



# What Words Do Learners Need to Achieve 600+ TOEIC Scores?

**Kasin Janjaroongpak**

Faculty of Global Communication

Thai-Nichi Institute of Technology

771/1 Pattanakarn Rd.

Suan Luang, Bangkok, Thailand, 10250

Tel: +66087-982-6315

Email: [kasin@tni.ac.th](mailto:kasin@tni.ac.th)

---

**Abstract:** This study aims to establish essential TOEIC words for intermediate learners. The word list used to trace the TOEIC word list is from Browne and Culligan (2016a). Two groups of participants, the group with 600 or more TOEIC scores, B1 level, and the group with unknown TOEIC scores, respond to 800 selected words from the TOEIC word list in a YES/NO test. The compared responses are classified by frequency rank, CEFR level, suffix, and theme before being arranged in five 40-word sub lists primarily based on frequency rank. 188 unique words are extracted. Most of them are of unknown CEFR level and the highest concentration of frequency rank is at 701-800, approximately 20%, suggesting that one in five words known to high-achievers is exclusive. Suffixes from all word classes are present, notably noun suffixes. The prevalent themes among words in the word list belong to the hospitality or service industry, thanks to its abstract nature. To achieve target TOEIC scores effectively, practitioners and learners should consult relevant TOEIC level-specific word list.

**Keywords:** TOEIC, vocabulary, word list, CEFR, frequency

*Received: May 30, 2024 Revised: December 27, 2024 Accepted: April 23, 2025*

---

## Introduction

### *Vocabulary*

Vocabulary holds a status of a deciding factor in academic success, such as listening (Wang & Treffers-Daller, 2017), GPA (Masrai & Milton, 2018), orthography (Karakoç & Köse, 2017), and TOEIC (McLean et al., 2020; Papol, et al., 2020; Zahrani et al., 2021). Duc (2020) found that learners had problems with both basic and TOEIC-oriented vocabularies.

Undergraduate learners in Thailand generally achieved English proficiency at A2 level and some of them were at A1 (Mafu & Sappapan, 2023; Waluyo et al., 2024). The English proficiency goal of the Ministry of Education as well as the private sector in Thailand was set at B1 (Franz & Teo, 2018). It is necessary to fill this vocabulary gap before attaining 600 scores because 98 percent vocabulary coverage is the established threshold for 70 percent comprehension (Nation, 2006).

### *ESP Word List*

The Word list has been created since the General Service List (West, 1953). Coxhead was the pioneer in specialized word lists (Coxhead, 2000). This list was also led to another robust specialized word list (Gardner & Davies, 2014). Over the last decades, there have been various attempts to provide specialized word lists for many disciplines (Martínez et al., 2009; Nation, 2016). In Thailand, Laosrattanachai and Ruangjaroon (2021) compiled Tourism, Hotel, and Airline Business Word Lists from various sources while Rungrueang et al. (2022) put forth Food Service Word List later reviewed by three experts.

### *TOEIC*

TOEIC is developed by ETS and it claims that the test reflects real-world working life English communication. There are two types of TOEIC, one for the test involving speaking and writing and the other for the test measuring listening and reading comprehension. There are 200 multiple-choice items equally divided between listening and reading sections. The listening part lasts approximately 45 minutes and there are four parts while the reading section takes 75 minutes and there are three parts. Each section scores between 5 to 495.

### *TOEIC Word List*

Although several commercial TOEIC vocabulary learning books have been in circulation, only a few peer-reviewed TOEIC word lists are available. Culligan (2019) compiled the TOEIC Service List which included both reading and listening sections. Junmat and Bunchutrakun (2021) traced the Academic Word List from the TOEIC reading section, while TOEIC listening word list has been later developed (Sengchuen, 2022). Although the TOEIC Service List is comprehensive, neither is it checked against TOEIC scores to verify its credentials nor specifies the level of vocabulary. Many studies scope the vocabulary size for TOEIC (Chujo & Oghigian, 2009; Kaneko, 2017) but they do not incorporate person-informed input and do not discuss level-specific word list. Level-specific vocabulary is important (Chujo & Utiyama, 2006). This study attempts to create the TOEIC word list that is appropriate for expected scores, 600.

### *Research Questions*

1. What are words that the participants with 600 or more scores know, while the unknown score group do not?
2. For a 40-hour course, what is the sequence of words to learn based on frequency and theme?

### **Methodology**

The present study involved three major components: learners-informed input, word list creation, and proposed practical use of the word list being created.

The word list used as a reference in this study was from Culligan (2019). This list was adopted by many studies related to vocabulary (Carney, 2021; Sheridan & Markslag, 2017). The list contained frequency-included TOEIC-oriented vocabularies from respectable publishers, such as Barron's, HarperCollins, and Oxford University Press.

