

A Comparability Study of Two Standardized English as a Foreign Language Tests

Inn-Chull Choi[†] & Youngsun Moon
Korea University

ABSTRACT

This study investigates the degree of comparability between TEPS and TOEIC regarding test content and construct considering that these elements are important since they are frequently utilized as official English tests in Korea. The New TEPS, an updated version of the original TEPS that reflects the needs analysis, and TOEIC, which was revised in 2016, were analyzed based on the following theoretical frameworks that focus on language elements. This study systematically analyzed test attributes based on: 1) Test Task Characteristics (type, length, speed, etc.) of each test section and item distribution based on the type and content category, 2) Various corpus indices of the test content, and 3) Test Usefulness (construct validity, authenticity, and interactiveness), comparing the overall test comparability between the New TEPS and TOEIC. Results indicate that although the tests' objectives and development foundations differ, there were similarities in content and construct. Nevertheless, a handful of discrepancies were also identified, necessitating the thorough examination of the tests' compatibility when interpreting test results. In addition, based on the theoretical framework of Test Usefulness, the strengths and weaknesses of each test's content and construct were examined. This study also postulates a suitable direction to develop a valid communicative proficiency test.

Keywords: comparability, standardized EFL tests, Test Task Characteristics, Test Usefulness, corpus analysis

1. Introduction

Official English tests are considered to provide a reasonable, reliable evaluation of English language learners' abilities in EFL environments, and in Korea, many examinees take various official English tests for several different purposes. The official English language test scores submitted by the examinees are presented using different scales of each test, so the assessment institutions that must evaluate the ability of examinees in an equitable manner use conversion tables to compare the

[†] Corresponding author: icchoi@korea.ac.kr



test results. This process of utilizing grade compatibility charts and using the converted test scores for the same purpose is based on the assumption that the official English test results are compatible, regardless of the purpose, content, and composition of each test. However, the TEPS and TOEIC are official English tests that are considered to have discrepancies in their use and composition and should preferably be used for each such purpose.

The TEPS is an EFL test that assesses a balanced combination of Cognitive Academic Language Proficiency (CALP) and Basic Interpersonal Communicative Skills (BICS). It is widely used as a qualification for university graduation or admission, or for civil service qualification. The test has been proven to be reliable and valid for almost 20 years since 1999. The TEPS Examination and Business Division of Seoul National University (JSPS) has been enhancing the validity and practicality of the TEPS by considering constructive criticism opinions (test difficulty, total test time, examinations, etc.). To further optimize the number of test items, the original test version with a total number of 200 test items was reduced to 135 (40 items / 40 minutes for Listening, 30 items each for Grammar and Vocabulary in 25 minutes, 35 items / 40 minutes for Reading).

On the other hand, the TOEIC expresses the attributes of an English for Specific Purpose (ESP) test, which focuses on English related to work used in business, excluding CALP. Although the two tests are expected to differ in content and composition, the scores of both tests are commonly interpreted as if they were interchangeable (especially in many cases of employment in Korea) based on grade compatibility charts or score conversion tables. Under the circumstances, it may be well worth the effort to investigate the extent to which the two tests are comparable in the test input and rubric in order to validate the current practice of interpreting the two scores on the basis of the conversion tables. Therefore, in this study, we used both quantitative and qualitative analysis methods in various aspects of test content and construct, focusing on the language aspects of the test (construct validity, authenticity, interactivity, etc.) to look in depth into the degree of compatibility of the New TEPS and TOEIC.

2. Theoretical Background

2.1. Test construct

The attributes of a test can be divided into content and construct. Both are

important elements that determine the test characteristics, and the theoretical background of the test construct will be introduced first, while the content analysis will be mentioned in Section 2. Test Analysis using Corpus Analysis. In language testing, the aspect of the Test Method Facet (Bachman 1990), or Test Task Characteristics (Bachman and Palmer 1996) plays a pivotal role in that, unlike the ability evaluation in other disciplines, the evaluation means and the subject are languages. The authenticity and validity of the test depend on how the Test Task Characteristics, which provide a theoretical framework of the content and a form of language assessment, approach the language function and content of the actual communication situation and the test taker's level of competence. In this respect, it is necessary to consider and examine each detailed aspect of the Test Task Characteristics in depth, in order to analyze the composition and rubric of the test. In other words, to compare the constitutional systems of the two tests, it is necessary to derive the patterns of the two tests, and to do this, we need to analyze the key elements of the test construct, which are the characteristics of the measurement elements, measuring factors for each test area, and frequency of each language element category.

2.1.1.1. Measuring factors by test area

The language area to be measured is a core element of the test construct, and it will be described first. It is desirable to measure the following receptive skills and linguistic competence factors that constitute the basis of the communicative competence. In particular, it is desirable to evaluate vocabulary and grammar in diagnostically-oriented testing, which plays a pivotal role in bridging the gap between education and evaluation in an EFL context.

2.1.1.1.1. Listening

The listening area is largely divided into micro-listening and macro-listening (Rivers and Temperley 1978). Micro-listening refers to the basic perception listening skills through an analysis of the language elements, such as pronunciation, vocabulary, and identification of the phrases. Macro-listening refers to the ability to listen through chunking in a pseudo-paragraph, not just to grasp realistic content, but also to communicate broadly through analogy, logical reasoning, and analogy of tone. Listening ability factors include: 1) pronunciation discernment (consonants and vowels, sandhi), 2) understanding of intonation, pause and stress, 3) under-

standing of spoken language, such as fillers, false starts, hesitation, repetition and redundancy, etc. 4) vocabulary / idiom knowledge, 5) grammar skills (syntactic skills), 6) ability to grasp details, and 7) ability to identify larger meanings (Ur 1984). In general communication situations, listening comprehension is macro-listening, so micro-listening and macro-listening techniques must be balanced.

2.1.1.2. Reading

The reading area is composed of 1) vocabulary (idiom), 2) grammar skills (syntactic skills), 3) ability to grasp details, 4) identification of the larger meaning, 5) logical reasoning, 6) analogy of attitude and tone, 7) evaluating and appreciating the overall flow, 8) ability of speeded reading, 9) world knowledge (non-verbal), and so on. It is preferable to focus on the evaluation of reading comprehension that emphasizes the overall content, such as contextual identification and logical reasoning, as well as grasping factual details, so that it is advantageous when using an interactive reading process rather than the bottom-up reading method used mainly by EFL beginners (Dubin, Eskey, Grabe and Savignon 1986; Goodman 1967; Smith 1982).

2.1.1.3. Grammar

Grammatical competence is distinguished into tacit/implicit grammatical competence, which is the actual language use ability for communication, and explicit grammatical knowledge, which is not directly related to communication ability, but is the knowledge of spoken and written words. This is a concept that is in line with Krashen's (1985) conscious learning and potential acquisition. Since there are differences in the grammatical rules applied to the context of colloquial and literary texts, it is desirable to evaluate them independently. In an EFL test, it is reasonable to measure the internalized (subconscious) grammar competence necessary to efficiently communicate spoken and written language.

2.1.1.4. Vocabulary

Vocabulary competence is the ability to use stemming (abbreviation) and idiomatic skills. It is desirable to refer to the criterion of difficulty of the vocabulary based on a corpus analysis because there are considerable differences in subjective diffi-

culty, according to the semantic network and frequency of use. As with grammar, it is important to categorize and evaluate vocabulary since there is a large difference between spoken and written contexts. It is desirable to include idiomatic expressions, such as vocabulary, collocation, idioms, and phrasal verbs, as well as written language vocabulary, since these are frequently used in colloquial communication. In addition, it is also essential in a vocabulary evaluation to assess the ability to distinguish and accurately use confusing words within a given context (polysemy and confusing words).

2.1.2. Test task characteristics

In order to maximize construct validity, authenticity, interactivity, and reliability, the multi-faceted test methods to consider for each content area based on the theoretical framework of Test Usefulness (Bachman and Palmer 1996) (see IV. 4. Test Usefulness Analysis Result), are as follows.

2.1.2.1. Listening

2.1.2.1.1. Number of questions and examination time

The most important variable for reliability, validity, and authenticity in a listening comprehension test is the pronunciation rate to read the test content. According to Rivers (1980), the normal pronunciation rate is (160-190) Words Per Minute (WPM), but the speed varies depending on various variables (Choi 1992). The difficulty level is adjusted according to the examinees' level, in order to maximize discrimination power and reliability, while reflecting the pronunciation speed in natural communication situations as much as possible.

2.1.2.1.2. Speech and text languages

Although a listening test is intended to measure the comprehension of spoken language, the listening comprehension test that is the most typical involves listening to questions and choosing the correct answer by reading written choices. However, this test type introduces serious problems. First, since listening and reading skills are blended, the language ability that is measured becomes ambiguous, which causes problems in terms of content validity and construct validity (due to adulterated

scores). Many examinees find it difficult to read and understand the choices in the shortly given time between test items to choose the answer. This is an example of the fact that the variables for reading comprehension are involved in the examination and in turn affects the evaluation of listening comprehension, which is a problem for test validity. Also, in this test type, test results may vary depending on which test strategy has been adopted. It is difficult to obtain a valid test result when variables related to the test strategy cannot be controlled. To solve this problem, it is appropriate to present all listening scripts, questions, and choices orally, from beginning to end.

2.1.2.1.3. Listening times

The method of listening two times to recorded scripts at normal speed and then solving the problem has several important purposes. The primary purpose is to improve the test reliability by increasing the discrimination power, and adjusting the degree of difficulty according to the examinee's skill level. More importantly, the secondary goal is to make artificial tests resemble natural communication processes. That is, when giving a somewhat lengthy textual speech to test takers, the listening passage is played twice, so students can grasp the overall meaning (macro-listening) during the first time, listen to the question, and then listen for the required information during the second time (micro-listening). This is a reasonable listening test method that reflects cognitive and psychological processes of natural listening comprehension.

2.1.2.1.4. One passage one item

In existing tests, it is common to produce long passages and present 3-5 questions. This type, which is frequently used for reading comprehension, has serious problems, including local dependence (the degree to which the probability of getting the previous question right or wrong affects the probability of the next item being right or wrong). Types with multiple problems per passage have a high local dependence, and thus have a lot of problems regarding the test reliability. Therefore, to increase the reliability of the test, local independence should be maintained by limiting test types to one item per passage. This is a nonsensical test pattern since it is difficult to understand the passage and solve the problems for long passages with multiple test items, even for Native English speaking professors, because the passages are

excessively long (Choi 1993).

