

CEFR-Referenced Item Specification Analysis of TOEIC Incomplete Sentences Part on Intermediate Thai Learners

Kasin Janjaroongpak

College of General Education and Languages, Thai-Nichi Institute of Technology, Thailand

Email: kasin@tni.ac.th

Received: 12/09/2022

Revised: 21/11/2022

Accepted: 01/12/2022

Abstract

While there have been many studies exploring the reading section in TOEIC, there are a few studies examining item specification. The objective of this study is to analyze item specification of the part-5 test scores of 52 mostly intermediate learners taking intermediate TOEIC-formatted test. The analysis was conducted by identifying the item specification in each test item as well as its CEFR level in test items whose correct responses are low. The result shows that the CEFR levels of the test items can predict, to a large extent, each test taker's proficiency and the distribution of the test items with low-correct responses is in line with the CEFR levels of the test items and learners' TOEIC scores. Test takers at the B1 level by and large cannot answer test items at the B2 or above level. The test takers at the B1 level may not necessarily answer all the B1 level questions should they be at B1 threshold. The item specification analysis also reveals that the most frequent type of errors found in the study is word choice and tense-aspect, respectively. Along with CEFR, other possible explanations for errors in each test item are proposed. Pedagogical implications are that instructors should focus on vocabulary learning in context, collocation and colligation, verb tense, respectively.

Keywords: CEFR, TOEIC®, Incomplete Sentences, Item specification

1. Introduction

TOEIC® has been used in an expanding circle for years thanks to its practicality and being fit for purpose. Companies especially international ones expect their potential candidates to submit TOEIC® scores along with relevant documents in their recruitment processes. Traditionally, total scores at 550, 275 for reading scores and 275 for listening scores respectively, are considered acceptable as a reliable yardstick for operational level or entry level. This score level is equivalent to B1 level on CEFR scale (Educational Testing Service [ETS], 2021). Owing to its prevalence, people utilize TOEIC® as a measure for general English (Powers & Powers, 2015). TOEIC® as a reliable informant of language ability in business English was strongly correlated with employability (Wilson et al., 2004). Not only is TOEIC® applicable for employment but also acceptable for certain higher education institutes in Japan (Hokuriku University, 2021).

To be successful in TOEIC®, candidates had to be competent in both English in general and English for Specific Purpose.

In this vein, a sizeable number of TOEIC® test takers in Thailand are anxious about answering incomplete sentences but they do not know exactly what grammar point they are weak at or what are words that they have to know to perform well in this high-stake examination. Part 5, Incomplete Sentences, in TOEIC is similar to cloze test with multiple choices. For instance, the question is “This is ___ cat” followed by four possible choices, “a. a, b. an, c. the, d. of”. Other studies on TOEIC in Thailand have focused on either teaching method effectiveness or opinion elicitation in a form of questionnaire. Little research has been done on problematic item specification in TOEIC Incomplete Sentences part, which accounts for 30 test items, in Thai learners. If instructors have a firm general overview on learners’ problems in this test section, they will be able to direct their attention and resources more effectively to address their students’ challenges in sitting this high-stake standardized test. For instance, if the main problem of the intermediate Thai learners generally lies in subject-verb agreement, instructors could prepare teaching materials in advance focusing on this issue rather than bombarding their students with grammar points that the students have been mastered such as connecting word.

2. Research Objectives

The purpose of this study is to (A) explore whether errors made in the part 5 incomplete sentence to the test by intermediate learners when checked against CEFR level of each test item can predict learners’ proficiency level (B) tease out lexicogrammar points that intermediate learner, B1, find challenging and explore possible explanation.

3. Literature Review

CEFR and TOEIC

Though the mapping between CEFR and TOEIC has been gaining traction, the studies on this are limited. Tannenbaum and Wylie (2013) laid out groundwork for mapping between TOEIC and CEFR but do not offer details on what type of questions belong to A1 and what type of questions is of C1. Runnels (2016) found moderate correlations between TOEIC listening and CEFR-J (a modified version of the CEFR) self-assessment scores but no correlation was observed for reading while Richard (2020) explored CEFR-J self-assessment and TOEIC listening and reading scores. A possible research gap from this study is that this study uses CEFR self-assessment scores rather than objective CEFR-informed references such as English Grammar Profile or English Vocabulary Profile. In addition, the listening section and the reading section in this study are not analyzed based on its components, part 1 to 4 for listening and part 5 to 7 for reading. Part-based analysis might provide more insights into the relation between CEFR and each section.

Related Studies

Error Analysis

Only a few studies explored TOEIC in light of error analysis. Papol et al. (2020) provided an overview of error analysis based on participants’ responses to both listening and reading sections while Wangmanee & Vongtangswad (2022) conducted the research in the same fashion

with two groups of participants. The common features between the two studies are that the reading section is of particular concern in that the reading scores significantly fall behind listening scores. The classification of language skills required in the reading section of the two studies might not offer insights into specific problems that learners are encountering. The five classifications in reading comprehension are vocabulary, part of speech, tense, comprehension, and inference. The fourth domain, comprehension, is redundant with the whole section, reading comprehension, itself. In addition, problems identified in the reading section do not provide proficiency level associated. For instance, the study reports that vocabulary is one of the problems in the reading section but it does not specify what those words are and at what difficulty level they are.