### *Vocabulary Test*

The entire word list contained more than 1,200 entries. The entries were first sorted by frequency rank. Next, the first 800 most frequent words were selected, forming a checklist for participants to check in the yes/no test. Close-end tests reliably assess vocabulary knowledge, especially in

recognizing many words within a limited period (The notion of asking learners to decide between words they know and words they do not know was adopted from the landmark papers (Chiang, 2018; Dang et al., 2022; Read, 2004; Schmitt et al., 2011). This was to reduce the burden on participants, avoiding guessing randomly adopted by Chiang (2018). In addition, the first 800 most frequent words in the list reached 80.9% at 5K vocabulary level, while 78.3% at 5K vocabulary level was reached when calculating with all words in the list, which covered 1259 entries. Given that the 5K vocabulary level is the critical threshold, applying either 800 words or 1259 words should produce, potentially, relatively the same result, but the 800-word test is more practical and manageable for participants.

### *Participants*

This study aimed to trace a unique TOEIC word list for intermediate learners. Purposive sampling was used. The main criterion was the level of TOEIC scores. Participants with official TOEIC scores of 600 were assigned to the first group, while those without scores belonged to the second group. Participants in this study were 187 L1-Thai non-English major undergraduate students from a higher education institute in Thailand. The participants were divided into two groups: one intermediate group, 12 students whose TOEIC scores were 600 or more, and one unknown-TOEIC-score group, 175 students. The different number of informants between the two groups should not be an issue because the control group all holds TOEIC scores of 600 or more, namely, B1 or above levels. Research conducted with different numbers of informants in each category is possible (Li, 2024; Schmitt et al., 2011). Despite the lack of official test results for the latter group, learners attending this institute generally gained TOEIC scores of 400. The participants were between 18 and 22 years old. Genders were reported not to be a significant factor in language proficiency (Chiang, 2018). All participants have studied English for twelve years. The exempt consent was observed as the study is integral to matriculation assessment.

### *Procedure*

The checklist of the selected words was created in a Google form. Participants were asked to complete the checklist in a close-end format. If they know the word in question, they answer “YES” and vice versa. After completing the checklist, the results were exported to a Microsoft Excel spreadsheet. Next, participants with official TOEIC scores submitted their scores, which were added to a column in the spreadsheet. According to Educational Testing Service (2022), the reliability of the TOEIC scores, as measured by the KR-20 reliability index, was at 0.90 and up. The validity of TOEIC scores came from language-testing experts and test-taker language-ability self-assessment results moderately correlating with their TOEIC listening and reading scores. The data were divided into two groups: those with official TOEIC scores of 600 and those without official scores. The statistical analysis function, “mode,” was used. The mode value was used in extracting the unique entries when working with person-informed responses (Laosrirattanachai & Ruangjaroon, 2021).

Then, word class listed in the Cambridge Business English Dictionary (Cambridge University Press, 2011) was assigned to each word. If the word did not appear in the Cambridge Business English Dictionary, the first part of speech listed in the Cambridge Advanced Learner's Dictionary was applied (McIntosh, 2013).

To make the word list aligned with received practice in the vocabulary book (HarperCollins, 2012), the words were arranged thematically. Despite the lack of methods for assigning a word to a particular sub-theme, arbitrariness was possible because different word lists served different purposes (Todd, 2017). After words were assigned to relevant sub-themes, a subset of the word list was created. Each subset contained approximately 40 entries, similar to Todd (2017).

There were 12 sub-themes on the master list. Each sub-list contained a word from every sub-theme arranged by frequency classified in Culligan (2019) to make the sub-list balanced as much as possible. The first sub-list contained high-frequency words, while the last contained low-frequency words. For instance, the sub-theme “Roles” had “clerk” (frequency rank: 26) and “architect” (frequency rank:151) as its members, and the sub-theme “Actions” had “obtain” (frequency rank: 61) and “postpone” (frequency rank: 68) as its constituents.

CEFR based on McIntosh (2013) was included to make this word list more informative. Words without CEFR profiles were specified as “N/A” in their CEFR profiles. If a word has more than one CEFR level, the lowest level will be selected because it is the level at which learners should first encounter the word.

### Findings and Discussion

There are 188 unique entries, 23.5% of the 800 words, in the word list. Both reading and listening comprehension requires at least 90% lexical coverage to achieve appropriate comprehension level (Nation, 2006; Schmitt et al., 2011). Therefore, learners who wish to achieve 600 or more TOEIC scores should understand these words. Examples from the first sub list are presented in the table below. The exhaustive word list is attached in the appendix.

