Which Factor–Listening, Reading, or Background–is the Best Predictor of CEFR–Based Scores?

Kasin Janjaroongpak
College of General Education and Languages
Thai–Nichi Institute of Technology

Received : Jan 7, 2022
Revised : Feb 28, 2022
Accepted : Mar 31, 2022

Abstract

TOEIC has been used in many countries around the world especially in Asian countries and there are a sizeable number of studies on TOEIC. Two ongoing topics in TOEIC research have been (1) indicative power between reading and listening sections in relation to the total (2) indicative power between test takers' background and TOEIC scores. To address the first point, this study adapts the research design by Park et al. (2020). The key difference is that (1) this study categorizes scores by using the Common European Framework of Reference for Languages (CEFR) level rather than intervals and (2) the scores used are official scores issued by Educational Testing Service (ETS) not scores derived from classroom test. For the second question, this study compares the result between mean scores from this study with the mean scores reported by ETS (2020a). Despite different sample sizes, the score distribution as well as correlation among listening, reading, and total between the two studies are relatively the same except the correlation between listening and reading sections. Reading is the indicative factor for A1-B2 learners' scores while the importance of listening emerges when the score reaches B2 level. On background analysis, test taker profile in the present study, mostly, appears to be in line with the report from ETS (2020a) on years of English study, academic major, and daily English use requirement, suggesting that background can serve as a useful indicator of TOEIC scores. Discussion and pedagogical implications are provided.

Keywords: CEFR, TOEIC, reading, listening, correlation

Introduction

The Test of English for International Communication (TOEIC) has been widely used for decades globally thanks to both its reliability and practicality especially in business. Companies in non-English speaking countries have been using TOEIC as one of the main criteria for recruitment. As a result, job seekers, both students and currently-employed skill workers, are expected to submit TOEIC test results to companies they are applying for. This puts pressure on higher educational institutes to take TOEIC into consideration. In Japan, the trend of examinees has been rising (Institute for International Business Communication, 2021). TOEIC has been designed to measure English in real-life working environments (Powers & Powers, 2015). TOEIC has been integrated into the curriculum in many countries where TOEIC has enjoyed its presence (Nam, 2016). In Taiwan, TOEIC has been used as an exit exam (Hsieh, 2017). Oliveri and Tannenbaum (2017) revealed that TOEIC scores are one of factors in recruitment decision-making.

One of the earliest attempts has been from Wilson (1989). Wilson (2000) conducted exploratory factor analysis on both listening and reading components of TOEIC. Wei and Low (2017) reported on score pattern changes. Perception of test takers over TOEIC was presented in Powers et al. (2008). Schmidgall (2017) discussed the actions taken by TOEIC to emphasize its argumentation over its validity.

In Thailand, there has been a study on content validity of TOEIC reading section (Imsa-ard, 2019). In'nami and Koizumi (2012) in Japan tested directional score trend in reading and listening sections. According to Zhang (2006), there seems to be a divergence between reading scores and listening scores. This study included more than 40,000 Japanese and Korean examinees. Another review of reading section proposed that there should be a robust examination on TOEIC reading section to promote greater fairness (Suzuki & Daza, 2004).

Despite a sizeable number of pieces of research on TOEIC validity, a study based on official TOEIC scores in Thailand, as far as I am concerned, has been limited. Apart from papers sponsored by ETS, TOEIC test writer, most research–both large- and small-scale studies–has used TOEIC scores obtained by administering a classroom examination in their analyses. This study tries to provide analysis whose result is derived from a credible source. In theory, it is possible that a mock test is as reliable as the real one (Furwana, 2019), but the outcome might be different thanks to factors involved (Gamer, 2012).

In Korean (Jee & Lee, 2009; Park et al., 2020), studies revealed that it was possible to predict, to a certain extent, total score from either listening score or reading score based on score level. Conducting this study with the Thai learners will explore whether such correlation exists in Thai learners and the result would help Thai instructors to allocate their resources efficiently for Thai learners.

For studies on background, ETS (2020a) provided mean scores based on a large collection of data but the mean scores of other learners might be the same or different from Thai learners. If a similar study is carried out, it will provide insights into the correlation between learners' background and test scores for Thai learners. This would help both administrators and practitioners create a more effective curriculum.

Objective

The objective of this study is to shed light on the correlation between TOEIC reading section, TOEIC listening section, and TOEIC total scores in intermediate Thai learners from a higher education institute. After gaining insight into contributing factors in TOEIC scores internally, another purpose of this study is to explore to what extent learners' background can predict learners' TOEIC scores.

Relationship between Language Skills in Standardized Test

The four skills in language have been linked, by and large, as they are components of language. The main question is the relationship among them.

In Relation to Listening

One of such research is Bozorgian (2012). He found that listening closely links with reading (r = .735) but less so with writing (r =.643) and speaking (r = .654). Also, the link between listening and overall performance is more meaningful (r = .887). The importance of listening skill seems to far outweigh other language domains. To some scholars, listening comprehension is a skill that should not be left behind (Hogan et al., 2014).

In Relation to Reading

While some focus researchers on the crucial role of listening, others look at relations between reading and listening. Hedrick and Cunningham (1995) argued that reading ability could be enhanced by listening to the text. Others have argued that listening and reading should be dealt with differently because of their fundamental differences despite some common features (Lund, 1995). Diakidoy et al. (2005) revealed that differences between reading and listening would first place them apart and they would move closer as the time goes by, improving the skills of students, but text type could be a factor in disparity between reading and listening. To establish this, Wolf et al. (2019) teased out relations between reading and listening comprehension.

TOEIC Research in Universities

In Korea, Lee and Jin (2009) have conducted a study on English-only classes to improve students' TOEIC scores and revealed that the English-only class focusing on active learning such as presentation and role playing had a meaningful positive impact on students' scores. Meanwhile, Ha (2012) held that promoting a class focusing on reading rather than listening was more beneficial to students compared with listening class alone. A study in Japan by Harada (2016) found that balanced teaching produced the best result in TOEIC scores. He also points out that vocabulary should be the focus of teaching. This finding supports another study (Komatsu, 2015). In Thailand, a method-oriented approach to the teaching of TOEIC has been conducted (Lertcharoenwanich, 2020; Suvarnaphaet & Desgres, 2017).