2.1.2.1.5. Picture presentation

The test method to select an appropriate description presented in spoken language while looking at photographs seems to be good in terms of the communication situation, so the face validity appears to be adequate. However, there is an actual problem in the construct validity. In human language, there is a conceptual linguistic device that processes static and dynamic aspects to better portray the communication situation. Static pictures do not match the dynamic aspects of language (such as action verbs and aspects), which raises questions of the validity of the test. Therefore, to describe the static situation and the dynamic situation of the language, it is desirable to present this item type as a computer-based test (animation method), rather than as a paper-based test.

2.1.2.2. Reading

2.1.2.2.1. Speed reading

It is reasonable to emphasize the speeded test aspect in measuring the ability to read, which is an important element of the EFL ability to actively cope with an information flood in modern society. The number of questions to be solved within the time limit should be determined to maximize the reliability and validity, and the number of vocabulary items per question should be determined according to the degree of difficulty.

2.1.2.2.2. Contents of reading and one passage one item principle

As in Listening, the principle of local independence (an essential premise of the IRT) is an important prerequisite for tests with a high reliability. According to research utilizing the factor analysis method (Choi 1991), the TOEFL reading test with high local dependence poses a serious problem in the test bias, especially in relation to the topic factor. In order to minimize the problems of local dependency, it is desirable to make a test based on the One Passage One Item principle. However, the items developed on the principle of One Passage One Item tend to be shortened, so while maintaining independence between the items, it is possible to present few

items as One Passage Multiple Items towards the end of the test, by slightly increasing the length of the passage.

In a language proficiency test, the validity can be enhanced only when the influence of the background knowledge variable on the test results has been minimized (Bachman and Palmer 1996). In this respect, it is very important to select the test content subject. For the test objectives, there may be some differences in the content of the test, but in an overall communication ability test, in addition to nonprofessional academic writing, it is desirable to diversify passage sources, including reading materials (newspapers / magazine articles, public interest / commercials, novels, etc.), which are inevitably important parts of language use in real life. According to the needs analysis of those who are engaged in international trade for the Business English test, it is also reasonable to measure the off-office English (field English), which is essential for deepening and maintaining actual human relationships, as well as office English for office situations.

2.1.2.3. Grammar

2.1.2.3.1. Speeded test

Since conscious learning must be given enough time to function as a monitor (Krashen 1985), time constraints can be used to more accurately measure the sub-conscious acquisition that underlies communication skills. Existing testing methods frequently utilized by test takers that determine which grammatical knowledge is being measured and then use memorized grammatical knowledge to answer, provoke very negative washback effects. Given the fact that time constraints are required to measure acquisition more accurately by deactivating learning, the speeded test method that takes time constraints in consideration of the communication situations is a highly plausible EFL test method (Bachman 1996; Choi 1993, 1994, 2002; Oller 1979). It needs to be noted that the “speeded” test method employed in language assessment per se is not to be confused with the “speed” test method (as opposed to “power test” from the perspective of educational assessment), which may deteriorate discriminability and reliability due to a highly restricted time constraint.

2.1.2.3.2. Test method

For the test method, the most common method is to fill in the gap and to find the grammatical error, and word order and paraphrasing are also used in overseas

tests (Eiken, UCLES test). The fill-in-the-blank type is often considered an easy problem, and it evaluates the ability to construct sentences and the basic ability of sentence composition. In the error-finding type, which used to be prevalent in old-fashioned grammar tests, the micro-grammar ability is assessed (in a so-called discrete point manner) by underlining only a partial phrase and finding the wrong portion. In order to overcome these problems, it is appropriate to separate the semantic paragraphs (e.g. by inserting slashes in-between sentences) that constitute the written and spoken dialogues to measure the implicit knowledge, which is the true grammatical ability used in conjunction with the context.

2.1.2.3.3. Grammar and phrase

The context of the stem should reflect the natural communication situation. To measure true communicative competence, it is desirable to measure the grammar and phrase of two different levels of written and spoken language separately. In order to measure the grammatical competence, it is suggested to be presented in the form of a discourse, but the passage should be composed of conversations measuring phrasal grammar

2.1.2.4. Vocabulary

2.1.2.4.1. Speeded test

As with grammar, it is also desirable for Vocabulary items to use a test method that rejects mechanical learning without connection to communicative situations. The speeded test method should be adopted to prevent the very inefficient learning method of deciphering the meaning of a word literally while analyzing the grammar analytically.

2.1.2.4.2. Test method

In general, there are two types: a fill-in-the-gap form, and a synonym type. The synonym selection type is a paradigm of discrete-point testing, in which decontextualized vocabulary knowledge is measured by choosing the synonyms of the underlined word, without relying on the context. On the other hand, the fill-in-the-gap form increases the reliability and validity of the test because it takes the aspect of the

Integrative test, which can only depend on the context to select the correct answer (Oller 1979).

2.1.2.4.3. Context of stem

The context of the stem should reflect the natural communication situation. To measure a true communicative competence, it is desirable to measure two levels of vocabulary, both spoken and written, as in grammar. The stem of the written vocabulary can be expressed in one or two sentences, but it is reasonable to measure the spoken vocabulary in a stem composed of dialogues.

2.2. Analysis of test content using Corpus analysis

To analyze the characteristics of the language and propositional content constituting the content of the test, an objective evaluation based on the subjective evaluation by language evaluation experts and corpus analysis may be done. To aim at an objective analysis, a linguistic corpus analysis will be used (Bachman, Davidson, Ryan, and Choi 1995).

Using the corpus analysis, we can compare and analyze the values of each index in terms of vocabulary, syntax, discourse, and writing. A corpus analysis can be conducted with various programs and tools. For Coh-metrix, analysis on a text yields 106 indices (Graesser, McNamara, Louwerse, and Cai 2004). By using the Lexical Complexity Analyzer (LCA) to calculate the index related to the vocabulary and the L2 Syntactic Complexity Analyzer (L2SCA) to calculate the syntactic complexity index, it is possible to grasp the difficulty and nature of the passage (Haiyang and Xiaofei 2010; Xiaofei 2010). Another corpus analysis tool is the AntWordProfiler, which uses a specific vocabulary list to help analyze the passage's word composition and complexity (Anthony 2014).

In addition, there are some studies that use corpus analysis to examine the difficulty level of the test (Choi 1994; Freedle and Kostin 1993; Hamada 2015; Sheehan, Kostin, Futagi, and Flor 2010). Hamada (2015) conducted a difficulty analysis of the reading section of the Japanese Eiken test using Coh-metrix. From the analysis of the first to third grade reading test passages using vocabulary, syntactic structure, and semantic structure indices, the vocabulary and syntactic indices, which are surface level indices, showed a high predictive power for the difficulty levels. Freedle and Kostin (1993) conducted a more in-depth analysis of the difficulty of the TOEFL

reading passages with corpus indices. Factors related to the passage were examined for whether they play a significant role in predicting the item difficulty. The majority of the sentence levels and discourse indices showed a significant correlation with the difficulty of the TOEFL reading questions and a regression analysis also revealed that 11 indices, including the length of the passage, the length of the sentence, and the indices indicating the indicative characteristics had a significant predictive power. On the other hand, Sheehan et al. (2010) tried to develop an automated passage analysis tool that can predict the difficulty of the passage. They developed an analysis tool called SourceRater, referring to the limitations of the existing corpus analysis, such as the narrow range of index criteria, and the improper handling of genre effects.

Choi's (1994) study of the Seoul National University Criterion-Referenced English Proficiency Test (SNUCREPT) in the development stage of the New TEPS that is being analyzed in the current study explored nine indices to analyze SNUCREPT. In addition to length-related indices, vocabulary concreteness, lexical density, topic of the passage, and the number of clauses in the sentence were also used. The readability index and the FOG index were also included to determine the difficulty of the passage. The indices showing a significant correlation to the degree of difficulty were the average number of clauses per sentence, and lexical density. The topic of the passages also showed meaningful correlations, indicating that passage topics do affect the difficulty level.

As can be seen from the above, in the result of the studies examining the relationship between the difficulty level and the corpus analysis, the majority of the indices were found to show significant relationships. Therefore, by using a variety of corpus indices that affect the difficulty of the test items, a comparison between the two tests can be performed using objective numerical values.

2.3. Test usefulness

One of the most important characteristics in developing and utilizing tests is the Test Usefulness. Since the ultimate goal of a test is to measure the competence of the examinee, how useful it is to achieve this goal should also be considered. However, it is difficult to objectively and quantitatively measure the abstract concept of Test Usefulness. Bachman and Palmer (1996) define the overall Test Usefulness using six factors: reliability, construct validity, authenticity, interactiveness, impact, and practicality. According to Bachman and Palmer (1996), it is more important to

maximize the sum of the six factors than to maximize each factor, and it is impossible to independently measure these factors because they are all related. It is also impossible to make general judgments about which factors are more important because the balance between the factors will differ, depending on the circumstances and the characteristics of each test.

The validity of the test is related to the relevance and significance of the meaning that can be derived from the scores obtained from the test (Bachman and Palmer 1996) because when using the test scores to make any conclusions or judgments about the examinee's ability, the way in which the interpretation is justified is important. In the case of an English test, it can be said that the test's construct validity indicates whether the test actually measures the intended linguistic abilities. The construct validity also relates to the generalization of the meaning obtained from the test scores. Therefore, when considering the validity of the score interpretation, the construct definition and the testing task should be considered. There are two characteristics of the task given in the test, one being the Target Language Use (TLU), which can be related to the authenticity factor mentioned above, and the other being the correspondence of the language ability actually used in the test. This is because the actual language use and language skills needed in the test should correspond to give meaning to the test scores regarding the language proficiency. The second characteristic is related to how much of the examinee's language ability is required by the tasks that are given by the test. This can be seen as the interactivenss between the language ability and the actual test task.