For CEFR profile, there are 3 words in A2 level, 1.60%, 17 words in B1 level, 9.04%, 38 words in B2 level, 20.21%, 41 words in C1 level, 21.81%, 10 words in C2 level, 5.32%, and 79 words in N/A group, 42.02%.

There are 12 sub themes selected from HarperCollins (2012). They are HR, Travel, Orders & Distribution, Roles, Actions, States & Processes, Qualities & Characteristics, Objects, Degree, Sales, Staffing, and Finance. The average number of each sub theme is approximately 16. While the complete list is in the appendix, the first sub list in Table 1 is used as an example.

The 40-word sub list is compiled based on two primary criteria: sub theme and frequency. Most words are pivoted around hospitality and words are getting more abstract as the frequency rank runs.

**Table 1**

*Sub List Example*

| Entries    | Sub Theme             | Frequency Rank | CEFR |
|------------|-----------------------|----------------|------|
| occupation | HR                    | 67             | A2   |
| recipient  | HR                    | 140            | N/A  |
| sincerely  | Travel                | 12             | N/A  |
| enclose    | Travel                | 30             | B2   |
| invoice    | Orders & distribution | 22             | N/A  |
| stationery | Orders & distribution | 209            | N/A  |

*Research Question 1:* What are words that the group with 600 or more scores know while the unknown score group do not?

There are 188 words that are exclusive to the group with 600 or more scores. Their profile–frequency rank, theme, CEFR level, and word components–is explored. Samples of the words in the list are presented in Table 2. The complete word list is in the appendix.

**Table 2**  
*Samples from Word List*

| Word           | Theme                       | Frequency Rank | CEFR |
|----------------|-----------------------------|----------------|------|
| acceptance     | Qualities & characteristics | 705            | C1   |
| accessible     | States & processes          | 437            | B2   |
| accommodate    | Travel                      | 221            | N/A  |
| accomplishment | HR                          | 405            | N/A  |
| accordingly    | States & processes          | 629            | N/A  |

### *Frequency*

Frequency is argued to be one of the important factors in L2 vocabulary acquisition (Nation, 2016; Schmitt, 2010). If learners repeatedly do not understand the words encountered, comprehension will suffer.

188 unique words were traced. Despite frequency cited as a crucial factor determining difficulty (Hansen, 2016), it is not always the case. Words that the unknown group are not familiar with can be of any frequency rank. Vocabulary learning is an activity by and large following frequency bands punctuated by tasks at hand (Milton & Alexiou, 2020). In a given frequency bracket, learners pick up some members in the bracket but leave out the others. One case in point would be the word, “invoice”, whose frequency rank is 22. Both groups know the word, “brochure”, whose frequency rank is 23, but only the high-achieving group knows the word “invoice.” Table 3 sums up the frequency rank overview.

**Table 3**  
*Frequency Rank*

| Frequency Rank | Number of Words | Percent |
|----------------|-----------------|---------|
| 1-100          | 15              | 7.98%   |
| 101-200        | 18              | 9.57%   |
| 201-300        | 26              | 13.83%  |
| 301-400        | 22              | 11.70%  |
| 401-500        | 27              | 14.36%  |
| 501-600        | 20              | 10.64%  |
| 601-700        | 24              | 12.77%  |
| 701-800        | 36              | 19.15%  |
| Total          | 188             | 100.00% |

Frequency rank of 1, “mister,” means the word is the most frequent word in the list while frequency rank of 800, “spreadsheet,” is the least frequent word in the list. Presumably, the first bracket, 1-100, accounts for the lowest proportion in the master list thanks to its high frequency. Unique words in this bracket mostly belong to the workplace such as sincerely, cabinet, clerk, enclose, and lease. Another interesting pair is the word, “reception” (frequency rank: 49), and “receptionist” (frequency rank: 75). Only the high-achieving group knows “receptionist” despite the only difference in suffix between the two words, thanks to proficiency difference (Schmitt, 2010).

In the 101-200 frequency rank, 6 words belong to the Travel sub theme, accounting for almost half of the bracket. The 1-200 frequency rank bracket accounts for approximately 20% of the total, which implies that 1 in 5 words of the master list appears a lot but the unknown score group does not know it.

The 201-300 bracket takes up 13.83% of the total, more than half of the 1-200 combined. From this point onwards, the percent of words in each bracket will be more than 10% suggesting that low frequency words are of particular concern to the unknown score group.

The next bracket, 301-400, contains four adverbs ending with “-ly”: annually, accurately, promptly, and efficiently. At the same time, the first bracket has only one adverb, sincerely, whose ending is “-ly” but there are 8 nouns followed by 4 verbs and 2 adjectives, suggesting the priority of adverbs.