CEFR and Standardized Tests

A study of CEFR and standardized tests was undertaken by Wudthayagorn (2018). She mapped and analyzed various standardized tests into the CU–TEP, a test developed by a university in Thailand. The main concern in creating a comparison table has been the point at which the threshold should lie. Nakanitanon (2021) has taken on the idea of cut-off scores and explored them further based on FRELE-TH, a test developed by Chiang Mai Rajabhat University in Thailand. The method used was Yes/No Angoff method. Apart from CU–TEP, there was another attempt to use CEFR in the same light (Athiworakun et al., 2018). In a paper (Waluyo, 2019), the result illustrated that a large number of participants were at A1-A2 level by comparing the WU–TEP test results with CEFR. WU–TEP was an in-house test developed by Walailak University in Thailand. By comparing WU–TEP and CEFR, both practitioners and learners would gain accurate understanding of learners' proficiency. For TOEIC, a study to establish a link between TOEIC and CEFR was reported in Tannenbaum and Wylie (2013).

To establish decisive factors contributing to success in standardized test, many researchers have conducted correlational studies on either relation among components of the test or relation between external factors and the total scores of the test or subset of the test. In Korea, there was a large-scale internal correlation study on TOEIC scores (Park et al., 2020). Pearson Correlation was used to determine the relation by dividing scores in 100-interval fashion. Two studies from Japan indicated that CEFR could be used to tease out the correlation of TOEIC scores. The first study, with 57 English-major participants, showed that there was a correlation between CEFR-J and listening score but not reading (listening r = .23; reading r = -.14) (Runnels, 2016). The second study, with 54 non-English-major participants, found, inconclusively, a greater correlation (listening r = .29; reading r = .50) on both ends (Richard, 2020). Clearly, CEFR was another possible means to examine the correlation among TOEIC scores.

Regardless of types of tests, language proficiency achievement is an interplay of various factors (Alyousif & Alsuhaibani, 2021; Cheng & Lee, 2018; El-Omari, 2016; Gu, 2015; Lehnert et al., 2018; Shi, 2021). In Korea, a study examined the correlation between learning strategies and years of studying English to predict English proficiency (Magno, 2010). The study suggested that time spent in formal English classes and compensation strategies were a significant factor in learner's English language ability. Another research investigated the correlation between background and test performance (Manna & Yoo, 2015).

Studies on Skill Predictors

Reading and Listening Predictor

For reading, recently, several studies have uncovered factors influencing reading skill: anxiety (Mardianti et al., 2021), self-esteem (Rosalina & Nasrullah, 2019), word form (Aziz et al., 2019), and vocabulary (Manihuruk, 2020). Thanks to the challenges in listening research, fewer correlational studies have been undertaken: personal profile (Kim & Petscher, 2021), factor analysis (Golen, 1990), and vocabulary (Hwang & Cabell, 2021).

Despite reading and listening both being receptive skills, relations between the two skills are debatable. Wong (2021) argued that there was a positive correlation between reading and listening while Sok et al. (2021) found that they both shared predictors of aptitude and phonological working memory, skill-specific predictor, and motivation. Disconnect between reading and listening was also reported in previous studies (Gauthier, 1988; Olejnik, 1978).

Many factor analysis studies on reading and listening have explored their relations to factors in a context of skill in general but factors might be different when it comes to reading and listening as a component of score in a standardized test.

TOEIC Score Predictor

In correlational research, a few studies on correlations between major and non-major students' performance on TOEIC have been conducted. The studies usually came in the form of attitudinal studies and did not investigate the score distribution. Robb and Ercanbrack (1999) compared the result of direct test preparation course on English-major students and non-English-

major students and argued that TOEIC reading course was beneficial for non-majors but not for the counterparts. Hong and Phan (2020) studied non-major students' self-efficacy beliefs and TOEIC performance in Vietnam. Despite a positive result, the study provided total TOEIC scores rather than score distribution and used a mock test rather than an official score report. Another perceptual research employing non-major students' questionnaire responses pointed out that the ramification of TOEIC being an exit exam was not recognized as strong but the participants were concerned with the shift towards teaching for testing rather than teaching for learning (Nguyen & Gu, 2020). This potential negative attitude derived from non-English major students on TOEIC was also voiced in a study from Vietnam (Phan et al., 2019). Not only were the scores not improved but the participants also were finding the test unfavorable. The significance of attitude towards the test and the test performance of non-English major students was further confirmed by a study from Thailand (Puengpipattrakul et al., 2007). This supported an earlier study (Wilson et al., 2005). In Korea, a study also cast the same light on relations between background and TOEIC (Shin & Lee, 2012). This study sifted through the relationship between cultural integration by meaning through exposure to American media and participants' performance on TOEIC reading part. In Taiwan, a study noted an interesting finding in that students with a business background outperformed their technology-oriented peers in the TOEIC listening comprehension test (Huang et al., 2015). In Indonesia, a study found that learning habit strongly correlated with the TOEIC performance by using Pearson Product-Moment Correlation (Zakaria et al., 2017).

Background and Proficiency

For reading, studies on relations between reading and learners' background, as far as I am concerned, cover various fronts: role of background (Smith et al., 2021), word use (Wood et al., 2021), working memory (Shin et al., 2019), no correlation between background and reading (Roohani et al., 2017), and positive correlation between background and reading (Al-Noori, 2014). On listening, little research has been conducted to discern the relation between listening and learners' background: culture (Al-khresheh, 2020), role of linguistic knowledge (Long, 1990), and positive relation between background and listening challenges (Hadist et al., 2022; Hasan et al., 2017; Sadighi & Zare, 2006).