Although there have been a number of studies conducted to investigate multifaceted constructs of language tests, little research has been conducted focusing on the comparability of standardized EFL tests in terms of the test usefulness (Bachman and Palmer 1996) and sophisticated constructs (Bachman et al. 1995). In this study, we analyze the Test Usefulness of the New TEPS and TOEIC using construct validity, authenticity, and interactivenss factors. Taking each factor into consideration, the characteristics, advantages, and disadvantages of the New TEPS and TOEIC tests will be analyzed, and the Test Usefulness of both tests will be objectively compared.

2.4. Research questions

The research questions of the current study are as follows:

- 1) Are there differences between the New TEPS and TOEIC test method attributes

(type, length, speed, etc.) and item distribution by category? If there are differences, what are the characteristics of each test?

- 2) Are there differences between the New TEPS and TOEIC corpus analysis results? If there are differences, what are the characteristics of each test?
- 3) Do the Test Usefulness analysis results differ? If there are differences, what are the characteristics of each test?
- 4) As a result of analyzing the language factors, what is the compatibility of the two tests, the New TEPS and TOEIC?

3. Research Method

3.1. Research data

The test data to be analyzed in this study are the New TEPS 1st pilot test from Seoul National University's TEPS headquarters, and the TOEIC simulation test conducted by leading educational institutions. The New TEPS 1st pilot test was used to analyze the New TEPS test, which was launched in May, 2018. Based on the needs analysis of the examinees and test users, the New TEPS 1st pilot test is composed of a test system that is basically similar to that of the original TEPS in terms of the philosophy of test development. Slight differences in the test rubric include the reduced number of items and test time, and the changed composition of grammar and vocabulary. The present research employed the mock TOEIC based on the de facto equivalency between the mock and actual TOEIC, given the fact that many institutions interpret the scores of mock and actual TOEIC as highly equivalent on the basis of the similarity of the test rubric and input.

3.2. Classification of items by category

The New TEPS and TOEIC items were classified by type and were compared accordingly. First, the items in the listening section were divided into five categories. The questions were divided into items on the subject of the listening passage, questions on detailed information, inference questions, questions that can be solved by understanding the idiom and English expressions, and other items. The other items included questions to choose the appropriate explanation for the picture in Part 1 of the TOEIC listening area. These questions were kept separate because they are

questions that deal with phonetic transitions of visual materials.

The grammar and vocabulary areas of the New TEPS are presented as separate sections, but the grammar and vocabulary areas of the TOEIC are included together in Part 5 of the reading comprehension section. Therefore, the total of 30 questions in the TOEIC test reading comprehension area, Part 5, are composed of 14 questions on grammar (## 2, 4, 5, 6, 10, 11, 13, 19, 20, 23, 26, 27, 28, and 30) and 16 questions on vocabulary (## 1, 3, 7, 8, 9, 12, 14, 15, 16, 17, 18, 21, 22, 24, 25, and 29).

In the case of the grammatical domain, it was divided into phrases and parts of speech, and then the phrases were divided into tense, subject-verb agreement, subjunctive, voice, relative, participial clause, and structural variation, while the parts of speech were divided into noun, verb, verbal, gerund, infinitive, participle, adverb, adjective, pronoun, preposition, conjunction, and determiners. The New TEPS had 30 grammatical items, but in the case of the TOEIC, there were only 14 items, and so there were many categories that did not have a single item.

The items in the vocabulary part were classified into general and form confusion items, and meaning confusion items. Meaning confusion items included questions that asked about words that have meanings that are easy to confuse as they are used, such as 'oath, pledge, vow, swear'. General and form confusion questions and meaning confusion questions were again classified into spoken and written language, according to the characteristics of the passage.

Finally, the items in the reading part were divided into vocabulary and expressions, grammar, topic, details, inference, and questions about the flow of passage. Vocabulary and expressions and grammatical items were separated because they were questions to find the correct word or grammatical form in the blank in Part 6 of the TOEIC, even though it was included in the reading section. As described above, the items of the New TEPS and TOEIC were classified according to their respective categories. Based on the results, the content of the two tests were compared by looking at the frequency of each category.

3.3. Test time and WPM

The New TEPS and TOEIC listening area test papers, scripts, and voice files were analyzed to calculate the words per minute (WPM), which is a very important factor in the listening comprehension area. In the case of Part 4 and 5 of the New TEPS, because the order of the script is 'passage - question - passage - question -

announcement', the script was also written in this way. In this script, the number of words was calculated for the entire listening area, and for each part. On the other hand, the length of the listening file was calculated by using the start and end times of each area. However, due to the characteristics of the listening test, there is a pause between questions, between the passages and the questions, between the questions and the announcements, and between the individual announcements. The pause times were measured collectively, and were excluded when calculating the WPM.

The analysis of grammar, vocabulary, and reading comprehension time was performed by examining the total number of vocabulary items in each area after examining the time limit and number of questions given for each area. The reading comprehension areas of the New TEPS and TOEIC are very different, and both the number of questions per minute and number of words per minute of each test were appropriately calculated. The New TEPS is divided into grammar, vocabulary, and reading comprehension, so the number of vocabulary items per minute for each part was calculated respectively, but the number of vocabulary items per minute was calculated at once for the TOEIC since there is no separation between the sections.

3.4. Corpus analysis method

To obtain the corpus indices, the original text of the New TEPS and TOEIC test were converted into a text file for each area. A corpus analysis was conducted considering the nature of the test texts after converting the instructions, the number of choices, and blanks. Text conversion was done collectively in accordance with the following principle:

- Delete the choice number (a) (b) (c) (d)
- Unify the blanks with '____' or fill them with correct answers, and then delete the blanks
- Batch conversion of unrecognized symbols
- Delete the speaker in the script of the file 'M:', 'W:'
- Delete 'Q:' indicating the question
- List non-level words in accordance with the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA)
- Delete all Direction parts
- Delete the symbol <, and > that causes error in Coh-metrix

Text files for listening, grammar, vocabulary, and reading section were made respectively. Because there are corpus analysis tools with a word limit, text files were also made by dividing each section into parts so it was necessary to obtain the index of the entire section after calculating each part first. The New TEPS listening section is divided into five parts, the grammar section has four parts, the vocabulary section has two parts, and the reading section has four parts. On the other hand, the TOEIC's listening comprehension section is divided into 4 parts, and reading comprehension is divided into 3 parts. Since the first part of the reading section in the TOEIC corresponds to the grammar and vocabulary section, the reading comprehension area was actually divided into two parts. Table 1 below shows the New TEPS and TOEIC area text file numbers:

Table 1. Text files of the New TEPS and TOEIC

	Listening	Grammar	Vocabulary	Reading
New TEPS	5	4	2	4
TOEIC	4	1	1	2

The New TEPS and TOEIC listening and reading comprehension passages that are not short were also analyzed by each passage. A corpus analysis was performed for the New TEPS listening area on the long conversations of part 3, short monologues of part 4, and long monologues of part 5, from ## 21 to 40. For the listening comprehension area of the TOEIC, the long conversations of part 3, and the long monologues of Part 4 were analyzed. In the case of reading comprehension, a corpus analysis of all items included in the New TEPS and TOEIC was conducted.

A total of five corpus analysis tools were used for the above text files.

3.4.1. Coh-metrix (Graesser et al. 2014)

Coh-metrix is a corpus tool that measures the consistency and coherence of the passages. Among the various indices, only the indices needed for this study were used.

3.4.2. Readability

Readability is an analytical tool that measures the readability of the passage. The indices used in this study are Kincaid Flesch's reading ease, and the FOG index showing the English education grade needed to read the passage.

3.4.3. AntWordProfiler (Anthony 2014)

AntWordProfiler is a program that analyzes the word structure and complexity of the passages, and the file that serves as the criterion of analysis can be uploaded directly. In this study, it was used to identify the academic / non-academic nature of the test passages.

3.4.4. Lexical Complexity Analysis (LCA) (Haiyang and Xiaofei 2010)

LCA is a program that analyzes the lexical complexity of a text using various measurement methods. In this study, many indices related to vocabulary (number of words, vocabulary variation, vocabulary characteristic, etc.) were obtained using LCA.

3.4.5. L2 Syntactical Complexity Analysis (L2SCA) (Xiaofei 2010)

If the LCA is an analysis tool for lexical level corpus analysis, L2SCA is a tool at the syntax level. It analyzes various syntactic complexity indices, based on clause, phrase, and T-unit.

Of the five corpus analysis tools above, Coh-metrix, LCA, and L2SCA are limited in the amount of text that can be analyzed at once. Therefore, the results of each section of each part analyzed within a limited amount were added or calculated. For example, the total reading section of the New TEPS exceeded the limit of the analysis program, so it was divided into four parts, and then each value was added or averaged according to the nature of the index.

From the corpus analysis tools above, a total of 22 corpus indices were used in this study (see Table 2).