The 401-500 bracket takes the second largest share in the master list. One of the striking features in this bracket is words associated with abstract or qualities prevalent in the following brackets: spacious, exemption, mandatory, residential, accessible, durable, malfunction, and patience. The last bracket, 701-800, takes up the largest share of the list, 19.15%, which is approximately three times of the first bracket. There are 36 entries in the bracket.

To sum up, words that the unknown TOEIC score group does not know tend to be low frequency words, especially the words in the 701-800 frequency rank bracket. In addition, the stark contrast between the two groups can be seen from the frequency rank of 201-300 brackets .

### *Suffix*

Despite a small number of words in the word list, many suffixes appear in the word list. In the word list, there are 13 noun suffixes, 11 adjective suffixes, 3 verb suffixes, and 1 adverb suffix as described in the four following tables. The reporting tables follow the patterns used in Suhandoko & Ningrum (2020).

**Table 4**

#### *Noun Suffix*

| <b>Noun Suffix</b>           | <b>Frequency</b> | <b>Sample of Words</b> |
|------------------------------|------------------|------------------------|
| -al                          | 6                | rental                 |
| -sion/-tion/-xion / -(at)ion | 18               | occupation,            |
| -ist                         | 2                | receptionist           |
| -ee                          | 1                | attendee               |
| -ance/-ence                  | 4                | inconvenience          |
| -ar/-er/-or                  | 7                | inspector              |
| -age                         | 1                | beverage               |
| -ant/-ent                    | 1                | recipient              |
| -ary/-ery/-ory               | 3                | itinerary              |
| -ment                        | 3                | refreshment            |
| -ity/-ty                     | 2                | authority              |
| -acy/-asy/-cy                | 2                | occupancy              |
| -ness                        | 1                | effectiveness          |

**Table 5**

#### *Adjective Suffix*

| <b>Adjective Suffix</b> | <b>Frequency</b> | <b>Sample of Words</b> |
|-------------------------|------------------|------------------------|
| -al/-ial/-ual           | 7                | residential            |
| -ive                    | 3                | administrative         |
| -able/-ible             | 4                | affordable             |
| -ary/-ory               | 4                | complimentary          |
| -y                      | 1                | lengthy                |
| -ly                     | 2                | quarterly              |
| -ous                    | 4                | generous               |
| -ic                     | 3                | enthusiastic           |
| -ate                    | 2                | inappropriate          |
| -ent                    | 1                | insufficient           |
| -ed                     | 2                | unattended             |

**Table 6***Verb Suffix*

| Verb Suffix | Frequency | Sample of Words |
|-------------|-----------|-----------------|
| -ise/ ize   | 3         | authorize       |
| -ate        | 11        | accommodate     |
| -ify        | 2         | clarify         |

**Table 7***Adverb Suffix*

| Adverb | Frequency | Sample of Words |
|--------|-----------|-----------------|
| -ly    | 16        | sincerely       |

For noun suffix, the highest frequency is -sion/-tion/-xion / -(at)ion suffix while there are four suffixes that return only one frequency, -ee, -age, -ant/-ent, and -ness. For adjective suffix, the highest frequency is -al/-ial/-ual suffix whereas -y and -ent suffixes have only one each. The highest verb suffix is -ate suffix but there are only two for -ify suffix. The only adverb suffix is -ly. In terms of number of suffixes, adverb and verb suffixes can be introduced before noun and adjective suffixes as there are only four suffixes to be discussed and their frequency is relatively high compared with noun and adjective suffixes. For instance, the total frequency of adjective suffixes is 33 while that of verb suffix is 16, approximately 50%, despite a clear difference between the number of adjective suffixes and verb suffixes, 11 and 3 respectively.

More than 60% of the words in the word list, 117 from 188, are suffixed words from 29 suffixes, while there are 3000 words in the reference word list and 1251 of them are suffixed words with 41 suffixes: 22 noun suffixes, 13 adjective suffixes, 4 verb suffixes, and 2 adverb suffixes, suggesting that suffix knowledge, especially the breath of suffix knowledge, is important for understanding upper intermediate TOEIC words.

For noun suffixes, the shared suffixes, 13 suffixes, are -age, -al, -ance/-ence, -ary/-ery/-ory, -acy/-asy/-cy, -ee, -ant/-ent, -ist, -ment, -ness, -ar/-er/-or, -sion/-tion/-xion / -(at)ion, and -ity/-ty. Despite a substantial different number of words, 3000 and 188, all noun suffixes in the word list of the present study appear in the noun suffix list of the reference study suggesting the diversity of noun suffixes at this level.