To sum up, there have been various studies on relations among language skills but only a limited number of pieces of research are looking into the relationship among skills in standardized tests as well as the relation between each skill and the total score. Previous relevant studies abroad focused on score relation by dividing scores into 100-score interval. This study, in Thailand, will shed light on score relation categorized by CEFR. The reason why this study compares the result with Park et al. (2020) is because both Korea and Thailand are considered expanding circle (Bolton & Kachru, 2006). In terms of environment, exposure, and necessity of English in daily life, the two countries are comparable. For age, participants in both studies are first-year and second-year undergraduate students. This study is an attempt to explore such relations in Thai test takers.

Research questions

- Is the correlation among listening comprehension, reading comprehension, and TOEIC total scores similar to the study from Korea (Park et al., 2020) should the scores be grouped by CEFR?
- Is the mean score of participants grouped by their background: academic major, daily English use requirement, and most frequent used language skill similar to the report from TOEIC (ETS, 2020a)?

Methodology

To make this study comparable with Park et al. (2020), the data collection and analysis set forth by Park et al. (2020) were adopted. The participants in this study came from students enrolling in a TOEIC preparation compulsory course in a university in Thailand. Simple sampling was used. Owing to the fact that each faculty had its unique characteristics, participants from each faculty were expected to come from different backgrounds. For instance, students from Faculty of Engineering were generally less familiar with English compared with their peers from Faculty of Information Technology. The former group involved with physical activities while the latter was spending much time coding in English. Forty-six students were from a business-related faculty, 24 were from science-oriented faculty, and the other 31 were from an IT faculty. All of them were second year students. On language proficiency, based on their TOEIC scores, the participants were mostly from A2 to B2 level. The exempt consent was applied thanks to the score submission being a part of matriculation assessment. The scores were collected during the semester. All personal confidential details were removed and deleted before the analysis took place. There were 101 scores submitted and they were grouped by their respective CEFR based on TOEIC-CEFR mapping provided (ETS, 2020b). After obtaining the scores, descriptive statistics were calculated, notably, means and the standard deviations for the listening and reading sections to scrutinize relationships between listening and reading sections and the total TOEIC scores. Then, Pearson's correlation, available in Microsoft Excel, was used to gauge relationships between the scores. Pearson's correlation has been widely applied in studies focusing on linear relational strength among factors in question (Ha, 2012). The value, in R, was from -1 to 1. In this study, it was utilized to uncover the relationships between the listening section and the total, the reading section and the total, and the listening section and the reading section. First, scatter plot was created to confirm linearity among the sets of components followed by correlation coefficient. The correlation was stronger when the two variables were coming close to a straight line. Positive correlation was observed when one variable was increased, the other variable increased at the same time we saw negative correlation when one variable was increased but the other variable decreased. In this vein, the correlation coefficient at 1 was an absolute positive correlation while -1 was an absolute negative correlation.

Result and Discussion

Research Question 1

Is the correlation among listening comprehension, reading comprehension, and TOEIC total scores similar to the study from Korea (Park et al., 2020) should the scores be grouped by CEFR?

Table 1
TOEIC Scores Descriptive Statistics

Skills	This study			Korean study (Park et al., 2020) with adaptation		
	Mean	Standard Deviation	N	Mean	Standard Deviation	N
Listening	324.950	93.132	101	324.5242	69.02404	11,328
Reading	269.405	97.763	101	292.1941	72.97826	11,328
Total	594.356	180.356	101	616.7183	133.06549	11,328

Descriptive statistics processed by Microsoft Excel under the function data analysis provided other basic statistical analysis such as standard error, median, and mode but in this study only mean, SD, and count or total would be presented to replicate the analytical steps taken by Park et al. (2020).

Table 2
TOEIC Scores Pearson Correlations

Skills	This study			Korean study (Park et al., 2020) with adaptation		
Listening	1	.785**	.941**	1	.756**	.933**
Reading	.785**	1	.947**	.756**	1	.941**
Total	.941**	.947**	1	.933**	.941**	1
N	101	101	101	11,328	11,328	11,328

Despite the difference in participant number, the overall Pearson correlations showed a similar pattern between the two studies with minor differences.

Table 3
Thailand Mean Scores (ETS, 2020a)

Skills	Mean	SD
Listening	279	105
Reading	206	102
Total	485	200

In the Korean study, the regional mean scores were used but this study would not resort to the regional mean scores because they remotely related to the case of Thailand. From Table 3 (ETS, 2020a), the mean scores were at 279 in listening, 206 in reading, and 485 in total were far lower than scores distribution of participants in this study because of, arguably, sample differences and field of studies.

Table 4

Mean Scores Comparison (Listening, Reading, and Total) by CEFR

Skills	B2-C1 (N = 8)	B1-B2 (N = 24)	A2-B1 (N = 26)
Listening	460.625	347.391	229.808
Reading	406.250	318.478	187.692
Total	866.875	665.870	417.500
Gap	54.375	28.913	42.115

Table 5

Mean scores of the Listening, the Reading, Total, and the score gap between the Listening and the Reading

Skills	Score						
	900 (N = 144)	800 (N = 902)	700 (N = 1,978)	600 (N = 2,990)	500 (N = 3,027)	400 (N = 1,672)	(N = 615)
Listening	467.11	432.16	388.98	342.64	293.47	247.45	200.6
Reading	460.41	410.5	358.86	308.67	260.78	210.53	161.39
Total	927.53	842.67	747.84	651.31	554.25	457.98	362
Gap	6.7	21.66	30.11	33.96	32.69	36.91	39.21

The samples used in Table 4 were the samples after exclusion. The exclusion took place to select only samples clearing CEFR threshold. CEFR A2 must achieve at least 110 in listening and 115 in reading. Samples with listening scores at 110 but reading scores at 95, for instance, would be excluded. As a result, only 58 entries were processed. The first column, A2-B1, appeared to be in line with the pattern found in Table 5 in Park et al. (2020). The gap was approximately at 40 points and the trend remained constant in B1-B2, 785 scores equivalent, in that the gap was in the range of 20, compared with 800 band intervals. The discrepancy lay in the highest level, B2-C1. The Korean study found that there was only a slight difference between listening and reading, 6.70, while this study reported 54.375 points. Different number of participants might be a factor because this study included only 8 entries while the other took in 144 samples. On nature of score distribution, both studies showed that participants scored more on listening compared with reading.