Table 2. Corpus indices utilized in the current study

Variable	Meaning
<i>Vocabulary indices</i>	
Token	The number of tokens in the passage
Type	The number of types in the passage
Mean number of syllables in word	The average number of syllables in one word, indicating the length of word

Table 2. Continued

Variable	Meaning
Type Token Ratio (TTR)	The ratio of the number of types to the number of tokens, or words in a text
Standardized TTR	The standardized TTR is used to complement the TTR which is sensitive to the passage length
Lexical density	The ratio of the number of lexical words to the number of words
Lexical sophistication	The ratio of the number of sophisticated word types, defined as words beyond the most frequent 2,000 words, to the total number of word types in a text (Laufer 1994)
<i>Syntactic indices</i>	
Number of sentences	The number of sentences in the text
Number of clauses	The number of clauses in the text
Number of T-units	The number of t-units in the text
Number of complex nominals	The number of complex nominals in the text
Syntactic complexity	The inverse of the syntactic simplicity index which is based on the degree to which the sentences contain shorter lengths and simpler syntactic structures (negative values indicate syntactic simplicity)
Mean number of words in sentence	The average number of words per sentence (MLS)
Mean number of clauses in sentence	The average number of words in one sentence, indicating the length and complexity of the sentence
F-minus	The inverse of the readability index, Flesch Kincaid reading ease, indicating the complexity of the passage
FOG index	The readability index of a passage, indicating the years of education required to understand the passage
<i>Discourse indices</i>	
Narrativity	The narrativity index shows the degree of everyday, oral conversation of the text, based on word familiarity, world knowledge and oral language (negative values indicate less familiar topics of the text)
Connectivity	The connectivity index shows the degree to which the text contains explicit adversative, additive, and comparative connectives to express relations in the text (negative values indicate the lack of connectives in the text)
Temporality	The temporality index shows how many temporality cues there are, as well as how consistent the temporality is (negative values indicate the lack of temporality cues and inconsistent temporality)
<i>Pragmatic indices</i>	
Concreteness	The degree of how concrete, or easy to visually represent, the words within the text are (negative values indicate that the text contains more abstract words which may be challenging to understand)
Academic/nonacademic	The coverage rate of the passage based on the Academic Word List, indicating the academic character of the passage
Academic/practical topic	The topic of the passage either classified as academic, such as passages based academic topics such as history, social science, and science, or as practical, such as advertisements, e-mails, news articles, or invoices

3.5. Data analysis method

First, the test items in the New TEPS and TOEIC were categorized according to the item type. The frequency distribution of the question types of the New TEPS and TOEIC items were examined, and the similarities and differences between the two tests were found. Based on this result, the degree of compatibility in terms of test items was examined. After comparing the two tests through the classification of test items, the test time and WPM of each test were analyzed and compared. Next, the test passages were refined in a form suitable for corpus analysis, and the corpus indices were obtained from various corpus analysis tools. The items in the listening and reading sections that are not short were also analyzed per item. After examining all 22 indices, the New TEPS and TOEIC figures were compared to determine the compatibility of the content and the difference in the difficulty of the passages between the two tests. Finally, the Test Usefulness analysis of the New TEPS and TOEIC was conducted based on three factors: construct validity, authenticity, and interactiveness.

4. Results & Discussion

4.1. Comparison of test construct

4.1.1. Listening

First, as shown in Table 3 below, the New TEPS and TOEIC listening sections are divided into five parts and four parts, respectively. Part 1 of the New TEPS consists of 10 questions inferring a response in a short dialogue, and Part 2 consists of 10 questions inferring a response in a long dialogue. The next 10 questions of Part 3 are the types of questions that are solved by listening to long conversations. Finally, in Part 4 and Part 5, after presenting a passage in the form of a sentence, Part 4 is a type of listening to one short monologue followed by one item, and 4 questions in Part 5 follow a type of two questions on one long monologue. The biggest difference between the TOEIC and the New TEPS that can be seen in the part division is the six questions selecting the appropriate description according to the picture in Part 1. As pointed out in Chapter 2, this method seems at a glance to have good face validity, but a method to select a description corresponding to a static photograph is not preferable in the construct validity. The other components of the

TOEIC listening test are composed of 25 questions to a short conversation in Part 2, 39 questions to solve the problems of the long conversation in Part 3, and 30 questions to solve the problems of a long monologue in Part 4. One of the more noticeable differences between the two tests is that Part 3 and Part 4 of the TOEIC are a set of passages and questions that have three questions to be solved for a single passage.

Table 3. Number of items and item types of the listening section

New TEPS	Part 1	Part 2	Part 3	Part 4	Part 5
Number of items	10	10	10	6	4
Item type	Inferring a response in a short dialogue	Inferring a response in a long dialogue	Questions on a long dialogue	Questions on a short monologue	Questions on a long monologue

TOEIC	Part 1	Part 2	Part 3	Part 4
Number of items	6	25	39	30
Item type	Picture description	Choosing a response in a short dialogue	Questions on a long dialogue	Questions on a long monologue

The item type in the listening section includes questions that can be solved only by knowing native English expressions, the items that ask the topic of the passage, the items that ask detailed information after listening to the passage, the items that infer from the contents of the passage, and other items. The total number of items in the listening comprehension area of the New TEPS is 30 items, and of the TOEIC is 100 items. Table 4 below shows the results of classifying the items by each category:

Table 4. Comparing the frequency of the items of each subskill component of the listening section

	Expression	Topic	Detail	Inference	ETC	Total
New TEPS	6 (15%)	7 (17.5%)	8 (20%)	19 (47.5%)	-	40
TOEIC	-	9 (9%)	36 (36%)	45 (45%)	10 (10%)	100

Since the total number of the TOEIC questions was three times that of the New TEPS, the percentage of item numbers to the total test was also compared. The

commonality between the two tests is that the proportion of the inference questions in the listening comprehension area is close to half. Among the inference problems, there were common types of questions that infer the correct response after listening to a dialogue. In the case of the New TEPS, the questions in Part 1, items 1-10, were dialogues in two turns, and items 11-20 in Part 2 were the types that infer the correct response in a conversation of four turns. Items 7-31 in Part 2 of the TOEIC were a type to pick the right answer in the conversation of two turns. In addition to inferencing the right response, there were also questions in the New TEPS and TOEIC that examined the ability to infer from the content of the passages. For example, in the New TEPS, questions such as 'What can be inferred from the conversation?' corresponded to inferential questions. In the TOEIC, questions did not directly ask to infer from the passage, but there were items that did require inference, such as, 'Why does the woman say "that couldn't be better"?'.

In both tests, there were questions that asked main ideas and details, but the New TEPS had a relatively high percentage of topic questions, while the TOEIC had a higher percentage of questions asking for details. The six items categorized as the expression items in New TEPS can also be classified into topic, detail, or inference categories, but cannot be solved without understanding the English expression. For example, there are questions including idiomatic expressions from the relatively easy phrase 'in a mood for ~', to difficult phrases, such as 'spare moment', and 'nip in the air'. In the TOEIC, there were no items found to be necessary to understand English expressions in order to solve the questions, but items 1-6 to select the appropriate description for a given picture that is a question type which is not present in the New TEPS were classified as other items.

4.1.2. Grammar

The New TEPS and TOEIC grammar sections are very different from the aspect of the test construct (Table 5). The grammar questions of the TOEIC are given along with the vocabulary questions in Part 5, where grammar items are included in the reading comprehension, while the grammar section of the New TEPS is given as an independent area containing 30 items. The grammar section of the New TEPS consists of four parts. Part 1 and Part 3 are colloquial passages in the form of dialogues, while Part 2 and Part 4 are texts of the written language. The 10 questions in Part 1 and the 15 questions in Part 2 are fill-in-the-blanks type, and the 2 questions in Part 3 and the 3 questions in Part 4 are question types to find the choice with

a grammatical error. On the other hand, the 14 grammatical items of the TOEIC are mixed with 16 vocabulary questions, so the examinees solve the grammar and vocabulary questions without any separation. The question types are also unified as fill-in-the-blanks type, so they are not shown separately in the table below.

Table 5. Number of items and item types of the grammar section

New TEPS	Part 1	Part 2	Part 3	Part 4
Number of items	10	15	2	3
Item type	Gap-filling in spoken language	Gap-filling in written language	Finding the grammatical error in spoken language	Finding the grammatical error in written language

In the grammatical section, the item classification was divided into phrases and parts of speech. Then, the phrases were divided into 7 categories: tense, subject-verb agreement, subjunctive, voice, relative, participial clause, and structural variations. The parts of speech were divided into 12 categories: noun, verb, verbal, gerund, infinitive, participle, adverb, adjective, pronoun, preposition, conjunction, and determiners. Table 6 shows the distribution of frequency for each category. Among the categories of the parts of speech, verb, verbal, gerund, participle, and pronoun were omitted from the table since there was no item classified for those categories for both the New TEPS and TOEIC.

Table 6. Comparing the frequency of items of each subskill component of the grammar section

Phrase	Tense	Subject-verb agreement	Sub- junctive	Voice	Rela- tive	Participial clause	Structural variations	Total
New TEPS	8 (26.7%)	2 (6.7%)	2 (6.7%)	-	2 (6.7%)	-	4 (13.3%)	18 (60%)
TOEIC	5 (35.7%)	-	-	-	-	-	-	5 (35.7%)
Part of speech	Noun	Gerund	Infinitive	Adverb	Adjective	Preposition	Conjunction	
New TEPS	1 (3.3%)	2 (6.7%)	1 (3.3%)	-	1 (3.3%)	4 (13.3%)	1 (3.3%)	
TOEIC	2 (14.3%)	-	-	2 (14.3%)	3 (21.4%)	-	-	

Verb, Verbal, Gerund, Participle, and Pronoun omitted

First, the frequency of questions for each category of the New TEPS is 60% phrase and 40% parts of speech. The frequency distribution of 18 items corresponding to the phrase was evenly distributed in the categories, except for voice and participial clause. In particular, the items related to tense had the highest frequency, with 8 items. In the case of parts of speech, 12 questions were evenly distributed among the categories, excluding the verbs, verbals, gerunds, participles, and pronouns that were omitted from the table, and adverbs. The most frequent among the categories of the parts of speech was preposition, which included four questions.

On the other hand, the grammar area of the TOEIC is quite insufficient in evaluating the applicant's grammar knowledge because the number of questions is only 14, which is far less than that of grammar categories. As for the item classification, the grammar section of the TOEIC shows a skewed distribution towards only a few categories, and most of the categories do not include a single item. Of the 14 items, 5 items were classified as phrase, but all 5 items were about tense. The other nine questions were questions of the parts of speech, and two or three items were distributed in nouns, adverbs, adjectives, and determiners.

4.1.3. Vocabulary

Next, the items in the vocabulary section of the New TEPS and TOEIC were compared (Table 7). Like for grammar, the vocabulary section of the New TEPS consists of 30 independent questions, while 16 items in the TOEIC are included in Part 5 of the reading, along with the grammar questions. The New TEPS vocabulary area consists of all fill-in-the-blank types, but it is divided into Part 5, where colloquial passages are given, and Part 6, where a single written sentence is given. The vocabulary items of the TOEIC are not presented separately in the table below because they are not given as an independent part.