For verb suffixes, the three verb suffixes are -ise/ -ize, -ate, and -ify. All verb suffixes in the present study overlap with the reference verb suffix word list. Some verbs in the present list suggest that English verbs can be formed by functional shift such as audit, invoice, clerk, enclose, lease, obtain, which is a frequent occurrence in English online article (Vaneva & Bojadjev, 2020) and it is understudied (Lindstromberg & Boers, 2023). One notable difference between the two lists is that the most frequent verb suffix in the reference study is -ize while -ate suffix is the most frequent one in the present study. It seems that -ate is the suffix relatively specific to B1-B2 level whereas -ize is a more general verb forming suffix.

For adjectives, every adjective suffix overlaps except the -ent suffixes. Although the different numbers of suffixes between noun and adjective suffixes are negligible, noun suffix should be prioritized as noun is the nomination of things, while adjective is simply an additional information on those things. Al-Jarf (2019) found that a considerable number of learners in the study mismatched between noun and adjective suffixes.

For adverbs, -ly is by far the most prominent adverb forming suffix compared with -wise or -ward. One of the benefits of studying words with -ly suffix might not necessarily be on the adverb itself,

but the word, specifically, adjective that the -ly suffix attached such as the “temporarily” and “temporary” pair.

The finding on suffixes suggests that every suffix is not created equal. The most frequent suffixes should be introduced before the less frequent ones. To achieve proper coverage, each suffix type should be adequately covered. Schmitt (2010) proposed 85% coverage as a mastery level threshold.

### CEFR

Recently, word lists in records include various dimensions of vocabulary to make the word lists relevant and practical (Lea & Bradbery, 2020; Mäntylä & Huhta, 2014) especially TOEIC word list (Sengchuen, 2022).

**Table 8**

#### *CEFR Level Breakdown*

| CEFR | Word | Percent |
|------|------|---------|
| A2   | 3    | 1.60%   |
| B1   | 17   | 9.04%   |
| B2   | 38   | 20.21%  |
| C1   | 41   | 21.81%  |
| C2   | 10   | 5.32%   |
| N/A  | 79   | 42.02%  |

Considering the scores at 600 or more, A2 and B1 levels altogether account for approximately ten percent, suggesting that before obtaining scores at this level learners need to first master vocabulary at prior levels. Around 40 percent of words in the list are at B2 and C1 levels. This corresponds with the score level itself in that 600 scores are considered between B1 and B2, resulting in many unknown words taking up at B2 end.

It is of great interest to investigate why N/A group takes up nearly half of the list. In terms of frequency rank, it seems to be almost evenly distributed. For instance, the frequency rank of 1-100 contains 7 entries while the next frequency rank of 101-200 returns 10 entries. For each category, every category contains N/A entries, but only some categories contain more entries than others. The categories containing fewer N/A entries are Degree, 1 entry, Object, 4 entries, Staffing, 3 entries while categories including more N/A entries are Actions, 9 entries, Finance, 8 entries, HR, 6 entries, Orders & distribution, 5 entries, Qualities & characteristics, 5 entries, Roles, 8 entries, Sales, 7 entries, States & processes, 11 entries, and Travel, 12 entries. One of the common features among the categories with more N/A entries is its abstraction and abstract words are often associated with low frequency words. One case in point would be the N/A words in the frequency rank of 700-800. The categories with this frequency range are Sales, Orders & distribution, Qualities & characteristics, Objects, Actions, Roles, States & processes. It seems to be difficult to assign CEFR level to these entries because they appear few and far between in public discourse, resulting in difficulty in determining its relevance to the text and meaning contributing to the text. Additionally, from the practicality standpoint, assigning a CEFR level to a word is a herculean task; therefore, selecting words to undergo such a process should be worthwhile and duly serve pedagogical and social purposes.

The part of speech of the N/A words include all major parts of speech, noun, verb, adjective, and adverb. As expected, nouns take up more than half of N/A words, 48 entries, followed by verbs at 27 entries. Lindstromberg and Boers (2023) found that in general the distribution among part of speech of a given corpus was roughly 50% noun, 25% verb, and 25% adjective. Adverbs sometimes take up a negligible percentage in some corpus. The number of verbs in the N/A words is interesting



in that it accounts for roughly 34% of all the N/A words, suggesting that determining CEFR level of verbs is relatively difficult. Verbs, especially important ones, are polysemous. To illustrate, the verb, “run”, according to Cambridge Dictionary, contains at least 9 meanings and each meaning, most of the time, has its unique CEFR level. The word, “run,” meaning go quickly, for example, is A1, while “run” meaning show is C1.

The two main pieces of information provided in the word list are CEFR levels and frequency rank which are competing priorities. Pintard and François (2020) argued for combining the two, while Sohsah et al., (2015) took CEFR levels, expert opinions, and machine learning into consideration. I would argue that frequency rank should be prioritized as it is more accessible and practical for practitioners followed by expert opinions.

*Research Question 2:* For a 40-hour course, what is the sequence of words to learn based on frequency and theme?