TOEIC Part-Specific Analysis

There has been one study focusing on Question-Response lexical distractors in the listening section of TOEIC (Taladngoen & Esteban, 2022). The study revealed that repeated words, similar-sounding words, and word associations were the commonest lexical distractors, which suggests that exploring the most problematic types of questions are important. The types of distractors identified in this research are similar to types of distractors found in Rogers (2006). These distractors concern the nature of listening in that the aural property of the word is used to discriminate test takers' proficiency levels. For reading, these distractors are absent. Therefore, analyzing a part of TOEIC reading section might offer insights into the nature of the TOEIC reading section. Besides, there has been no study focusing on test types of TOEIC part 5, the reading section, and CEFR levels of the questions in TOEIC part 5. The present study is an attempt to explore this point.

TOEIC and its Reading Section

Maliwan (2018) reported that students' average TOEIC scores were 200-400, which is markedly lower than the expected scores of 405-600 (Papol, et al., 2020). To address this, Papol, et al., (2020) and Wangmanee & Vongtangswad (2022) describe problems in a cognitive fashion, such as inferencing, reading comprehension, and prediction. Though an analysis is beneficial for giving an overview of the problems, it does not provide insights into problematic test types in TOEIC that affect test scores. On this front, Taladngoen and Esteban (2022) studied the nature of test types in TOEIC part 2, which belongs to the listening section. Research on test types in the reading section has been limited.

Wei and Low (2017) conducted rigorous research on TOEIC score change patterns. The unconditional means model for listening and reading scores revealed that the estimated mean of all repeaters' listening scores was 333.75 while the estimated mean of reading scores was 279.30. Harada (2016) reported that in summer 2016 the mean listening scores of participants were 273.4 while the mean reading scores were 150.8. Wu (2013) also reported that the posttest listening score mean was 280.56 while the posttest reading score mean was 177.16. The difference between the mean listening score and reading score was also observed in studies in Thailand (Papol, et al., 2020; Wangmanee & Vongtangswad, 2022). Every study cited indicates that reading score always lags. To achieve better scores, test takers need to overcome reading section challenges.

Papol, et al. (2020) reported that more than half of all reading questions were problematic to test takers. For part 5 (Incomplete Sentences), 15 questions are not a problem while the other are. If this section is carefully analyzed, test takers may significantly score higher. In addition, it

takes 30 seconds to a couple of minutes to correctly answer a test item in part 5 but it may well take more than 10 minutes to answer the 5-question set in part 7 (Passage). Though there are more test items in part 7, 54 questions out of 100, a closer look suggests that it comprises 29 single-passage questions, 10 double-passages questions, and 15 multiple-passages questions. Therefore, it is possible to say that part 5 takes the largest share in the reading section. Trew (2007) pointed out that part 5 primarily tests test takers' lexicogrammatical integrity. Therefore, this study focuses on part 5 of the test.

Chiang (2018) established the link between vocabulary and TOEIC listening and reading comprehension. In addition, Taguchi (2015) and Wang (2015) supported the importance of vocabulary in determining TOEIC scores. Essentially, vocabulary is a key ingredient in understanding language.

Kamijo (2010) analyzed text types, text structures, and question types in the TOEIC reading section but offered neither CEFR measures nor the relationship between questions and answers. My study would explore how these two components could uncover test takers' problems in answering TOEIC test items.

4. Research Questions

- By means of analyzing mistakes made by learners in Part 5 (Incomplete Sentences), to what extent can CEFR classify item specification in relation to total TOEIC-formatted scores of the test?
- What is the distribution of types of errors found?
- Apart from CEFR, what are possible explanation of the errors made by learners?

5. Research Method

Participant

The participants were 52 business-concentration students enrolling in TOEIC preparation course at a higher education institute in Thailand. At the time when the study was conducted, only these participants met the screening criteria: being at intermediate level.

Sampling

This paper uses purposive sampling because the study focuses on intermediate business students and these students were upper-elementary to intermediate learners because they were 3rd or 4th year students who have passed previous English courses which required A2 level to complete the course.

Proficiency level

Considering past records of students with relatively the same stature, proficiency level of these learners was assumed by and large to be at upper elementary to intermediate level. In addition, according to CEFR mapping provided by ETS (2021), the average score of participants from the test administered were at 287.40 for listening and 183.94 for reading. This combination passed the threshold for B1, 275, in listening while failed to clear the bar in reading, 275.

