications; Springer: Cham, Switzerland, 2022; pp. 107–119.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.