Table 7. Number of items and item types of the vocabulary section

New TEPS	Part 5	Part 6
Number of items	10	20
Item type	Fill-in-the blank of spoken English passage	Fill-in-the blank of written English passage

The items in the vocabulary section were classified into general and form confusion items and meaning confusion items, and each category was divided into

colloquial and literary forms. General and form confusion questions are questions that check whether examinees know the meaning of a word while meaning confusion questions are questions that select correct words in context, or ask about words that are easily confused. For example, item 38 in the New TEPS was a meaning confusion item for which (b) draw, which means to neither win nor lose, is the correct answer. Even if the examinee knows the representative meaning of ‘draw’, ‘to sketch, or depict’, it was an item that one could solve only by knowing the alternative meaning of the word ‘draw’. After categorizing items into these two categories, the passages were again divided into dialogue forms of spoken English and monologue forms of written English. Table 8 below shows the results of classification of the items by category in the vocabulary section.

Table 8. Comparing the frequency of items of each subskill component of the vocabulary section

	General and form confusion		Meaning confusion		Con- junction	Total
	Spoken	Written	Spoken	Written		
New TEPS	5 (16.7%)	19 (63.3%)	5 (16.7%)	1 (3.3%)	-	30
TOEIC	-	10 (62.5%)	-	3 (18.8%)	3 (18.8%)	16

Comparing the distribution of the vocabulary section, the most noticeable point is that the TOEIC does not have questions that present a colloquial type of passage. None of the vocabulary items in the TOEIC are in a spoken form of the language but are all in the form of selecting the appropriate word to fill in the blank of a single sentence. The TOEIC test emphasizes evaluating abilities needed for everyday life and work, focusing on communicative competence. The fact that the TOEIC does not assess vocabulary and colloquial expressions is a noticeable disadvantage. On the other hand, the New TEPS is composed of 30 items with 10 items of spoken language passages in a conversation format, and the remaining 20 items are written language passages that give a single sentence with a blank. Next, both the New TEPS and TOEIC had more general and form confusion questions, and meaning confusion questions comprised about 20% of the questions. The TOEIC vocabulary section also included items that asked about conjunctions in a way that is unclear as to whether they are in the grammatical or vocabulary domain. It seems that this is because questions of vocabulary and grammar in the TOEIC Part 5 were given without distinction.

4.1.4. Reading

The parts of the reading section in Table 9 show the differences between the New TEPS and TOEIC. The New TEPS is divided into four parts. First, Part 1 consists of a fill-in-the-blank item type that selects a phrase that fits in the context of a passage. Next, Part 2 with an academic passage and Part 3 with a practical passage both have question types of reading a passage, needing to solve one and two questions, respectively. Finally, Part 4 is the type of question to select a sentence that does not fit the flow in a given paragraph. On the other hand, the TOEIC's reading section is divided into two parts, Part 6 and Part 7, and the problem type and composition are very different from that of the New TEPS. Although Part 6 is a fill-in-the-blank type like the New TEPS, the TOEIC has many context-independent questions, such as choosing the correct word in the blank or choosing the correct grammar type, which makes it a different type from the New TEPS that requires examinees to find the correct phrase, based on their comprehension of the passage. The TOEIC Part 7 consists of one area, but Part 7 can be divided into two parts, Part 7A and Part 7B. Part 7A is a one-passage multiple-item type with one passage and several questions, while Part 7B is a multiple-passage multiple-item type with two or more passages with multiple questions. All of the passages in the TOEIC are practical, which is different from the New TEPS that also includes academic passages, as well as practical passages.

Table 9. Number of items and item types of the reading section

New TEPS	Part 1	Part 2	Part 3	Part 4
Number of items	10	13	10	2
Item type	Fill-in-the-blank of a passage	Questions on an academic passage with one item	Questions on a non-academic passage with two items	Choosing the sentence that does not fit
TOEIC	Part 6	Part 7A	Part 7B	
Number of items	16	29	25	
Item type	Fill-in-the-blank of a sentence	Multiple questions on one non-academic passage	Multiple questions on several non-academic passages	

Finally, the items in the reading comprehension area were classified by category. The items in the reading section were divided into vocabulary, grammar, topic, detail,

inference, and flow. The total number of items for the New TEPS is 35, and 70 for the TOEIC, which is twice the number of the New TEPS. Table 10 below shows the results of classifying the reading area items:

Table 10. Comparing the frequency of items of each subskill component of the reading section

	Vocabulary	Grammar	Topic	Detail	Inference	Flow	Total
New TEPS	-	-	8 (22.9%)	10 (28.6%)	13 (37.1%)	4 (11.4%)	35
TOEIC	7 (10%)	7 (10%)	6 (8.6%)	28 (40%)	16 (22.9%)	6 (8.6%)	70

As in the listening section, the number of questions in the TOEIC was twice as large as the New TEPS, so comparisons were made by referring to the ratios in each category. First of all, the TOEIC includes items that ask questions about grammar and vocabulary in the reading area, unlike the New TEPS. In the TOEIC reading part 6, there are 7 word items to select the appropriate words in the blank, and 7 to select the choice with the appropriate grammar form. Although a reading passage is given for these test items, since these items are far from requiring reading comprehension knowledge, it is desirable to classify them into vocabulary and grammar sections.

Next, the distribution of the items in the New TEPS shows that the items asking the topic, the items about details, and the inference items are appropriately distributed. Four items related to the flow of passages include two items that select the appropriate conjunctions in the blank, and two items that select sentences that do not match the flow in the passage, thus showing a balanced weight among each subskill. Meanwhile, the TOEIC items were more concentrated in detail questions and showed relatively less weight for flow questions. To be specific, the most frequently asked questions were items asking for details, at 40%, and questions asking about the topic comprised 8.6%, which was less than that for the New TEPS. In addition, the items related to the flow were all different from the New TEPS, in that they select the position to insert the given sentence.

In summary, comparing the frequency of distribution of the questions by category of the parts of the New TEPS and TOEIC, there were more differences than similarities. The most noticeable difference between the two test constructs was that in the grammar and vocabulary areas of the New TEPS, questions are given in each

area, whereas the TOEIC gives grammar and vocabulary items together included in the reading section. Also in the listening and reading areas, all the TOEIC passages are composed of practical topics and forms, while the New TEPS includes parts that are given with academic passages like the original TEPS.

The results of categorizing the question types also revealed that the two tests differ more. There was a large difference in the total number of items in all areas, and the results were also different when the ratios of the item types were compared. In the case of the listening comprehension area, the item type to select the most appropriate response for a conversation was common, but there were more differences. Part 4 and Part 5 of the New TEPS involves solving problems from academic passages, and except for the four questions in Part 5, they all follow the one passage-one item form to improve local independence. On the other hand, all of the TOEIC passages were practical, and Part 3 and Part 4 gave one listening passage with three question types following a one passage multiple item form. This type of question form always implies a risk of local dependency.

In the case of grammar, the New TEPS consisted of grammatical categories with grammatical items of both colloquial and literary subjects, but for the TOEIC, which had half the number of items of the New TEPS, most of the grammatical categories did not include a single question, so it would be quite insufficient in evaluating the examinee's grammar skills. The fact that the vocabulary section of the TOEIC consists of a type that fills the blank of a single sentence in the written language is contrary to the aspect of the TOEIC being a test that emphasizes evaluating communication ability. On the other hand, the New TEPS is composed of a balanced content of both spoken and written language, also including 10 questions that evaluate colloquial vocabulary using a dialogue form. The reading comprehension section included question types that are common to both exams, including topics, details, reasoning, and questions about the flow, except for the TOEIC vocabulary and grammatical type questions (Part 6). However, the item distribution by category showed the difference between the two tests, and although the item type was the same, the composition of the questions and the topic and characteristic of the passages were different.

In conclusion, when analyzing the test type and the number of questions by category, TEPS measures BICS and CALP in a balanced manner compared to the TOEIC, which appears to be attributed to the difference of test rubric and item types based on each of the tests' distinct objectives. Also, the New TEPS can be seen to measure diverse vocabulary and grammar linguistic elements, and it evaluates listen-

ing comprehension and reading comprehension in a varied and balanced way. Thus, while the TOEIC is an ESP test measuring Business English, TEPS is a test to measure General English with more focus on measuring practical communication skills, while still measuring academic literacy ability.

4.2. Test time and WPM

4.2.1. Listening

The New TEPS’s listening area provides all content as spoken language while the TOEIC presents only passages and questions as spoken language, and choices as text, as in the existing test system. Specifically, for the monologues in Part 4 and Part 5 of the New TEPS, the passages and questions are heard, and after listening to the passages and questions one more time, the choice is heard. However, in the TOEIC, passages and questions can only be heard once, and questions and choices are presented in written form. There may be a difference in the WPM between the two tests and the examinee’s perceived speaking rate due to this difference in method. Nonetheless, it is necessary to compare the results of the WPM of the speaking speed, which has a great influence on the test difficulty level for examinees during listening comprehension. Table 11 shows the WPM analysis results of the total listening duration and of each part.

Table 11. WPM analysis result of the listening section

Part	New TEPS	Part	TOEIC
Total listening section	146.81	Total Listening section	152.42
Part 1 (dialogue)	140.57	Part 1 (picture description)	147.27
Part 2 (dialogue)	146.60	Part 2 (dialogue)	134.26
Part 3 (dialogue)	151.82	Part 3 (dialogue)	155.77
Part 4 (monologue)	145.79	Part 4 (monologue)	159.20
Part 5 (monologue)	144.29	-	-

The WPM calculated for the total listening duration of the New TEPS is 146.81 WPM, and 152.42 WPM for the TOEIC, which is a difference of about 5.61 words per minute. This is attributed to the fact that the choices in the New TEPS are presented in oral mode, and the pronunciation speed depends on reading with

articulate pronunciation. Overall, both tests are considered to have a similar range of speaking speed, slightly below the normal rate of speech (ranging from 160 to 190 WPM).

In the analysis of the speaking rate for each part, it can be seen that the speeds of Part 1, Part 2, and Part 3 of the New TEPS are becoming slightly faster with 140.57, 146.60, and 151.82 WPM respectively, while the speed of the TOEIC Part 2 and Part 3 is 134.26 and 155.77 WPM respectively. Also, the speed of Parts 4 and 5 of the New TEPS is of about 145 WPM, and of Part 4 of the TOEIC is about 160 WPM, which shows a remarkable difference. By comparison, the fastest part of the New TEPS is Part 3, which solves the problem of a long conversation. The TOEIC was the fastest in Part 4 which is based on monologues, but this result is the opposite of a normal communication situation. The pronunciation speed of literary language in an actual communication situation tends to be slower than the pronunciation speed of colloquial speech. In this respect, Part 4 and Part 5 of the New TEPS are of desirable speed.