Therefore, these learners sat between A2 in reading and B1 in listening. Their language performance should be considered comparable to those with the same proficiency level elsewhere. These students had enrolled at business administration school of a multinational-oriented vocational higher education institute. TOEIC was introduced from their first year and business textbooks used in their core courses were mostly in English. This study proceeded with exempt consent from participants as the test result was an integral part of course evaluation. The scores came from a mock test from a respectable intermediate commercial TOEIC course book in Thailand (Rogers, 2006). Total score conversion, 990, was carried out. In theory, it would be best if the official test paper were used instead of a test formatted after TOEIC test paper. However, the test paper used in this study was taken from a respectable publisher. Both the validity of reliability of the test should be of acceptable standard.

Despite a sizeable number of approaches to categorize TOEIC test items, this paper adopted classification guideline from a test preparation book (Rogers, 2006). Another measure used was English Profile from Cambridge University which divided into English Vocabulary Profile and English Grammar Profile. For vocabulary, if CEFR level of the word was not assigned, BNC would be consulted for the frequency of the word.

Participants sat this test, their final examination, as a part of matriculation assessment. Exempt consent is sought. The full test paper was a TOEIC-formatted 200-item test paper. This study analyzed only part 5, Incomplete Sentences. All responses were checked by computer software for processing multiple choice answer sheet. The responses were classified in a form of Excel file. The file detailed responses from each participant in each test item from test item number 101 to 130. Percentage of correct responses in each test item was calculated and percentage of correct response ratio of lower than 55% was highlighted. a formula, MODE, then, was applied to each test item concerned to see which choice was selected the most. MODE was a function in Excel that took out the most repeated value within the range selected. This function was used to search for the answers most selected by learners i.e., representation. The reason why the threshold was at 55% derived from the goal of 600 TOEIC scores which was considered a desirable proficiency level, B1. According to ETS (2021), the score of 550, 275 from reading and 275 from listening, was acceptable because ETS (2022) indicated that candidates at 250 reading score could understand easy and some medium vocabulary while those at 200 listening score could grasp the gist of recordings. The threshold set at 600 to avoid standard error of measurement for TOEIC, ± 25 scores, and the standard error of measurement was doubled to make sure that this group was clearly different from the other group. Next, the most selected choice was compared with the correct answer and the test items which most participants incorrectly answered would be of particular attention. For instance, most participants might select choice 2 while the correct answer was 4. Next, item specifications were assigned to test items returning lower than 55% correct responses. At this stage, CEFR level from both grammar and vocabulary profiles was used to classify complexity level of the test items primarily by looking at the blank in each test item as well as its collocations. After assigning CEFR, item specification based on Rogers (2006) was assigned on each test item.

6. Research Results and Discussions

At 55% correct response threshold, there were 12 out of 30 test items. Particularly, the average percentage of test items with correct responses of below 55% was at 35%.

Research Question 1

To a certain extent, CEFR could be used as a bellwether for learners' proficiency judging from Incomplete Sentences responses. In general, the correct and incorrect responses are in line with CEFR level of each test item. Questions with mostly incorrect responses were in test item at B1-C2 level while questions at A1-A2 level were mostly answered correctly. Before we looked into insights from CEFR, we were looking at overview of correct and incorrect percentage on each test item in Table 1.

CEFR Mapping
Table 1 Response Ratio

| Test Item No. | CEFR | Item specification | Mapping between correct answer and most-selected response | Correct Response Ratio |
|---------------|------|--|---|------------------------|
| 101 | B2 | Tense-Aspect | Match | 31% |
| 102 | C1 | Word Choice | Mismatch | 29% |
| 104 | C1 | Word Choice | Match | 38% |
| 107 | B1 | Connecting word | Mismatch | 27% |
| 108 | B1 | Tense-Aspect | Match | 42% |
| 111 | B1 | Word Form | Match | 38% |
| 115 | B2 | Word choice | Match | 54% |
| 116 | A2 | word choice | Match | 54% |
| 117 | N/A* | word choice | Mismatch | 17% |
| 119 | B1 | Tense-Aspect | Mismatch | 25% |
| 121 | C2 | word form | Mismatch | 23% |
| 122 | B2 | Gerunds, Infinitives, and Simple Forms | Mismatch | 21% |
| 123 | N/A | word choice | Match | 35% |
| 125 | B2 | tense-aspect | Mismatch | 23% |
| 126 | B1 | word choice | Match | 40% |
| 127 | B1 | tense-aspect | Match | 42% |
| 128 | B2 | word form | Match | 38% |
| 130 | A2 | connecting word | Match | 52% |

*N/A means no specific CEFR classified neither on grammar nor on vocabulary based on Cambridge University (2010).

All test items were listed in the appendix. Test items in Table 1 were arranged by first presenting marked test items followed by unmarked items. This table was intended to show

discrepancy or congruity between correct answer and the most-selected response from all participants. For instance, in number 102, “mismatch” means most-selected response was “B.” but the correct answer was “A.” and participants who correctly answered this test item was at 29% (15 out of 52). In Table 1, among the selected test items, there were 7 mismatched entries which could be viewed as the most problematic for these test takers.