This study follows the number of entries in a subset of a word list and hours of study by Todd (2017) and Laosrirattanachai and Ruangjaroon (2021). Samples of 5 sub lists are presented in the table below. Category is based on HarperCollins (2012), the frequency rank is based on Culligan (2019), and CEFR level is adapted from McIntosh (2013). Although the word list is traced from a relatively small number of participants, it should be reliable thanks to informants’ proficiencies.

**Table 9**

*Samples of Sub Lists*

| Entries      | Category           | Frequency Rank | CEFR Level | Sub List No. |
|--------------|--------------------|----------------|------------|--------------|
| occupation   | HR                 | 67             | A2         | 1            |
| dial         | Actions            | 157            | B1         | 2            |
| periodically | Degree             | 649            | N/A        | 3            |
| separately   | States & processes | 626            | B2         | 4            |
| entrepreneur | Roles              | 683            | N/A        | 5            |

For the first sub list, most of the words are concrete words rather than abstract words as they are more imaginable and concrete and imaginability and concreteness are strongly correlated with learnability (Pichette et al., 2012). For the next sub list, the highlighted feature is dual meaning words, such as specification, allocate, and frustrate. Various studies report that polysemous words are a challenging concept for L2 English learners (Lee, 2021; Yuananda & Prayogo, 2022).

In the fourth sub list, many words are concerned with negativity, such as deduct, bankrupt, and inappropriate, suggesting that words are getting critical and nuanced. Another common group is about events such as reopen, beforehand, and alternate.

Many themes in the word list of this study are associated with service industry or hospitality, such as HR, Travel, Roles, Sales, Staffing, and Finance, accounting for 95 entries out of 188. One possible explanation is that these themes are of abstract nature, which is more difficult to acquire, compared with concrete words as reported in Lana and Kuperman (2024).

### **Pedagogical Implications**

This study reveals that an exclusive TOEIC vocabulary list exists and learners aiming for 600 TOEIC scores or more should direct their attention to B2 words followed by C1 words after mastering basic and pre-intermediate vocabulary. At this stage, it is challenging for both practitioners and test takers to address the difficulty levels of TOEIC vocabulary as unknown CEFR level words take up the largest share of the word list. Learners' lexical size should be measured before using the word

list (Schmitt, 2010) and word lists should be incorporated into appropriate courses (Smith et al., 2020). When a class meets, two to three words from the sub list should be discussed and accountability measures such as dictation should be observed.

Many lexis in the word list of this study are suffixed words suggesting its importance in relation to vocabulary learning (Mochizuki & Aizawa, 2000), for B1 level in particular (Mäntylä & Huhta, 2014). Outstanding thematic clustering of this word list primarily concerns leisure industry followed by abstract.

### **Limitation of the Study**

The study's limitation lies in the sample size of both the control and experiment groups. The word list used to assess TOEIC vocabulary knowledge should be a complete word list or a more authoritative word list should be used to measure learners' potential different levels of vocabulary knowledge. The close-end test should incorporate a non-word approach, ensuring that the responses truly reflect learners' vocabulary knowledge as much as possible.

### **Recommendation**

Future research should address the limitations identified in this study, such as the effectiveness of the word list in this study in relation to TOEIC scores. The word list could also be traced with other standardized proficiency tests, such as IELTS and TOEFL, which are generally associated with academic English. In addition, the study should divide participants into three groups, elementary, intermediate, and advanced levels so that the results could suggest potential different lexical profiles among these groups. To move beyond single word level, a level-specific TOEIC list of multi-word expressions should be explored.

### **REFERENCES**

- Al-Jarf, R. (2019). Freshman students' difficulties with English adjective forming suffixes. *Journal of Foreign Language Teaching and Applied Linguistics*, 6(1), 169-180.
- Cambridge University Press. (2011). *Cambridge business English dictionary*. Cambridge University Press.
- Carney, N. (2021). Referencing word and phrase lists in an online vocabulary notebook. *Language Teaching Research*, 28(1), 253-272.
- Chiang, H.-H. (2018). English vocabulary size as a predictor of TOEIC listening and reading achievement among EFL students in Taiwan. *Theory and Practice in Language Studies*, 8(2), 203-212.
- Chujo, K., & Oghigian, K. (2009). How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary. *The Journal of Asia TEFL*, 6(2), 121-148.
- Chujo, K., & Utiyama, M. (2006). Selecting level-specific specialized vocabulary using statistical measures. *System*, 34(1), 255-269.
- Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, 34(2), 213-238.
- Culligan, B. A. (2019). Evaluating corpora with word lists and word difficulty. *Vocabulary Learning and Instruction*, 8(1), 29-38.