Table 6
R values comparison between Listening & Total and Reading & Total

Skills & Total	B2-C1	B1-B2	A2-B1
Listening & Total	.883	.665	.745
Reading & Total	.778	.769	.854

Table 7
R values between the Listening & Total and the Reading & Total (Park et al., 2020)

Correlation	Score						
	Over 900	Over 800	Over 700	Over 600	Over 500	Over 400	Over 300
Listening & Total	.642**	.580**	.509**	.505**	.441**	.450**	.572**
Sig.(2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Reading & Total	.636**	.684**	.565**	.476**	.490**	.482**	.510**
Sig.(2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**Correlation is significant at the 0.01 level (2-tailed)

Next, the *r* values between listening-total and reading-total were calculated by the function correlation in data analysis from Microsoft Excel. Statistical interpretation referred to the guideline provided by Evans (1996). *R* value at lower end, 0.00-.19, was considered weak while the other end, .60-.79, was strong, and very strong for .80-1.0. Despite the different values, the overall trend of the correlations was relatively similar to the counterpart study. The B2-C1 band reported the *r* value at .883 for listening and .778 for reading with total scores. Despite the similar pattern in that listening score was closer than reading, the figure reported was different (.642 for listening and .636 for reading in 900 band score).

Table 8
Listening-Reading r values

Correlation	B2-C1	B1-B2	A2-B1
Listening and Reading	.392	.034	.289

Table 9
R values between the Reading & the Listening (Park et al., 2020)

Reading and listening	Score						
	Over 900	Over 800	Over 700	Over 600	Over 500	Over 400	Over 300
Correlation	−.184**	−.198**	−.423**	−.519**	−.566**	−.566**	−.414**
Sig.(2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**Correlation is significant at the 0.01 level (2-tailed)

While the counterpart study found a moderate negative correlation in most bands except 800 and 900 intervals, this study found a weak positive correlation except for B1-B2 band which reported very weak correlation. In Jee and Lee (2009), there was a correlation in group with mean score of 705 at .50 *r* value but a weak correlation in lower band score at .20 and negative correlation at -.28 for the group of mean score at 94. This research was conducted with 599 participants. The similar point was that in the higher band score there was a strong to relatively strong correlation (.50 and .392).

To summarize, despite some differences, it appears that the correlation between reading, listening, and the total TOEIC scores of this study is similar to that of Park et al. (2020). On mean scores, despite the different sample size between the two studies, the mean scores are relatively comparable especially in listening skill. This implies that listening skill of Thai learners in the samples is on par with Korean learners in the other study. One way to interpret this is that Thai learners are, surprisingly, keen on listening on par with learners from developed countries. It is possible to argue that access to internet-mediated media is at play. Reading is the skill that Thai learners need to catch up. On correlation, given the different sample size, the correlation between the two studies shows the same trend. One possible application is that for correlation instead of aiming for a large sample size it is possible, to a certain extent, to use 100-sample study to bolster the claim. Also, the agreement between the two studies lent support to the argument that reading is slightly more indicative of the total TOEIC scores. Test takers who aim to excel at TOEIC are encouraged to prioritize reading over listening, considering time and energy one has to invest. Compared with mean scores of Thailand, learners in this study have higher mean scores on both listening and reading. One possible interpretation is that participants are from a university championing language learning, resulting in taking on many language-oriented undergraduate students.

Motivation is instrumental in achieving language learning success. On gap comparison, the lower level, A2-B2, shows the same trend as that of the other study in that the gap is wider on the lower end and it is narrowing as the score moves higher but the difference lies in scores at the advanced level. The fact that the gap, with listening score higher than reading score, is around 40 scores both A2-B1 and B1-B2 might be explained by the TOEIC scoring system in that it takes fewer correct answers to score points for listening compared with reading. For instance, 1 correct answer in listening might score 5 points while it takes 3 correct answers in reading to score 5 points. Though different score range has different requirements, listening, by and large, is an easier part compared with reading. We should expect a gap to be wider exclusively at B2-C1 because of vocabulary requirement. Milton and Alexiou (2009) proposed that a learner requires 3,250 words to reach B1 but it takes 4,500 words to attain C1, which is almost two times the vocabulary size of B1, resulting in a wider gap between those who just reached B2 and full C1. More research is called for to establish this. On correlation between listening and total scores and reading and total scores, this study suggests that those who aim at A2-B1 level should prioritize reading over listening because it correlates with the total score more than listening and they should continue to do so on B1-B2 level. Then, they should move to listening when they are on B2-C1 level by the same token.

Despite different *r* value, the Korean study shows that the correlation is stronger in listening on over 900 scores. Also, the scores on over 700 scores correlate strongly with reading. Both studies recommend starting with studying for reading and listening later. One key advantage of categorizing into CEFR rather than 100-score interval is that it describes a more accurate picture of learner's interlanguage. One notable example would be the indeterminate nature of *r* value on over-300 to over-400 scores. The fact that these two levels produce relatively the same *r* value is because they both fall under A2. This rings true until mid over-500 scores–550 scores which is the start of B1. Also, by grouping as B1-B2, we can see that reading will be a deciding factor until the upper-intermediate level, B2 or 785 TOEIC scores. Therefore, learners at B1 should pay more attention to reading. At the same time, over-500 scores gives little direction for learners because the *r* value is at .49 for reading and .44 for listening meaning that neither listening nor reading is indicative of the total scores. Upon a closer look, it reveals that the clear indication emerges when the score reaches over-800 level at .58 and .68 respectively, suggesting that reading is a leading factor. This interpretation is in line with .66 and .76 respectively reported in this study. On the correlation between listening and reading scores, both studies reveal that the correlation between the two skills are negligible. This implies that both reading and listening need to be developed individually–killing two birds with one stone is not applied.