In addition, the New TEPS maintained a similar overall WPM throughout the parts, but the TOEIC showed a speed difference between the slowest and fastest areas. Also, the fact that the pronunciation rate of all areas for both tests is less than the actual communication situation of 160 to 190 WPM seems to be due to the test development principle of maximizing the discriminative power and reliability, considering that both tests are taken mostly by EFL examinees.

4.2.2. Grammar, vocabulary, and reading

As noted in the theoretical background, time constraints are inevitable in an actual communication environment, so testing to assess language ability should also use language skills in a similar situation. To measure subconscious acquisition, which is the core of lively communication skills, it is desirable to perform a speeded test (Bachman 1996; Choi 1993, 1994, 2002; Oller 1979) In this regard, to measure the degree of speediness of the New TEPS and TOEIC, the number of questions and length (vocabulary), along with the test time for grammar, vocabulary, and the reading part of the two tests, were analyzed and compared (Table 12).

Table 12. Test time, vocabulary and item number of the grammar, vocabulary, and reading section

	New TEPS			TOEIC		
	Grammar	Vocab	Reading	Grammar	Vocab	Reading
Test time	Total 25 min		40 min	Total 75 min		
Item number	30 items	30 items	35 items	Total 30 items	70 items	
Item number/min	2.4 items		0.88 items	1.33 items		
Word number	822	528	4,120	268	304	6,580
WPM	54 WPM		103 WPM	95.36 WPM		
WPM multiplied by 1.5	81 WPM		154.5 WPM	143 WPM		

As seen from the comparison of the test construct for each section, the composition of the grammar, vocabulary, and reading of the New TEPS and TOEIC was very different. The New TEPS grammar and vocabulary areas are comprised of 30 items each, but the time limit for the two sections is 25 minutes, and the reading section is given 40 minutes for 35 items. On the other hand, the TOEIC allows 100 questions to be solved within 75 minutes, without further time restriction regardless of grammar, vocabulary, or reading sections. As a result of analyzing the number of items to be solved per minute, the New TEPS showed that the grammar and vocabulary area had 2.4 questions per minute, and that the reading area had 0.88 items. Considering that grammar and vocabulary skills are used from subconscious acquisitions, the temporal pressure of 2.4 questions per minute will activate the acquired, rather than conscious knowledge of the examinees, which allows for a more accurate measure of lively knowledge. On the other hand, the TOEIC showed that the average grammar, vocabulary, and reading item speed was 1.33 items per minute. However, it is not possible to control the time spent by examinees in each area because the time limit is given for the three sections altogether. As a result, it is impossible to know how much time is spent for each area, and various test strategies used by each examinee will most likely affect the examination results, resulting in mixed scores (adulterated scores), which complicate the process of making a valid interpretation of the test results. The speed of information processing, which is very important in the reading process, was analyzed by the number of words per minute (WPM).

The New TEPS grammar and vocabulary area had a significantly higher number of words than the TOEIC, whereas for reading, the TOEIC showed a higher number

of words. The result of the grammar and vocabulary section of the New TEPS was 54 words per minute, and the reading section was 103 words per minute. Compared to this, only a calculation of three sections together was possible for the TOEIC of 95.36 words per minute. In the EFL test situation, since the characteristic of the passage is not easy or interesting, some logical cognitive ability is required. Therefore, in order to process the information of the passage and choices, the examinee must process the information of the given text and choice more than once at a minimum. Therefore, it is reasonable to interpret the actual amount of words that examinees interpret per minute to be about 1.5 times more than the number of WPM from Table 12 above. Therefore, the analysis results were calculated as 1.5 times, to obtain the actual number of vocabulary items to be read by the examinees. The results showed the grammar and vocabulary area of 81 words per minute, and the reading area of 154.5 words for the New TEPS, and the TOEIC of 143 words.

Extremely complex multidimensional elements - such as the genre, material, and form of the passage as well as the linguistic and theoretical difficulty hidden in the passage (type of information, distribution of information, and degree of contextualization; Bachman et al. 1995), the cognitive difficulty directly related to the examinee's background knowledge and cognitive ability, and the linguistic and cognitive difficulties of the choices - are concurrently linked to the reading process. Also, the speed of reading required depending on the attributes of these passages and options can be very different, so it is difficult to determine the desired reading speed in an absolute manner. If the level of educated native speakers handling information of ordinary writing in daily life is assumed to be 180~300 WPM (Dubin and Bycina 1991; Higgins and Wallace 1989; Jensen 1986; Nuttall 1996), the speed required to process logical text information in an EFL test context can be estimated to be approximately 150 WPM, according to the examination group proficiency level and examination purpose (Boudjella, Sharma and Sharma 2017).

4.3. Corpus index comparison

4.3.1. Listening

4.3.1.1. Total listening section

First, the results of the New TEPS and TOEIC listening area corpus indices were examined. Using the vocabulary, syntax, discourse, and pragmatic indices given in Table 13 below, the two tests were compared using objective numbers:

Table 13. Comparison of the corpus indices of the listening section

Index (22)		New TEPS	TOEIC
Vocabulary (7)	Token	2,952	5,259
	Type	1,316	2,084
	Mean number of syllables in word	1.51	1.39
	Type token ratio (TTR)	0.48	0.45
	Standardized TTR	0.88	0.88
	Lexical density	0.52	0.48
	Lexical Sophistication	0.24	0.23
Syntax (9)	Number of sentences	360	752
	Number of clauses	430	685
	Number of T-units	381	604
	Number of complex nominals	252	379
	Syntactic complexity	0.63	0.65
	Mean number of words in sentence	8.10	7.51
	Mean number of clauses in sentence	1.10	1.05
	F-minus	23.00	22.80
Discourse (3)	FOG index	6.30	7.00
	Narrativity	0.86	-0.01
	Connectivity	-0.67	0.05
Pragmatics (3)	Temporality	0	0.38
	Concreteness	-0.57	0.31
	Academic/non-academic	3.49	3.88
	Academic/practical topic	-	-

From vocabulary related numbers in the listening comprehension area, the total length of the TOEIC listening comprehension area can be seen to be close to twice the New TEPS based on the total number of tokens. Figures related to vocabulary variation show the nature of the vocabulary used. The average number of syllables in a word is an index that indicates the average word length. Long words are often more difficult than relatively short words and due to the nature of the listening comprehension area, words with longer length are difficult to understand by the examinees. The words used in the New TEPS can be seen to be longer because the mean of the number of syllables in the New TEPS was 1.51, which is higher than 1.39 in the TOEIC. Also, the TTR of the New TEPS is higher than that of the TOEIC, but the standardized TTR values are the same, so it is difficult to say that the difference in the lexical variation values is significant. The vocabulary density

index is an index that represents the ratio of content word to function word. The higher the content word ratio, the more difficult it is because the information in the passage is composed more densely. Since both the lexical density and the lexical sophistication are shown to be higher for the New TEPS, the difficulty of the text of the New TEPS is higher due to the nature of the vocabulary.

Next, the corpus indices related to the syntax for the listening comprehension area were examined. As was seen from the vocabulary indices, the TOEIC is much longer, so the number of sentences is also higher for the TOEIC. All length-related syntactic indices were higher in the TOEIC, but in terms of complexity indices, the New TEPS was higher in terms of difficulty. In particular, the sentence length of the New TEPS was 8.10, which was about 0.6 words longer than the 7.51 for the TOEIC. The mean length index of the sentence is significant because, as was mentioned in regard to the vocabulary indices, due to the nature of the listening area, it is more difficult for examinees to understand the sentence when the sentence becomes longer. Also, the syntactic complexity in the listening comprehension area was analyzed to be more difficult than the New TEPS, although it was a small difference. Among the syntax indices, only the syntactic complexity value was slightly higher for the TOEIC. From the readability indices, the difference between the two tests can be inferred to be insignificant, as the test that is more difficult alternates, depending on the calculation method.

According to the corpus indices related to discourse of the listening area, the New TEPS seems to be closer to colloquial language because it showed a higher narrative result. Considering that the New TEPS was provided with both academic and practical materials, while the TOEIC was all practical in nature, it is noteworthy that the passages of the New TEPS are closer to real life language. As for the connectivity and temporality, it would be more difficult to follow the flow of the passages and tense of the New TEPS since it had lower results.

Finally, the indices related to pragmatics in the listening comprehension area showed that the concreteness of the New TEPS passage was relatively lower than that of the TOEIC and the index showing the academic nature of the passage was high in the TOEIC. Since the method of analysis was to measure the ratio of the words included in the Academic Word List, the TOEIC was found to have higher results, even though the New TEPS covered more academic subjects. The classification of academic and practical subjects will be discussed in the item analysis because the classification of the items to hear and solve conversations with less than four turns was ambiguous.

4.3.1.2. Item analysis of the listening section

After comparing the corpus indices of the entire listening area of the two tests, the corpus analysis was performed separately for each of Parts 3, 4, and 5, in which the length of the passages were relatively longer. Using the vocabulary, syntax, discourse, and pragmatic indices given in Table 14 below, the two tests were compared using objective numbers.

Table 14. Comparison of the corpus indices per item of the listening section

Index (22)		New TEPS	TOEIC
Vocabulary (7)	Token	119.94	195.96
	Type	76.22	116.35
	Mean number of syllables in word	1.50	1.46
	Type token ratio (TTR)	0.64	0.60
	Standardized TTR	0.91	0.90
	Lexical Density	0.53	0.50
	Lexical Sophistication	0.27	0.25
Syntax (9)	Number of sentences	13.50	20.83
	Number of clauses	14.28	23.57
	Number of T-units	12.44	20.30
	Number of complex nominals	11.39	14.91
	Syntactic complexity	0.87	0.75
	Mean number of words in sentence	9.19	9.48
	Mean number of clauses in sentence	1.05	1.14
	F-minus	27.68	25.33
	FOG index	7.26	7.54
Discourse (3)	Narrativity	0.05	0.29
	Connectivity	-0.67	-0.07
	Temporality	-0.13	0.23
Pragmatics (3)	Concreteness	0.11	-0.46
	Academic/non-academic	3.97	4.11
	Academic/practical topic	1.78	2.00

Among the vocabulary related corpus indices, the values representing the number of words were calculated by the average length of each item, unlike when analyzing

the entire listening area. The average number of tokens per question for the TOEIC is 195.96, which is significantly longer than the New TEPS of 119.94. The TOEIC is not only longer in total test length, but the questions and choices in each item are also longer. As a result of analyzing only the items with long passages, the average number of syllables in a word showed a 0.04 difference between the two tests, and the New TEPS was higher. The TTR indices were similar to the results for the analysis of the total listening comprehension area since the TTR showed a higher result for the New TEPS, but the standardized TTR was almost the same. Among the corpus indices related to the lexical characteristics, lexical density and lexical complexity were high in the New TEPS, which can be regarded as dense information transmission. For this reason, it seems that the passages of the listening area of the New TEPS could be more difficult for examinees to understand.