- Dang, T. N., Webb, S., & Coxhead, A. (2022). Evaluating lists of high frequency words: teachers' and learners' perspectives. *Language Teaching Research*, 26(4), 617-641.
- Duc, L. P. (2020). Difficulties in test of TOEIC reading comprehension by the final year students of high quality classes at faculty of mechanical engineering, *HaUI. TRƯỜNG ĐẠI HỌC THỦ ĐỨC HNỘI*, 102-111. Retrieved from <https://easychair.org/publications/preprint/Wd6x>
- Educational Testing Service. (2022). *ETS examinee handbook*. ETS.
- Franz, J., & Teo, A. (2018). 'A2 is normal' – Thai secondary school English teachers' encounters with the CEFR. *RELC Journal*, 49(3), 322-338.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Hansen, P. (2016). What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development. *First Language*, 37(2), 205–225.
- HarperCollins. (2012). *COBUILD key words for the TOEIC test*. HarperCollins.
- Junmat, W., & Bunchutrakun, C. (2021). Academic word list of reading part in TOEIC sample test. *International Journal of Management and Applied Science*, 7(4), 11-14.
- Kaneko, M. (2017). Vocabulary size targets for the TOEIC test. *JACET Journal*, 61(2017), 57–67.
- Karakoç, D., & Köse, G. D. (2017). The impact of vocabulary knowledge on reading, writing and proficiency scores of EFL learners. *Journal of Language and Linguistic Studies*, 13(1), 352-378.
- Lana, N., & Kuperman, V. (2024). Learning concrete and abstract novel words in emotional contexts: Evidence from incidental vocabulary learning. *Language Learning and Development*, 20(2), 158-173.
- Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 50-86.
- Lea, D., & Bradbery, J. (2020). *Oxford advanced learner's dictionary*. Oxford University Press.
- Lee, Y. J. (2021). Polysemous words in English movies, learning obstacles or gifted talent? *Journal of English Teaching through Movies and Media*, 22(4), 14-26.
- Li, C. (2024). Exploring aural vocabulary knowledge for TOEIC as a language exit requirement in higher education in Taiwan. *International Review of Applied Linguistics in Language Teaching*, 62(4), 1853-1875.
- Lindstromberg, S., & Boers, F. (2023). Word classes in second language acquisition. In E. V. Lier (Ed.), *The Oxford handbook of word classes* (pp. 876-886). Oxford University Press.
- Mafu, A., & Sappapan, P. (2023). A study of undergraduate students' problems toward taking an English proficiency test based on CEFR. *Journal of Positive School Psychology*, 7(8), 242-251.

- Mäntylä, K., & Huhta, A. (2014). Knowledge of word parts. In J. Milton, & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 45-59). Palgrave Macmillan.
- Martínez, 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.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44-57.
- McIntosh, C. (2013). *Cambridge advanced learner's dictionary*. Cambridge University Press.
- McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, 37(3), 389-411.
- Milton, J., & Alexiou, T. (2020). Vocabulary size assessment: Assessing the vocabulary needs of learners in relation to their CEFR goals. In M. Dodigovic, & M. P. Agustín-Llach, *Vocabulary in curriculum planning: Needs, strategies and tools* (pp. 9-27). Palgrave Macmillan.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291-304.
- Nation, I. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-59.
- Papol, R., Boonsue, W., Nakcharoen, J., Sriratanaban, S., Silachai, N., & Khamsri, T. (2020). The analysis of listening and reading problems in taking TOEIC of English major students, faculty of humanities and social sciences, Chiang Mai Rajabhat university. *Journal of Buddhist Education and Research (JBER)*, 6(2), 239-251.
- Pichette, F., De Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, 33(1), 66-82.
- Pintard, A., & François, T. (2020). Combining expert knowledge with frequency information to infer CEFR levels for words. In N. Gala, & R. Wilkens (Eds.), *Proceedings of the 1st workshop on tools and resources to empower people with READING Difficulties (READI)* (pp. 85-92). European Language Resources Association.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-227). John Benjamins.
- Rungrueang, T., Boonprasert, P., Poempongsajaroen, S., & Laosrirattanachai, P. (2022). Corpus-based approach to generate a word list for food service. *THAITESOL Journal*, 35(1), 57-76.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan London.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.