Research Question 2

Is the mean score of participants grouped by their background: academic major, daily English use requirement, and most frequent used language skill similar to the report from TOEIC (ETS, 2020a)?

To make the data set comparable, there were three domains to be investigated in this study: academic major, score by daily English Use Requirement, and Score by Most Frequent Used Language Skill. The following comparison will use all 101 participants because I wanted to include data as much as possible to compare with ETS report and mean comparison gave a proper overview.

Table 10

Mean Score Comparison Based on Academic Major

Faculty	EST Report						Gap	
	Listening	Reading	CEFR	Listening	Reading	CEFR	Listening	Reading
Business-related	309	259	A2	334	279	B1	25	20
Engineering	330	269	A2	316	258	A2	14	11
Sciences	345	284	B1	325	272	A2	20	12

There were fairly small differences between participant mean scores and the scores from the report. Participant's Business-related group underperformed its counterparts around 20 points per skill while the other groups from participant domain relatively outperformed its peers around 10 points. One Sample T-test was conducted on each pair and found no statistical significance. As a result, it was possible to interpret that the mean scores of participants were in line with global

standard of their peers in the professions. Employers had specific expectations on TOEIC in their respective field of professions (Puengpipattrakul et al., 2007).

Table 11
Mean Score Comparison Based on Daily English Use Requirement

Percentage	EST Report						Gap	
	Listening	Reading	CEFR	Listening	Reading	CEFR	Listening	Reading
1%-10%	329	269	A2	321	262	A2	8	7
11%-20%	309	259	A2	344	284	B1	35	25
21%-50%	345	284	B1	363	302	B1	18	18

The criteria to decide which topic to be included in the questionnaire was set out by ETS (2020a) including the choices available in the questionnaire, from none to 100%. The necessity to use language effectively or the drive to get message across successfully was particularly vital in survival. Therefore, Daily English Use Requirement should be incorporated. In Table 11, participants were divided into three groups to make the data comparable with the report. The first group was categorized as low English use requirement group because the main source of Daily English Use was reading textbook when they were studying in a few classes. The next group was placed into 11%-20% group because the participants had, at least, one class with a textbook in English every second day. The last group was assigned into 21%-50% because the textbooks used in almost every subject studied were in English. In particular, activities promoting listening such as English-only classes might help boost the listening section (Lee & Jin, 2009).

Table 12
Mean Score Comparison Based on Most Frequently Used Skill Language

Skill	EST Report						Gap	
	Listening	Reading	CEFR	Listening	Reading	CEFR	Listening	Reading
Four skills	396	337	B1	372	311	B1	24	26
Listening oriented	287	240	A2	319	255	A2	32	15
Reading oriented	338	277	B1	319	271	A2	19	6

The first group, reading, is arguably the proxy for learning style practiced in Thailand and is similar to the score distribution in the report group. It might be interpreted that the learning outcome of learners whose primary learning style is reading should fall into approximately 600 scores. Despite reading-oriented class, reading score precedes listening scores in both groups. The same pattern is shown in Ha (2012). For listening group, though the differences are not statistically significant, they remind the participants in this bracket that they should work on their reading scores to improve the overall score, provided that their mean scores fall into 500-interval, as

suggested in Park et al. (2020). The all-around group has the best overall scores and score distribution is the same in both groups, with the participant group outperforming its counterpart. This indicates that teaching English in a skill-balanced manner yields the highest overall TOEIC score, which is further supported by Harada (2016).

Table 13
Credit Comparison

Credit	Listening	Reading	CEFR
English-oriented (21 credits)	396	337	B1
Non-English-oriented (9 credits)	318	262	A2

After running mean comparison analysis, the report shows that the difference between the two groups, both listening and reading, is statistically significant. By and large, the scores rise in tandem with the number of credits. Wei (2013) pointed out that some background domains are predictive of TOEIC scores and the domain of English study time at 4-6 years is a contributing factor (variance explained at 46.21% for listening and 34.09% for reading) on global level but when the researcher looks into individual details by using random coefficient to tease out the correlation, the variance is reported at approximately 4%. It appears that number of years of studying English plays a useful role in TOEIC scores but it is not necessarily the unique one.

To summarize, on gap between mean score of participants in this study and the report by ETS, the score is relatively the same, 25-point difference. This means that language proficiency of this group is in line with Thai TOEIC mean scores. Comparison in this manner could be an informed bellwether for universities in Thailand to use as one of the key performance indicators for English proficiency should business English be their language goal. However, the score requirement might be subject to various reasons such as level of position, specialization, or type of company. On Daily English Use Requirement, the rule of thumb is the higher the requirement level, the higher the score but the participants in this report (11%-20%) show a disconnect in that their listening and reading scores are lower than that of 1%-10% bracket. It appears that these learners have not reached their end stage of language learning in university while the participants from the ETS report are from working people. Once the students in 11%-20% bracket reach their fourth year, the score should be improved. More research is required. On Most Frequently Used Skill Language, two key points emerge. First, despite reading being used most, reading score lags behind listening scores on both participants in this study and the participants in the ETS report. This suggests that frequency alone is not necessarily a determining factor in improving reading. Therefore, to improve reading, instructors and administrators need to explore other factors such as vocabulary, grammar, or metacognition rather than adding more hours of reading classes. Second, the four-skill group outperforms both listening-oriented and reading-oriented groups across the board. This means that learners who aim for high scores should attend language class that incorporate four skills rather than skill-specific course.

Pedagogical Implications

Between skills and background, reading skill is the best predictor of CEFR-based TOEIC scores. First, reading score correlates closer to the total TOEIC scores compared with listening score or background. Manna and Yoo (2015) reported the effect size of approximately 0.3 for many background factors. In addition, reading score is a laggard factor in every score range meaning that if a test taker wants to achieve a high score, reading score must be first dealt with. Second, background information might give a false impression and should be considered in relation to reading skill.