Table 2 TOEIC Score 600 and above

| Test item No. | CEFR level |
|---------------|------------|
| 102 | C1 |
| 121 | C2 |
| 122 | B2 |
| 125 | B2 |

For the threshold at 550, B1, there were 7 incorrect test items, listed in the appendix, which most of them were in line with CEFR level but there was one test item whose CEFR was not assigned as described in Table 4.

Table 3 TOEIC Score 550

| Test Item No. | CEFR level |
|---------------|------------|
| 102 | C1 |
| 107 | B1 |
| 117 | No CEFR |
| 121 | C2 |
| 122 | B2 |
| 125 | B2 |
| 128 | B2 |

Despite the test item number 107, listed in the appendix, was at B1, test takers at B1 can find it challenging. There were various explanations for this. First, B1 was a broad range in CEFR continuum (Díez-Bedmar, 2018). It was speculated that the question might be at the higher tier of B1, so called B1+ (Council of Europe, 2001). Another possibility was because participants have just reached B1 level and their knowledge has not been fully developed on B1 level. In spite of unknown CEFR for test item number 117 listed in the appendix, its frequency, based on BNC corpus, revealed that the word, “enlarge”, was of low frequency, 2.43 per million words. Next, CEFR distribution of participants with 595 and below (B1-A1) was presented in Table 5.

Table 4 TOEIC Score 595 and below

| Test item No. | CEFR level |
|---------------|------------|
| 101 | B2 |
| 102 | C1 |
| 107 | B1 |

| Test item No. | CEFR level |
|---------------|------------|
| 111 | B1 |
| 117 | No CEFR |
| 119 | B1 |
| 121 | C2 |
| 122 | B2 |
| 125 | B2 |

One clear indicator supporting the argument was straight alignment between increasing number of error test items and CEFR profile of incorrect test items. For instance, the test item number 121, C2 level, was found in every threshold as well as that of test item number 122, B2. What was more, when grouping test takers according to score level error test items at A2 level did not show up. On the other hand, there were some discrepancies despite CEFR holding water in most cases. The error test item number 101 was at B2 as well as those of 122, 125, and 128; therefore, it should have appeared on 550-score threshold. One way to account for this deviation was that the test item number 101 was a test for passive construction in modal verb such as *should be done*, which learners received concentrated tutoring, resulting in being able to evade the distractors. Also, (Foley, 2019) argued that, after careful examination, CEFR equivalence from TOEIC might not necessarily be comparable to criteria shared.

Research Question 2

According to Table 1, there were 4 notable groups of issues that were challenging to test takers, connecting word (2 items), tense-aspect (5 items), word choice (7 items), and word form (4 items). The analysis on this part adopts item specification classification laid out by Rogers (2006) rather than technical linguistic analysis because it appears that terms used in commercial books are considered more accessible to general audience. Using these terms would help practitioners direct their attention to marked item specifications efficiently. Each test item is examined then its item specification is identified. CEFR level of each test item is assigned based on Cambridge University (2010) by looking at deciding factor in determining the correct answer. If the test concerns word choice, CEFR level of vocabulary in the four choices especially the correct answer is considered. If the test item asks about tense-aspect, grammar profile will be taken into consideration.

Tense-aspect

Test item number 101 involved passivization. 15 out of 52 was not mindful of passivity. From CEFR perspective, passive with modal verb was categorized as a B2 level. More to the point, Pienemann and Keßler (2011) argued in light of processability theory that passive construction took many steps to derive from base generation to surface form. This squared with findings claiming that passive construction has gone under various syntactic hierarchical steps from Wanner (2009). Depth of Processing has been discussed in Leow (2015). He argued that depth of processing was a deciding factor in success of SLA. Deeper level of processing required greater cognitive efforts and this affected comprehension mechanism (VanPatten, 2004). In light of cognitive processing theory, Ellis (2006) ascribed the problem to limited instantiation in L2 learners.

In item 108, the signaling device, “as soon as”, was not taken up by 21 out of 52 participants. They chose a tense with progressive form which caused redundancy in the sentence. This incident was not unique to the participants in this study as another study with Thai learners also shared the same result (Hinnon, 2014). The same incident was reported in a longitudinal study by Crosthwaite (2017).

Another major concern was conditionals in test item 119. On CEFR, it was on B1 level and 16 out of 52 chose, “has”, which was A1. Fixating on only language that individuals were well familiar with was a defining feature of low proficiency learner (Saville-Troike & Barto, 2016). Not only were conditionals a problem in theory but this grammar point was a problem for Thai learners (Katip & Gampfer, 2016). This test item was testing on structural meaning of past unreal conditional by providing a clause with perfect modal as a clue. This hint would be taken up only if, at least, learners’ knowledge on tense forms as well as mood has been fully developed. The distinction between theoretical and factual meaning of English verb was an identifying feature of English verb (Leech, 2004). Past infinitive was difficult both non-native and native speakers thanks to the constructed meaning of hypothetical past time. In light of L1 transfer interference hypothesis, Thai did not distinguish between subjunctive and indicative mood, resulting in difficulties in approximating English mood. This speculation was commensurate with findings from Yang and Huang (2004).