- Sengchuen, N. (2022). A corpus-based study of TOEIC listening vocabulary for developing TOEIC listening word list for the listening exam. *The Golden Trek: Humanity and Social Science Journal*, 29(1), 74-90.
- Sheridan, R., & Markslag, L. (2017). Effective strategies for teaching vocabulary: An introduction to engaging cooperative vocabulary card activities. *PASAA: Journal of Language Teaching and Learning in Thailand*, 53, 214-229.
- Smith, G. F., Kyle, K., & Crossley, S. A. (2020). Word lists and the role of academic vocabulary use in high stakes speaking assessments. *International Journal of Learner Corpus Research*, 6(2), 193-219.
- Sohsah, G. N., Ünal, M. E., & Güzey, O. (2015). Classification of word levels with usage frequency, expert opinions and machine learning. *British Journal of Educational Technology*, 46(5), 1097-1101.
- Suhandoko, & Ningrum, D. R. (2020). A corpus-based list of academic English derivational suffixes. *Indonesian Journal of Applied Linguistics*, 10(2), 481-490.
- Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, 45, 31-39.
- Vaneva, M., & Bojadjev, M. I. (2020). English zero derivation revisited: Nouning and verbing in online business articles. *International Journal of English Linguistics*, 10(6), 307-321.
- Waluyo, B., Zahabi, A., & Ruangsung, L. (2024). Language assessment at a Thai university: A CEFR-based test of English proficiency development. *REFlections*, 31(1), 25-47.
- Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, 65, 139-150.
- West, M. (1953). *General service list of English words*. Longman Green & Co.
- Yuananda, Y., & Prayogo, J. A. (2022). The polysemous words found in business section of the Economist magazine. *JoLLA: Journal of Language, Literature, and Arts*, 2(1), 77-90.
- Zahruni, N. A., Fahmi, F., & Pratolo, B. W. (2021). The challenges of taking TOEIC test and how to overcome: Perception of Indonesian vocational students. *Ethical Lingua: Journal of Language Teaching and Literature*, 7(1), 82-91.

### Appendix

|               |               |             |                |
|---------------|---------------|-------------|----------------|
| acceptance    | accessible    | accommodate | accomplishment |
| accordingly   | accumulate    | accurately  | administrative |
| affordable    | airfare       | aisle       | allocate       |
| alternate     | alternatively | ambitious   | amuse          |
| annually      | architect     | aspect      | attendee       |
| audit         | auditorium    | authentic   | authority      |
| authorization | authorize     | balcony     | bankrupt       |
| bargain       | beforehand    | behalf      | beverage       |
| briefcase     | bulletin      | cabinet     | clarify        |
| clerk         | commute       | compile     | complimentary  |

|               |               |   |   |
|---------------|---------------|---|---|
| comprehension | considerably  | consistently                              | consultation                              |
| consumption   | conveniently  | coordinate                                | correspondence                            |
| costly        | courier       | criteria (redirected to criterion)        | cruise                                    |
| dedication    | deduct        | deduction                                 | definite                                  |
| descriptive   | diagnose      | dial                                      | discontinue                               |
| dishwasher    | disruption    | dissatisfy (redirected from dissatisfied) | distribution                              |
| dividend      | duplicate     | durable                                   | duration                                  |
| eager         | editorial     | effectiveness                             | efficiently                               |
| enclose       | enthusiastic  | entrepreneur                              | exceptional                               |
| exemption     | expertise     | facilitate                                | familiarize                               |
| faulty        | finalize      | foreman                                   | fountain                                  |
| freight       | frustrate     | furnish                                   | garment                                   |
| generic       | generous      | grocery                                   | hazardous                                 |
| hike          | humidity      | immigration                               | inappropriate                             |
| inclusive     | inconvenience | influential                               | inquire                                   |
| inspect       | inspector     | instruct                                  | insufficient                              |
| intern        | intonation    | introductory                              | invoice                                   |
| irritate      | itinerary     | leak                                      | lease                                     |
| leisure       | lengthy       | ma'am                                     | malfunction                               |
| mandatory     | merchandise   | merge                                     | mislead                                   |
| nationwide    | nominate      | obtain                                    | occupancy                                 |
| occupation    | omit          | orientation                               | oval                                      |
| overdue       | paraphrase    | parcel                                    | patience                                  |
| payable       | periodical    | periodically                              | pharmaceutical                            |
| pharmacist    | plumber       | postpone                                  | proficiency                               |
| promptly      | proofread     | punctual                                  | quarterly                                 |
| questionnaire | rack          | receptionist                              | recipient                                 |
| recreation    | recruiter     | referral                                  | refreshment                               |
| rehearse      | reinforce     | relocate                                  | renewal                                   |
| rental        | renter        | reopen                                    | repairperson (redirected from repairman)  |
| repeatedly    | reschedule    | residential                               | respondent                                |
| retailer      | rumor         | satisfactory                              | sauna                                     |
| seldom        | seminar       | separately                                | sightsee                                  |
| sincerely     | skim          | spacious                                  | specification                             |
| spokesperson  | stationery    | steadily                                  | stockholder (redirected from shareholder) |
| strictly      | subtract      | temporarily                               | tray                                      |
| tuition       | unattended    | unspecified                               | urgent                                    |