The listening and reading scores on the second research question come from participants in this study who turned in their TOEIC scores issued by ETS and the mean scores obtained from ETS (2020a). Two key factors in developing listening and reading scores are, from this study, the number of skills frequently used and the level of daily English requirement.

Despite its limited predictiveness, background could offer some quick cues to both practitioners and professionals. For instance, a quick round of field question on skills frequently used at the beginning of a class might help instructors grasp their learners' ability.

This study shows that the score distribution grouped by CEFR, in general, squares with the score distribution by 100-interval but it helps classify test takers in a systemic and linguistic manner. The three stages of development, A1, A2, B1, B2, C1 and C2 in CEFR (Council of Europe, 2020) are easier for both practitioners and test takers to develop specific areas for improvement. The way to improve reading and listening skills concurrently is by channeling strenuous efforts in learning a shared aspect between the two skills, which is vocabulary (Wolf et al., 2019). For teaching language skills, all four skills should be included as they produce the highest mean scores. If a class is limited to only one skill, reading-concentrated class is recommended (Wei & Low, 2017).

Grouping learners based on their proficiency appears to be beneficial because the cohort has relatively the same problem in their learning journey. However, the fine line should be trodden carefully when it comes to division (Hallinan et al., 2003; Jones & Gerig, 1994; Kiss, 2017; Kurian & Mekoth, 2021; Magableh & Abdullah, 2021; Mazenod, et al., 2019; McGillicuddy, 2021).

Lastly, credits or hours of studying English, in this study, are a significant factor in TOEIC scores. Despite criticism on TOEIC listening-reading scores being modality-specific, some studies suggest that the scores in question suffice as evidence for requesting credit waiver (Hahta et al., 2000; In'nami & Koizumi, 2017; Powers, 2013).

Author

Kasin Janjaroongpak is an English Department Lecturer at the College of General Education and Languages, Thai–Nichi Institute of Technology. His research interests include English as a Foreign Language, English Linguistics, and Corpus Linguistics. He can be contacted at kasin@tni.ac.th

References

Al-Khresheh, M. H. (2020). The impact of cultural background on listening comprehension of Saudi EFL students. *Arab World English Journal*, *11*(3), 349-371.

Al-Noori, B. S. (2014). The effect of the reader's background on reading comprehension performance. *Advances in Language and Literary Studies*, *5*(6), 194-203.

Alyousif, R., & Alsuhaibani, Z. (2021). English language learning demotivating factors for Saudi high school EFL students. *English Language Teaching*, *14*(8), 29-39.

Athiworakun, C., Vathanalaoha, K., Thongprayoon, T., Rajprasit, K., & Yaemtui, W. (2018). SWU-SET as a CEFR standardized English test. *Journal of Language Teaching and Research*, *9*(2), 261-267.

Aziz, Z. A., Daud, B., & Ismar, R. T. (2019). Morphological awareness and its correlation with EFL reading comprehension of senior high school students. *TEFLIN Journal: A Publication on the Teaching and Learning of English*, *30*(1), 121-136.

Bolton, K., & Kachru, B. B. (2006). *World Englishes: Critical concepts in linguistics.* Taylor & Francis.

Bozorgian, H. (2012). The relationship between listening and other language skills in international English language testing system. *Theory and Practice in Language Studies*, *2*(4), 657-663.

Cheng, A., & Lee, C. (2018). Factors affecting tertiary English learners' persistence in the self-directed language learning journey. *System*, *76*, 170-182.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume.* Council of Europe Publishing. Retrieved from www.coe.int/lang-cefr

Diakidoy, I.-A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The belationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, *26*(1), 55-80.

Educational Testing Service. (2020a). *Report on test takers worldwide*. Retrieved from https://www.ets.org/s/toeic/pdf/2020-report-on-test-takers-worldwide.pdf

Educational Testing Service. (2020b). *TOEIC® Listening and Reading Test Scores and The CEFR Levels.* Retrieved from https://etswebsiteprod.cdn.prismic.io/etswebsiteprod/5290f06b-e1cf-4254-8776-4a7c757f5252_TOEIC-Listening-and-Reading-test-CEFR-correlation-table.pdf

El-Omari, A. H. (2016). Factors affecting students' achievement in English language learning. *Journal of Educational and Social Research*, *6*(2), 9-18.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences.* Brooks/Cole Publishing Co.

Furwana, D. (2019). Validity and reliability of teacher-made English summative test at second grade of Vocational High School 2 Palopo. *Language Circle: Journal of Language and Literature*, 111-119.

Gamer, M. (2012). Validity of the Concealed Information Test in realistic mock crime scenarios: comment on Bradley, Malik, and Cullen. *Perceptual and Motor Skills*, 427-431.

Gauthier, L. R. (1988). A study of listening comprehension as a predictor of reading gains. *Reading Improvement*, *25*(4), 276-81.

Golen, S. (1990). A factor analysis of barriers to effective listening. *Journal of Business Communication*, *27*(1), 25-36.

Gu, L. (2015). Language ability of young English language learners: Definition, configuration, and implications. *Language Testing*, *32*(1), 21-38.

Ha, M. (2012). Effects on the improvement of TOEIC scores of college English learners using listening-focused and reading-focused teaching methods. *Foreign Languages Education, 19*(4), 323-348.

Hadist, M. K., Hidjanah, Dtakiyatuddaaimah, & Dewanti, L. (2022). An analysis of EFL students' difficulties in listening comprehension at STKIP Muhammadiyah Bogor. *Jurnal Educatio FKIP UNMA, 8*(1), 174-179.

Hahta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* University of California Linguistic Minority Research Institute.

Hallinan, M. T., Bottoms, E., Pallas, A. M., & Palla, A. M. (2003). Ability grouping and student learning. *Brookings Papers on Education Policy*, *6*, 95-140.

Harada, S. (2016). Effects of the TOEIC course on students' TOEIC scores. *The Bulletin of St.Margaret's*, *48*, 1-17.