When comparing the syntactic indices of the listening comprehension per item, the number of sentences for the TOEIC was higher, which is a consistent result with the vocabulary index. However, the result of the syntactic complexity index was different from that of the analysis of the entire listening area. Unlike the result from the analysis of all items, the result from the analysis of longer items showed that the New TEPS syntactic complexity was 0.87, which was higher than that of the TOEIC. The average length of the sentences also showed a different result, as the average sentence length for the TOEIC was 9.48 words, and longer than the New TEPS of 9.19 words. When analyzing only the second half of the listening section with longer passages, the average sentence length of the TOEIC was longer and the syntactic complexity of the New TEPS was higher. The readability values, F-minus, and FOG index were the same as the results of the whole listening area. The F-minus of the New TEPS was higher, and the FOG index for the TOEIC was higher, so there was no significant difference in the readability levels between the two tests.

As for the discourse related indices, the New TEPS narrative was lower than that for the TOEIC. On the other hand, it is likely to be more difficult for examinees to hear and understand passages of the New TEPS because the connectivity and temporality values for the New TEPS were lower. Finally, comparing the corpus figures related to pragmatics, the New TEPS passage was found to be more concrete than the TOEIC. The TOEIC only consisted of texts of practical characteristics, but the percentage of the Academic Word List used in the passage was actually higher than that of the New TEPS. Although the New TEPS passage in the listening comprehension did have passages with academic topics, the TOEIC test was found to have a higher percentage of academic word use.

4.3.2. Grammar

Next, the corpus values of the New TEPS and the TOEIC grammar areas were compared (Table 15). Unlike the New TEPS, in which the grammar area is composed of 30 items, the items corresponding to the grammar area of the TOEIC are included with the vocabulary items in one part of the reading. The number of questions in the TOEIC grammar area is about half that of the New TEPS. Considering that the length of the listening comprehension area of the TOEIC is nearly twice as long as the New TEPS, the length of each section for the TOEIC test is distributed in a biased manner.

Table 15. Comparison of the corpus indices of the grammar section

Index (22)		New TEPS	TOEIC
Vocabulary (7)	Token	822	268
	Type	444	189
	Mean number of syllables in word	1.43	1.72
	Type token ratio (TTR)	0.54	0.71
	Standardized TTR	0.91	0.94
	Lexical density	0.51	0.54
	Lexical sophistication	0.28	0.32
Syntax (9)	Number of sentences	64	23
	Number of clauses	132	28
	Number of T-units	96	19
	Number of complex nominals	88	38
	Syntactic complexity	1.80	1.00
	Mean number of words in sentence	8.92	9.54
	Mean number of clauses in sentence	1.43	1.00
	F-minus	23.50	48.00
	FOG index	7.20	11.10
Discourse (3)	Narrativity	0.35	-0.91
	Connectivity	-0.80	-0.04
	Temporality	-1.10	0.37
Pragmatics (3)	Concreteness	-0.93	-0.34
	Academic/non-academic	3.16	10.00
	Academic/practical topic	-	-

First, the analysis results to compare the corpus values related to the vocabulary used in the grammar section were examined. The total number of tokens showing the length of the grammar area of the New TEPS is three times longer than the TOEIC, and lengthwise, the TOEIC test seems to be too short to assess the grammatical knowledge of the examinees. On the other hand, although the length of the TOEIC is much shorter, the TTR figures showed that the TOEIC vocabulary was used for more diverse purposes. The mean of the average length of the word is 1.72 for the TOEIC, which is longer than the New TEPS of 1.43. Also, the TTR and the standardized TTR as well as the vocabulary density and sophistication showed a higher level for the TOEIC. This shows that the words used in the TOEIC grammar area were longer and more varied, and the vocabulary conveying the content was more densely contained. In terms of the vocabulary indices of the grammar area, the length of the New TEPS was three times that of the TOEIC, but the variation level and difficulty of the vocabulary indicated that the TOEIC is more difficult.

Next, the corpus indices related to the syntax of the grammatical domain were looked at. As can be seen from the number of tokens of the vocabulary related corpus index, the number of sentences of the New TEPS is also three times that of the TOEIC. Therefore, other length-related corpus indices also showed significantly higher results for the New TEPS. On the other hand, the indices related to complexity were different from each other in terms of the degree of difficulty. The syntactic complexity index showed a significantly higher result for the New TEPS (1.80) than for the TOEIC (1.00). The average number of clauses in a sentence was 1.43 for the New TEPS and 1.00 for the TOEIC, so the syntactic complexity of the sentences was higher for the New TEPS. On the other hand, the mean length of the sentences in the grammar area was found to be longer for the TOEIC. Since both of the readability indices are higher in the TOEIC, unlike the listening area, the readability value of the TOEIC for the grammar section is higher, and there was a significant difference between the two groups because the F-minus for the TOEIC was twice as high, and the FOG index showed a difference of three years.

Among the corpus indices related to discourse of the grammatical section, the narrative index was higher for the New TEPS, so the sentences in the New TEPS is likely to be closer to the spoken language. Connectivity and temporality values were consistent with the listening section, with both indices showing higher scores for the TOEIC. Finally, comparing the pragmatic corpus indices, the concreteness of the passages was lower in both tests, but the result of the New TEPS was lower. On the other hand, the TOEIC grammar area showed a much higher ratio than the New TEPS in the use of academic vocabulary. Since the passages in the grammar

area are composed of one or two sentences, the distinction between academic and practical material was ambiguous, and therefore was not classified.

4.3.3. Vocabulary

Next, the vocabulary section of the New TEPS and TOEIC was examined (Table 16). Like the grammar area, the TOEIC vocabulary questions were given as part of the reading section with the grammar items, unlike the New TEPS, which has a dedicated section containing 30 items for vocabulary. Therefore, the number of questions in the TOEIC vocabulary section is about half that of the New TEPS, and it seems that the number of questions is insufficient to adequately measure an examinee’s vocabulary skill.

Table 16. Comparison of the corpus indices of the vocabulary section

Index (22)		New TEPS	TOEIC
Vocabulary (7)	Token	528	304
	Type	340	212
	Mean number of syllables in word	1.47	1.59
	Type token ratio (TTR)	0.64	0.70
	Standardized TTR	0.93	0.94
	Lexical density	0.54	0.55
	Lexical sophistication	0.31	0.31
Syntax (9)	Number of sentences	49	26
	Number of clauses	88	30
	Number of T-units	75	21
	Number of complex nominals	44	33
	Syntactic complexity	0.99	1.59
	Mean number of words in sentence	6.95	9.47
	Mean number of clauses in sentence	1.16	0.94
	F-minus	24.10	34.80
	FOG index	6.60	8.50
Discourse (3)	Narrativity	−0.18	−0.59
	Connectivity	−1.16	0.42
	Temporality	−1.08	0.76
Pragmatics (3)	Concreteness	−0.04	−0.80
	Academic/non-academic	3.44	4.59
	Academic/practical topic	-	-

Looking at the corpus index related to vocabulary, the total number of words in the TOEIC was much smaller than the New TEPS as both the number of tokens and type indices were greater for the New TEPS, similar to the grammar area. However, the difference in the length between the two tests for the vocabulary section was smaller than that of the grammar section. On the other hand, the TTR and the standardized TTR scores showed a higher level for the TOEIC. Also, the mean of the number of syllables in a word was higher for the TOEIC than for the New TEPS, so this shows that the words used in the TOEIC vocabulary section are relatively longer. The vocabulary sophistication result was the same for both tests, but the vocabulary density showed a slightly higher result for the TOEIC.

Next, the syntax related corpus indices were compared, and the length-related figures showed a much larger number for the New TEPS, similar to the vocabulary indices. In addition to the vocabulary indices mentioned above, the syntax indices of the vocabulary section of the test also showed similar results as in the grammar section. The length of the vocabulary section of the TOEIC was short, but the syntactic complexity value for the TOEIC (1.59) was higher than that of the New TEPS (0.99), and the average length of the sentences was also longer for the TOEIC. The readability indices show that the level of the TOEIC is more difficult, as the difference between the two tests of the F-minus and FOG indices were quite large, which is a similar result to the grammar section. In particular, the difference between TEPS and the TOEIC for the FOG index is more than two grades.

For the discourse related corpus indices, the New TEPS narrative index was higher than the TOEIC, but both figures were negative. Connectivity and temporality were consistent with the results of the listening and grammar areas, showing that the New TEPS figures were much lower. Finally, comparing the pragmatic corpus indices, although the concreteness of the passages was found to be abstract in both tests, the passage of the TOEIC seemed more abstract. The difference in the academic vocabulary usage rate in the vocabulary part was not large compared to the grammar section, but the result was the same in that the TOEIC result was higher. As in the grammar area, the items in the vocabulary section are also composed of one or two sentences, so no distinction was made between academic and practical materials.

In general, the results of the corpus analysis of the grammar and vocabulary area showed a similar pattern. Although the length of the TOEIC was much shorter and the number of questions was smaller than the New TEPS due to the characteristics of the test construct, several indices showed that the difficulty of the TOEIC was higher when the vocabulary diversity or syntactic complexity was examined and the readability indices also showed that the TOEIC was more difficult by a large margin.

To sum, the grammar and vocabulary section of the New TEPS was longer in terms of the length, but the TOEIC was more difficult in terms of the content.