Mood was the problem in test item number 125. To correctly answer this question, test takers had to have substantial knowledge on semantic property of “will” in relation to tense (Nuyts & Auwera, 2016). With the shift from expressing mood via verbal conjugation to periphrastic expression, the use of “will” was gaining grounds in English and crossing into functions of other grammatical construction, causing complexity in language use (Fries, 1925). From the result, 19 out of 52 chose “will be completing” instead of “will have completed”, which was the correct answer. These two aspects were problematic for learners because of their complexity. Gvozdanović (2012) proposed that mastery of English perfective aspect is understanding that perfective aspect denoted a certain period of time in the time continuum whereas progressiveness was present in almost every language, supporting a case of L1 transfer interference as found in Deuber (2010). Another possibility of the difficulty was low encountering of the construction in question. From BNC, the frequency of construction, “will + V1” was at 1,000 words per million whereas it was 10 words per million for “will + have + V3”.

In test item 127, subject-verb agreement was an issue especially temporal congruency. The test takers did not take cue in subordinate clause using present tense because they chose past tense in the main clause. 10 out of 52 participants chose, “complained” which was past tense. Tense was often reported as a problem for L2 English learners (Duan, 2011; Listia & Febriyanti, 2020).

Word Choice

Thanks to multifaceted nature of lexis, word choice appeared to be the major source of problem. On test item number 102, the word, “agree”, was chosen by 34 participants out of 52 was at A2 but “comply”, which was the correct answer, was C1. Test takers of low proficiency inclined to opt for choices they knew or they were comfortable with rather than ventured to look for other unknown choices. It seemed that they lacked linguistic resources to verify other viable

options. This action was considered an anchoring effect (Deylamsalehi et al., 2015; Furnham & Boo, 2011).

Item number 104 was testing on words in context. Learners misinterpreted the cue, “no longer in charge...” as a signaling device for the choice, “dismissed”. They disregarded the agent, “Samantha”, in the middle of the sentence. The full sentence was “We have _____ responsibility for the budget to Samantha, so Max is no longer in charge of it.”

For test items number 115 and 117, the so-called egg corn effect, a near homograph—a word with near similar written form but different meaning—was the source of problem. The test takers got confused between “content” and “contend”, which 13 out of 52 participants selected. Meanwhile, the word “expend” was confused with “expand” so 25 out of 52 participants chose the incorrect answer. The correct answer was “enlarge” which was less familiar to learners at A2-B1 level; therefore, they erred on the side of caution. From BNC, the word “enlarge” had 7.95 words per million while “expand” had 42.19 words per million. What was more, participants were mostly from Faculty of Business Administration. It was possible that anchoring effect was at play here. The test item number 116 seemed to fall under the same category. Many test takers associated the word, “anymore”, with a specific position in a sentence. 11 out of 52 chose this word because the question asked for a word at the end of the sentence. From BNC, the frequency of “anymore” at the end of the sentence was 1.55 words per million while the total tokens were 2.81 words per million. In this sense, anchoring effect had to be studied.

Given that TOEIC was based on business English, learners whose discursive knowledge was limited might find the test item difficult. There were 16 out of 52 participants choosing “expansion” instead of “reimbursement”. On CEFR, the former was of B2, resulting in CEFR level not being the key factor. From a dictionary, “reimburse”, which was by extension, “reimbursement”, was described as formal language while “expand” was not (Cambridge University, 2010). To confirm this, another dictionary also labelled “reimburse” in the same fashion (Longman, 2014). This test item was measuring the extent to which learners exposed to ESP, or register. One of the defining features of successful learners of English was an ability to identify not only semantic property of the word but also appropriate word uses in a specific context (Biber & Conrad, 2019).

The last point associating with word choice was a test item on collocational knowledge or lexical priming (Cantos & Almela, 2017; Hoey, 2005). 21 out of 52 selected “employment” because they took cue from “job opening” in the adverbial clause but upon closer examination semantic association between “position” and “application” was more salient compared with “employment”, which was a broad category of the two words in question.

Connecting word

14 out of 52 could correctly answer, “despite”, while 17 out of 52 chose the wrong choice, “even though” on test item number 107. It was possible to argue that the problem lay in syntactic property of these two words. The first one was preposition while the second one was conjunction; therefore, it was what followed that decided which one was the correct answer owing to the fact that they were both used to convey concession. A closer look suggested that CEFR could cast light on this. “Despite” was B1 while “even though” was A1. Learners at lower end might not be familiar with a concession word at B1 level. On generative grammar, this problem

could be attributed to structure raising and finiteness (Radford, 2016). Learners did not reach or acquire finiteness within VP, resulting in mapping error between VP dominated by IP with finiteness, [+fin], and VP under non-finiteness, [-fin].