Hasan, A., Gushendra, R., & Yonantha, F. (2017). The Influence of prior knowledge on students' listening and reading comprehension. *IJEE (Indonesian Journal of English Education), 4*(1), 1-15.

Hedrick, W. B., & Cunningham, J. W. (1995). The relationship between wide reading and listening comprehension of written language. *Journal of Reading Behavior*, *27*(3), 425-438.

Hogan, T. P., Adlof, S. M., & Alonzo, C. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology, 16*(3), 199-207.

Hong, N. X., & Phan, N. T. (2020). Students' self-efficacy beliefs and TOEIC achievements in the Vietnamese context. *International Journal of Instruction, 13*(4), 67-86.

Hsieh, C.-N. (2017). The case of Taiwan: Perceptions of college students about the use of the TOEIC® Tests as a Condition of Graduation. *ETS Research Report Series*, 1-12.

Huang, T.-Y., Chen, S.-C., Chen, S.-H. E., & Wey, S.-C. (2015). The Role of Background Knowledge and Proficiency in Vocational EFL Learners' Listening Comprehension. *International Journal of Education and Research*, *3*(9), 115-126.

Hwang, H., & Cabell, S. Q. (2021). Latent profiles of vocabulary and domain knowledge and their relation to listening comprehension in kindergarten. *Journal of Research in Reading, 44*(3), 636-653.

Imsa-ard, P. (2019). TOEIC Reading section: Evaluation of the four cardinal criteria for testing a test. *NIDA Journal of Language and Communication*, 91-110.

In'nami, Y., & Koizumi, R. (2012). Factor structure of the revised TOEIC® test: A multiple-sample analysis. *Language Testing*, 131-152.

In'nami, Y., & Koizumi, R. (2017). Using EIKEN, TOEFL, and TOEIC to award EFL course credits in Japanese universities. *Language Assessment Quarterly*, *14*(3), 274-293.

Institute for International Business Communication. (2021, October 10). *TOEIC program data & analysis 2020.* Retrieved from http://www.iibc-global.org/library/default/english/toeic/official_data/pdf/DAA_english.pdf

Jee, S., & Lee, S. (2009). A study on the correlation between listening comprehension and reading comprehension through an analysis of TOEIC score. *Journal of Language Sciences*, *16*(1), 111-128.

Jones, M. G., & Gerig, T. M. (1994). Ability grouping and classroom interactions. *The Journal of Classroom Interaction*, *29*(2), 27-34.

Kim, Y.-S. G., & Petscher, Y. (2021). Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening comprehension). *Annals of Dyslexia, 71*(2), 218-237.

Kiss, D. (2017). A model about the impact of ability grouping on student achievement. *The B.E. Journal of Economic Analysis & Policy*, *17*(3), 20170036.

Komatsu, Y. (2015). An attempt to improve students' TOEIC test scores focusing on Part 2 & Part 5. *National Institute of Technology,Ishikawa College Bulletin*, *47*, 77-86.

Kurian, J., & Mekoth, N. (2021). Deconstructing coping using cognitive influences on ability groups. *Studies in Higher Education*, *46*(2), 177-188.

Lee, E., & Jin, S. (2009). English-only classes and student improvement on TOEIC. *Pan-Pacific Association of Applied Linguistics*, *13*(1), 23-38.

Lehnert, T. E., Krolak-Schwerdt, S., & Hörstermann, T. (2018). Language and nationality attitudes as distinct factors that influence speaker evaluations: Explicit versus implicit attitudes in Luxembourg. *Language & Communication, 61*, 58-70.

Lertcharoenwanich, P. (2020). Washback Effect of Direct TOEIC Preparation class on students' motivation. *Humanities and Social Sciences Journal, Ubon Ratchathani Rajabhat University*, *11*(1), 304-316.

Long, D. R. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition*, *12*(1), 65-80.

Lund, R. J. (1995). A Ccmparison of second language listening and reading comprehension. *The Modern Language Journal*, *75*(2), 196-204.

Magableh, I. S., & Abdullah, A. (2021). The impact of differentiated instruction on students' reading comprehension attainment in mixed-ability classrooms. *Interchange: A Quarterly Review of Education, 52*(2), 255-272.

Magno, C. (2010). Korean students' language learning strategies and years of studying English as predictors of proficiency in English. *Teaching English to Speakers of Languages Journal*, *2*, 39-61.

Manihuruk, D. H. (2020). The correlation between EFL students' vocabulary knowledge and reading comprehension: A case study at the English Education Department of Universitas Kristen Indonesia. *Journal of English Teaching*, *6*(1), 86-95.

Manna, V. F., & Yoo, H. (2015). Investigating the relationship between test-taker background characteristics and test performance in a heterogeneous English-as-a-second-language (ESL) test population: A factor analytic approach. *ETS Research Report Series*, *2015*(2), 1-23.

Mardianti, N., Wijayati, P. H., & Murtadho, N. (2021). The correlation between students' reading anxiety and their reading comprehension in ESP context. *International Journal of Language Education, 5*(2), 15-29.

Mazenod, A., Francis, B., Archer, L., Hodgen, J., Taylor, B., Tereshchenko, A., & Pepper, D. (2019). Nurturing learning or encouraging dependency? Teacher constructions of students in lower attainment groups in English secondary schools. *Cambridge Journal of Education, 49*(1), 53-68.

McGillicuddy, D. (2021). "They would make you feel stupid" - Ability grouping, children's friendships and psychosocial wellbeing in Irish primary school. *Learning and Instruction, 75*, 101492.

Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller, *Vocabulary studies in first and second language acquisition*, 194-211. Palgrave Macmillan. doi:https://doi.org/10.1057/9780230242258_12

Nakanitanon, P. (2021). Linking an English proficiency test to the CEFR: Setting valid cut scores. *The New English Teacher*, *15*(1), 53-64.

Nam, E.-h. (2016). Analysis on the relationships between the perceptions and attitudes of Korean university students towards TOEIC and their TOEIC scores. *English Language & Literature Teaching*, 225-248.

Nguyen, H., & Gu, Y. (2020). Impact of TOEIC listening and reading as a university exit test in Vietnam. *Language Assessment Quarterly*, *17*(2), 147-167.

Olejnik, S. (1978). Factors affecting the relationship between listening and reading. *Reading Horizons*, *18*(4), 277-81.

Oliveri, M. E., & Tannenbaum, R. J. (2017). Insights into using TOEIC® test scores to inform human resource management decisions. *ETS Research Report Series*, 1-9.

Park, S., Kwak, E.-J., Tak, J.-y., & Tate, T. (2020). Investigation on TOEIC score trends in Korea and its pedagogical implications. *Cogent Education*, *7*(1), 1-13.

Phan, N. T., Quang, H. T., & Nhat, T. T. (2019). The effectiveness of a Toeic preparation Course on nonenglish majored students' Toeic scores at a university in Vietnam. *Journal Of Science And Technology*, *37*(1), 90-97.

Powers, D. E. (2013). *Assessing English-language proficiency in all four language domains: Is it really necessary?* Educational Testing Service.

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*, 151-167.

Powers, D. E., Kim, H.-J., & Weng, V. Z. (2008). The redesigned TOEIC® (Listening and Reading) Test: Relations to test-taker perceptions of proficiency in English. *ETS Research Report Series*, 1-16.

Puengpipattrakul, W., Chiramanee, N., & Sripetpun, W. (2007). The challenge facing Thai graduates and their potential for English-required workforce. *Kasetsart Journal of Social Sciences*, *28*(2), 288-297.

Richard, J.-P. J. (2020). Investigating CEFR-J self-assessment and TOEIC listening and reading scores. *グローバルマネジメント[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). Retrieved from http://www.cc.kyoto-su.ac.jp/information/tesl-ej/ej12/a2.html

Roohani, A., Dayeri, K., & Farhang-Ju, M. (2017). Role of background knowledge in Iranian EFL learners' reading comprehension test performance. *International Journal of Language Testing*, *7*(1), 28-39.

Rosalina, E., & Nasrullah. (2019). The Ccrrelation between self-esteem and student's reading comprehension. *English Language Teaching Educational Journal*, *2*(2), 70-78.

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.

Sadighi, F., & Zare, S. (2006). Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners. *The Linguistics Journal*, *1*(3), 110-126.

Schmidgall, J. E. (2017). Articulating and evaluating validity arguments for the TOEIC® tests. *ETS Research Report Series*, 1-9.

Shi, H. (2021). Exploring factors for English language learners' language competence. *World Journal of Education*, *11*(1), 86-93.

Shin, J., Dronjic, V., & Park, B. (2019). The interplay between working memory and background knowledge in L2 reading comprehension. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, *53*(2), 320-347.

Shin, M.-H., & Lee, E. (2012). The importance of learning language and culture integration: Focused on TOEIC reading comprehension. *English Language & Literature Teaching*, *18*(3), 207-221.

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, *42*(3), 214-240.

Sok, S., Shin, H. W., & Do, J. (2021). Exploring which test-taker characteristics predict young L2 learners' performance on listening and reading comprehension tests. *Language Testing*, *38*(3), 378-400.

Suvarnaphaet, K. M., & Desgres, D. (2017). Evaluation of the effectiveness of a lesson plan combining the Direct Teaching Method (DM) and the Task-Based Learning teaching (TBLT) approach in the context of an intensive TOEIC course for Thai students from Silpakorn University Animal Science. *Veridian E-Journal, Silpakorn University*, *10*(5), 35-45.

Suzuki, M., & Daza, C. (2004). A review of the reading section of the TOEIC. *TESL Canada Journal*, 16-24.

Tannenbaum, R. J., & Wylie, E. C. (2013). *Mapping TOEIC and TOEIC bridge test scores to the Common European Framework of Reference.* Educational Testing Service.

Waluyo, B. (2019). Examining Thai first-year university students' English proficiency on CEFR levels. *The New English Teacher*, *13*(2), 51-71.

Wei, Y. (2013). *Monitoring TOEIC listening and reading test performance across administrations using examinees' background information.* Educational Testing Service.

Wei, Y., & Low, A. (2017). Monitoring score change patterns to support TOEIC® listening and reading test quality. *ETS Research Report Series*, 1-20.

Wilson, K. M. (1989). Enhancing the interpretation of a norm referenced second-language test through criterion-referencing: A research assessment of experience in the TOEIC testing context. *ETS Research Report Series*, i-99.

Wilson, K. M. (2000). An exploratory dimensionality assessmentof the TOEIC test. *ETS Research Report Series*, 1-33.

Wilson, K. M., Nagara, S. K., & Woodhead, R. (2005). TOEIC ® /LPI relationships in academic and employment contexts in Thailand. *ETS Research Report Series*, *2004*(1), i-26.

Wolf, M. C., Muijselaar, M. M., Boonstra, A. M., & Bree, E. H. (2019). The relationship between reading and listening comprehension: Shared and modality-specific components. *Reading and Writing*, *32*(7), 1747-1767.

Wong, Y. K. (2021). Developmental relations between listening and reading comprehension in young Chinese language learners: A longitudinal study. *Journal of Psycholinguistic Research*, *50*(2), 261-273.

Wood, C. L., Schatschneider, C., & VelDink, A. (2021). The Relation between academic word use and reading comprehension for students from diverse backgrounds. *Language, Speech, and Hearing Services in Schools*, *52*(1), 273-287.

Wudthayagorn, J. (2018). Mapping the CU-TEP to the Common European Framework of Reference (CEFR). *LEARN Journal: Language Education and Acquisition Research Network Journal*, *11*(2), 168-180.

Zakaria, Rahma, M., Ridhwan, M., & Eriza. (2017). The correlation between students' English learning habits and their TOEIC institutional test scores in State Polytechnic of Sriwijaya. *Holistics Journal*, *9*(18), 31-41.

Zhang, S. (2006). Investigating the relative effects of persons, items, sections, and languages on TOEIC score dependability. *Language Testing*, 351-369.