Despite relative clause being present in L1 Thai syntactic parameter, some test takers chose the wrong answer but after rigorous scrutiny L1 interference might play a role in this problem. Thai used relative pronoun but the form of relative pronoun remained the same regardless of cases or syntactic functions it performed, unlike English which had an array of forms for its relative pronouns, who, which, when, why, whose, and so forth. Learners stick to “that” as an archetypal relative pronoun, resulting in not being sensitive to other forms of relative pronouns (Phoocharoensil & Simargool, 2010). In addition, relative pronoun encountering number also played a part in acquisition as what has been reported in a study with Thai learners (Amornwongpeeti & Pongpairoj, 2014). What was more, the correct answer for this test item was “who” while some test takers chose “whose”. There was a study claiming that “who” had the highest rate of error (33.6%) among other relative pronouns (Kusdianty, 2016).

Word Form

Morphological awareness raising accounted for the error test takers had made on test item number 111. 18 out of 52 selected a citation form of word family, “contract”. One possibility to explain the error was that there was negative L1 transfer in that Thai was not an inflected language. It observed neither inflectional nor derivational affixation. Low level Thai learners inclined to apply the citation form or the form that they were most familiar with across the board. This pattern was found in Chilean university students (Dissington, 2018). The issue on morphology was also listed as a problem in Gardner (2013). There were also findings proposing that premodification was one of learners’ problem when it denoted not only semantic property but also syntactic meaning (Feist, 2011). Two nuanced differences in premodification were descriptive or syntactic. Learners had to discriminate these differences and choose one that was fit for purpose. According to corpus studies (Flowerdew & Forest, 2015), L2 English learners had a problem with signaling nouns.

Another major problem in word form was marginal noun forming suffix. The test item number 121 asked test takers to choose appropriate word form of the word family, “withdraw”. The key difference between “withdrawal” and “withdraw” was that the former was C2 while the latter was C1. Almost more than half of the participants selected the wrong answer, C1. Another source of problem might be because suffix “-al” as a noun confined to only a handful of nouns. Usually, “-al” suffix was an adjective forming suffix. In terms of etymology, B1 learners were familiar with “-ing” as productive noun forming suffix because it was from Germanic branch but only advanced learners would know that “-al”, which came from “-aile” in French—a language from Romance branch—could also be a non-productive noun forming suffix (Hoad, 2003).

Gerunds, Infinitives, and Simple Forms

This item specification is asked in test item number 122. In form-function mapping principle (Walenta, 2019), learners who could not successfully map between form or surface and function would have difficulties in choosing the correct forms. Insights from corpus linguistics proposed that learners whose extended units of meaning or N-gram were limited would have limited command of language because English had a specific word order as well as preposition or

particle use (Cheng, 2012; DeKeyser, 2005; Hunston, 2011; Sinclair, 1996). attributed grammatical error on form to the complexity between form-function relationship in that learners could not identify the difference between “to + V1” and “-ing” when using in conjunction with “forget”. In general, it was not possible to add a form without significant semantic changes. One case in point would be a difference between, “take aim” and “take aim at” or “in the building” and “at the building”.

7. Conclusions

TOEIC part 5 error analyses in conjunction with the CEFR level of each test item could cast light on test takers’ proficiency to a certain extent. Learners at the pre-intermediate to intermediate level have issues with some B1-level questions and errors will be found at the B2 level. Generally, learners tend not to be able to answer questions beyond their CEFR levels. One possible implication is that instructors can ask students, especially in a multilevel class, to answer different CEFR level test items to gauge their students’ proficiency levels and provide appropriate treatment.

On item specification, the most problematic item specification identified in this study is word choice due to its lexicogrammar complexity. For ESP learners predominantly upper elementary and pre-intermediate learners, to be successful in studying ESP, studying vocabulary should precede grammar as it permeates the discipline, ranging from individual word to genre analysis. Instructors should provide the TOEIC word list with words at B2 and above level for B1 learners. After remembering the words, they should be exposed to those words in a context such as newspapers, articles, and commercials. Zahar et al. (2001) also noted the importance of frequency and context in vocabulary acquisition.

On plausible explanation for the errors, learners at this level struggle with grammatical features existing in English but absent in Thai. Not only should instructors point out these language differences but they also provide appropriate materials to the learners such as external reading. Learners who expose to the texts with problematic language features such as gerunds will have a chance to figure out their function in context, resulting in a better understanding of English grammar.

For the testing community especially ESP instructors, one possible implication is that if detailed statistical analysis such as rigorous item analysis, test specification, or item specification is unavailable, item specification and CEFR level could also be used to complement the conventional means.

One of the limitations of this study is diversity in learners’ profile. Future research should compare learners’ problems across proficiency spectrum. For instance, word choice is a major problem for low or high achievers or not. In addition, more samples could help strengthen generalizability of the study.

References

- Amornwongpeeti, A., & Pongpairoj, N. (2014). The Acquisition of English Restrictive and Non-restrictive Relative Clauses by L1 Thai Learners. *NIDA Journal of Language and Communication*, 19(21), 1-54.
- Biber, D., & Conrad, S. (2019). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Cantos, P., & Almela, M. (2017). Colligational effects of collocation: Lexically-conditioned dependencies between modification patterns of the noun cause. In M. Pace-Sigge, & K. J. Patterson, *Lexical Priming: Applications and advances* (pp. 232–250). Amsterdam: John Benjamins.
- Cheng, W. (2012). *Exploring corpus linguistics : language in action*. London: Routledge.
- 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.
- Chomsky, N. (2006). *Language and Mind*. Cambridge: Cambridge University Press.
- Council of Europe. (2001). *Common European framework of reference for languages : learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Crosthwaite, P. (2017). Does EAP writing instruction reduce L2 errors? Evidence from a longitudinal corpus of L2 EAP essays and reports. *International Review of Applied Linguistics in Language Teaching*, 56(3), 315-343.
- DeKeyser, R. (2005). *Grammatical development in language learning*. Malden, Mass.: Blackwell.
- Deuber, D. (2010). Modal Verb Usage at the Interface of English and a Related Creole: A Corpus-based Study of Can/Could and Will/Would in Trinidadian English. *Journal of English Linguistics*, 38(2), 105-142.
- Deylamsalehi, A., Barjasteh, H., & Vaseghi, R. (2015). Anchoring Effects of English Vocabulary Instruction: The Case of Pre-University Students. *Advances in Language and Literary Studies*, 6(3), 138-143.
- Díez-Bedmar, M. B. (2018). Fine-tuning descriptors for CEFR B1 level: insights from learner corpora. *ELT Journal*, 72(2), 199–209.
- Dissington, P. A. (2018). Addressing the Problem of Negative Lexical Transfer Errors in Chilean University Students. *Profile Issues in Teachers' Professional Development*, 20(1), 25-40.
- Duan, M. (2011). A Corpus-based Study of the Misuse of Tenses in the English Composition of Chinese College Students. *English Language Teaching*, 4(4), 173-180.
- Ellis, N. C. (2006). Cognitive Perspectives on SLA: The Associative-Cognitive CREED. *AILA Review*, 19, 100 - 121.
- Feist, J. (2014). *Premodifiers in English: Their Structure and Significance*. New York: Cambridge University Press.
- Flowerdew, J., & Forest, R. W. (2015). *Signalling Nouns in English: A Corpus-Based Discourse Approach*. Cambridge: Cambridge University Press.
- Foley, J. A. (2019). Issues on Assessment using CEFR in the Region. *LEARN Journal: Language Education and Acquisition Research Network Journal*, 12(2), 28-48.

- Fries, C. C. (1925). The Periphrastic Future with Shall and Will in Modern English. *PMLA*, 963-1024.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35-42.
- Gardner, D. (2013). *Exploring Vocabulary: Language in Action*. London: Routledge.
- Gvozdanović, J. (2012). Perfective and Imperfective Aspect. In R. I. Binnick, *The Oxford Handbook of Tense and Aspect* (pp. 781–802). Oxford: Oxford University Press.
- Harada, S. (2016). Effects of the TOEIC course on students' TOEIC scores. *The Bulletin of St. Margaret's*, 48, 1-17.
- Hinnon, A. (2014). Common Errors in English Writing and Suggested Solutions of Thai University Students. *Journal of Humanities and Social Sciences, Khon Kaen University*, 31(2), 165-180.
- Hoad, T. F. (2003). *The Concise Oxford Dictionary of English Etymology*. Oxford: Oxford University Press.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hokuriku University. (2021). *Hokuriku International Exchange Center*. Retrieved from Hokuriku University: <https://iec.hokuriku-u.ac.jp/en/41.html>
- Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.
- Kamijo, T. (2010). TOEIC Reading Comprehension: Analysis of the Texts and Questions, Practical Application and Its Implications. Kyoto: 「立命館高等教育研究」立命館大学教育開発推進機構紀要, (10), 173-185.
- Katip, P., & Gampper, C. (2016). English as a Foreign Language Student Productive Conditional Verb Form Errors in Thailand. *Language Education and Acquisition Research Network*, 9(2), 1-13.
- Kim, M., & Crossley, S. A. (2020). Exploring the Construct Validity of the ECCE: Latent Structure of a CEFR-Based High-Intermediate Level English Language Proficiency Test. *Language Assessment Quarterly*, 18(3), 434-457.
- Kusdianty, S. R. (2016). The Analysis of Students' Error in Using Relative Clause. *Indonesian EFL Journal*, 2(1), 60-70.
- Lawal, A. M., Idemudia, E. S., & Adewale, O. P. (2017). Academic self-confidence effects on test anxiety among Nigerian university students. *Journal of Psychology in Africa*, 27(6), 507-510.
- Lee, J.-Y. (2018). The Use of Test-Taking Strategies and Students' Performances in Answering TOEIC Reading Comprehension Questions. *Taiwan Journal of TESOL*, 15(2), 33-64.
- Leech, G. N. (2004). *Meaning and the English Verb*. London: Routledge.
- Leow, R. P. (2015). *Explicit Learning in the L2 Classroom: A Student-Centered Approach*. New York: Routledge.
- Listia, R., & Febriyanti, E. R. (2020). EFL Learners' Problems in Using Tenses: An Insight for Grammar Teaching. *IJET (Indonesian Journal of English Teaching)*, 9(1), 86–95.
- Longman. (2014). *Longman Dictionary of Contemporary English*. Harlow: Pearson Education Limited.
- Maliwan, S. (2018). TOEIC Preparation Course for Aviation Personnel Development Institute Students. *KASEM BUNDIT JOURNAL*, 19(June), 234–243.

- Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), 1-16.
- Nuyts, J., & Auwera, J. V. (2016). *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press.
- Papol, R., Boonsue, W., Nakcharoen, J., Sriratanaban, S., Silachai, N., & Kham Sri, 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*, 6(2), 239-251.
- Phoocharoensil, S., & Simargool, N. (2010). English Relativization and Learners' Problems. *Pan-Pacific Association of Applied Linguistics*, 14(1), 109-129.
- Pienemann, M., & K bler, J.-U. (2011). *Studying Processability Theory: An Introductory Textbook*. Amsterdam: John Benjamins.
- Powers, D. E., & Powers, A. (2015). The Incremental Contribution of TOEIC Listening, Reading, Speaking, and Writing Tests to Predicting Performance on Real-Life English Language Tasks. *Language Testing*, 32(2), 151–167.
- Radford, A. (2016). *Analysing English Sentences*. Cambridge: Cambridge University Press.
- Richard, J.-P. J. (2020). Investigating CEFR-J Self-Assessment and TOEIC Listening and Reading Scores. *Global Management*, 3, 21-32.
- Robb, T. N., & Ercanbrack, J. (1999). A Study of the Effect of Direct Test Preparation on the TOEIC Scores of Japanese University Students. *TESL-EJ*, 3(4), 25-54.
- Rogers, B. (2006). *Complete Guide to the TOEIC Test*. Florence, KY: Thomson.
- Runnels, J. (2016). Self-Assessment Accuracy: Correlations between Japanese English Learners' Self-Assessment on the CEFR--Japan's Can Do Statements and Scores on the TOEIC . *Taiwan Journal of TESOL*, 13(1), 105-137.
- Saville-Troike, M., & Barto, K. (2016). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- Sinclair, J. (1996). Textus. *The Search for Units of Meaning*, 9(1), 75-106.
- Taguchi, K. (2015). An Exploratory Study on Correlations among Vocabulary Size, Vocabulary Learning Strategy Use, TOEIC Scores and Self-Efficacy. *The Economic Review of Toyo University*, 41(1), 55-67.
- Taladngoan, U., & Esteban, R. H. (2022). Assumptions on Plausible Lexical Distractors in the Redesigned TOEIC Question-Response Listening Test. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 802-829.
- Tannenbaum, R. J., & Wylie, E. C. (2013). *Mapping TOEIC  and TOEIC Bridge™ Test Scores to the Common European Framework of Reference*. Princeton, NJ: Educational Testing Service.
- Trew, G. (2007). *A Teacher's Guide to TOEIC  Listening and Reading Test: Preparing Your Students for Success*. Oxford University Press.
- VanPatten, B. (2004). *Processing Instruction: Theory, Research, and Commentary*. New York: Routledge.
- Walenta, M. (2019). *Form-Function Mapping in Content-Based Language Teaching*. Cham: Springer International Publishing.
- Wang, S. (2015). An Empirical Study on the Role of Vocabulary Knowledge in EFL Listening Comprehension. *Theory and Practice in Language Studies*, 5(5), 989-995.

- Wangmanee, P., & Vongtangswad, M. (2022). Effectiveness of the Potential Promoting Strategies on the TOEIC Test of English Major Students, Chiang Rai Rajabhat University. *Journal of Human and Society, Sisaket Rajabhat University*, 6(1), 40-60.
- Wanner, A. (2009). *Deconstructing the English Passive*. Berlin: De Gruyter Mouton.
- Wei, Y., & Low, A. C. (2017). Monitoring Score Change Patterns to Support TOEIC® Listening and Reading Test Quality. *ETS Research Report, 2017(1)*, 1-22.
- Wilson, K. M., Komarakul Na Nagara, S., & Woodhead, R. (2004). TOEIC®/LPI RELATIONSHIPS IN ACADEMIC AND EMPLOYMENT CONTEXTS IN THAILAND. *ETS Reserach Report Series, 2004(1)*, 1-33.
- Wu, H.-l. (2013). Implementation of an elite sophomore English class at a university in Taiwan. *Hsiuping Journal of Humanities and Social Sciences*, 21(2013), 21-30.
- Yang, S., & Huang, Y. Y. (2004). The Impact of the Absence of Grammatical Tense in L1 on the Acquisition of the Tense-Aspect System in L2. *International Review of Applied Linguistics in Language Teaching*, 42(1), 49-70.
- Zahar, R., Cobb, T., & Spada, a. N. (2001). Acquiring Vocabulary through Reading: Effects of Frequency and Contextual Richness. *The Canadian Modern Language Review*, 57(4), 541-572.