4.3.4. Reading

4.3.4.1. Total reading

Finally, the New TEPS and TOEIC passages were examined by comparing the corpus indices of the reading area (Table 17). Part 5, which is the first reading area of the TOEIC, was analyzed as grammar and vocabulary areas, so only parts 6 and 7 were analyzed.

Table 17. Comparison of the corpus indices of the reading section

Index (22)		New TEPS	TOEIC
Vocabulary (7)	Token	4,120	6,580
	Type	1,821	2,410
	Mean number of syllables in word	1.62	1.57
	Type token ratio (TTR)	0.51	0.40
	Standardized TTR	0.90	0.87
	Lexical density	0.55	0.52
	Lexical sophistication	0.35	0.30
Syntax (9)	Number of sentences	288	761
	Number of clauses	379	603
	Number of T-units	263	481
	Number of complex nominals	530	644
	Syntactic complexity	2.09	0.75
	Mean number of words in sentence	15.66	13.04
	Mean number of clauses in sentence	1.50	1.27
	F-minus	41.70	35.60
	FOG index	11.20	9.50
Discourse (3)	Narrativity	-0.68	-0.04
	Connectivity	-1.52	-0.35
	Temporality	0.13	0.18
Pragmatics (3)	Concreteness	-0.48	-0.97
	Academic/non-academic	6.24	6.38
	Academic/practical topic	-	-

Table 17 above shows the vocabulary related index of the reading area. As for the difference of the number of tokens, the TOEIC reading section is much longer than the New TEPS, like the listening section. The 6,580 tokens of the TOEIC were found to be 1.5 times more than the 4,210 tokens of the New TEPS. As a result, the number of types also showed a longer result for the TOEIC. On the other hand, the indices related to the vocabulary variation in the reading section showed a similar pattern to those in the listening section. The mean number of syllables in a word was 1.62 for the New TEPS, and 1.57 for the TOEIC, showing that longer words were used in the New TEPS. In addition, the length of the reading area of the TOEIC was longer, but the corpus indices related to the vocabulary variation were mostly higher for the New TEPS. Since both the lexical density and sophistication were higher in the New TEPS, it appears that the TOEIC was easier to read and understand passages.

Next, the syntax corpus indices of the reading section were compared. The number of sentences for the TOEIC was more than twice the number for the New TEPS. In addition, all length-related syntactic indices showed higher results for the TOEIC. On the other hand, there were many cases in which the complexity indices indicated that the New TEPS was more difficult than the TOEIC, regardless of the length. The syntactic complexity index showed a big difference between the two tests as the result of the New TEPS (2.09) was more than twice than that of the TOEIC (0.75). The mean length of the sentences also differed by more than two words, with 15.66 for the New TEPS, and 13.04 for the TOEIC. The average number of clauses in a sentence was 1.50 for the New TEPS, and 1.27 for the TOEIC. As the syntactic complexity indices show, although the New TEPS is shorter than the TOEIC in terms of length, it seems that it was more difficult for the examinees to understand the reading passage of the New TEPS because the grammatical aspect of the passage was relatively difficult. The readability indices also showed a difference in the difficulty of the reading area. Both F-minus and FOG indices indicated that the New TEPS was more difficult, and the FOG index was found to differ by about two years.

Comparing the discourse related indices of the reading section, the score of the narrative index for the TOEIC showed that it was closer to colloquial language. The connectivity and temporality indices of the New TEPS were found to be lower than those of the TOEIC in all areas, including reading comprehension. It seems that the passages of the New TEPS were more difficult for the examinees to understand because the passage connectivity was poor, and hints indicating the tenses were infrequent. From the pragmatic corpus indices, both of the tests were abstract rather

than concrete, but the New TEPS passages were relatively more concrete. The use of academic vocabulary was higher in the TOEIC, but without much difference from the New TEPS. Comparisons were made for academic and non-academic material use in the item analysis.

4.3.4.2. Item analysis of the reading section

As in the listening area, the passages in the reading area were analyzed per item. 35 items of the New TEPS and 70 items of the TOEIC were analyzed individually, and the mean value was calculated and compared.

Index (22)		Revised TEPS	TOEIC
Vocabulary (7)	Token	139.27	347.89
	Type	86.37	179.47
	Mean number of syllables in word	1.63	1.56
	Type token ratio (TTR)	0.66	0.53
	Standardized TTR	0.91	0.89
	Lexical density	0.55	0.53
	Lexical sophistication	0.37	0.30
Syntax (9)	Number of sentences	8.90	26.32
	Number of clauses	11.90	31.74
	Number of T-units	8.03	25.53
	Number of complex nominals	17.63	34.47
	Syntactic complexity	− 1.40	− 1.05
	Mean number of words in sentence	16.24	13.25
	Mean number of clauses in sentence	1.36	1.22
	F-minus	45.85	34.83
	FOG index	12.31	9.63
Discourse (3)	Narrativity	− 0.69	− 0.31
	Connectivity	− 1.42	− 0.47
	Temporality	− 0.17	0.23
Pragmatics (3)	Concreteness	− 0.09	− 0.54
	Academic/non-academic	6.30	6.40
	Academic/practical topic	1.47	2.00

Table 18 above shows the indices related to vocabulary as a result of analyzing the items in the reading section. The average length of the TOEIC for each question was longer than the New TEPS, with a difference of over 200 words. It is estimated that the length of each question is longer due to the nature of the TOEIC, which has three questions for one passage. On the other hand, most of the New TEPS had one question to be solved for one passage, and only five passages had two questions for a passage, so the length of each item was shorter. However, the length of the words used in the New TEPS was longer on average. The difference between the TTR values of the two tests that were obtained was larger than that of the analysis results for the entire reading section. The lexical density and sophistication indices were also higher for the New TEPS, so it seemed to be more challenging for the examinees to understand the passages of the New TEPS.

Next, the syntax related corpus indices for the analysis results of each item were examined. The number of the TOEIC sentences was significantly higher than that of the New TEPS because there are many passages in the TOEIC that are in the form of an ordering form or an e-mail due to the practical characteristic of the TOEIC passages. Unlike when analyzing the whole reading area, the syntactic complexity result was negative in both tests. The mean length of the sentences of the New TEPS was three words longer on average. Both readability values indicated that the New TEPS was more difficult, and the FOG index showed that the New TEPS was about three grades higher than the TOEIC.

Comparing the discourse related indices, both tests showed a negative value for narrative, but the TOEIC showed a relatively higher value. Connectivity and temporality were also found to be lower for the New TEPS for the item analysis results. Thus, the connectivity and temporality indices showed lower results for the New TEPS in all areas, including the results for the listening and reading section analysis conducted per item. Finally, as a result of comparing the corpus indices related to pragmatics, the concreteness of the passages was lower for the TOEIC. The use of academic vocabulary did not differ much between the two tests, but the difference for academic and practical subjects was significant. The New TEPS used almost an equal amount of academic and practical materials in the reading passages while the TOEIC only used practical materials for all passages. Thus, the difference for the academic and practical material use of the passages was very clear in the listening and reading sections, for which an analysis was conducted per item.

One of the notable points in analyzing the items of the reading area is Part 7 of the TOEIC. Part 7 consists of the newly-introduced multiple passage-multiple item

type together with the existing one passage-multiple item type. However, considering the length difference of the passages and the difficulty of the questions, it is not desirable to include the two question types together in one part. This is supported by the fact that there is a very large difference in the average number of words in the multiple passage type (530.4 words), and the one passage type (298.8 words). The introduction of a multiple passage question type seems like an attempt to achieve the goal of increasing the difficulty, by measuring comprehension abilities with complex passages. However, it will be necessary to examine whether this attempt is a good test method to measure the Business English construct that the TOEIC aims to evaluate.

4.4. Test usefulness

4.4.1. Construct validity

Choi (1993, 1994), who verified the validity of the Seoul National University Criterion-Referenced English Proficiency Test (SNUCREPT), which was the model of the original TEPS test that is quite similar to the New TEPS, refers to the nature and characteristics of the test, and verifies the validity of the test (Bachman 1996). In addition, Choi (1995) conducted a validity test to verify the compatibility between the SNUCREPT and TOEIC. This study also compared the construct validity of the two tests to analyze the Test Usefulness. First, the validity of each test was investigated by examining the content relevance and coverage of the test. Using the content aspects and Test Task Characteristics considered during the development of the TEPS test, Choi (1994) found that the content relevance and scope of the test's actual language use supported the test validity. An analysis of the content based on Bachman's (1990) model was intended to verify the construct validity by explaining the characteristics of directions (instruction, test assignment, and test structure) and input (length, audiovisual, form, and language) based on the notion that construct definition is an integral component to contemplate when investigating construct validity.

In this study, the construct definition of each test was analyzed as well since the framework of test development was not attainable. The New TEPS consists of four parts: listening, vocabulary, grammar, and reading. There are a total of 135 questions, composed of 40 items of listening, 30 items each of vocabulary and grammar, and 35 items of reading. On the other hand, the TOEIC is divided into two areas, listening and reading, and vocabulary and grammar are combined together in the

first part of the reading. The TOEIC test is different from the New TEPS in terms of composition, with 100 questions in listening, 30 questions in grammar and vocabulary, and 70 questions in reading. The result of dividing the grammar and vocabulary section of the TOEIC showed that grammar had 14 items, and vocabulary had 16 items. The fact that the TOEIC evaluates grammar and vocabulary in one part with a small number of items, although grammar and vocabulary belong to a strictly different linguistic category from each other, can be regarded in terms of validity as a serious drawback.

Accordingly, the score of the TOEIC grammar and vocabulary is not presented separately for each section, but it provides a mixed score (adulterated score), which acts as a fatal weak point from the aspect of construct validity, as it raises a problem in interpreting the score. In addition, it is impossible to know whether the reading score for the TOEIC is low due to the weakness of the subject's reading skills (identifying topic, detail, inference, etc.), or the lack of ability of each language element of vocabulary and grammar. If the examinee scores highly on the vocabulary and grammar elements but the reading score is low, he or she should develop their reading skills, and vice versa. As such, in the EFL situation, there are many examinees that lack the ability of a specific language element. Therefore, it is desirable to provide such detailed diagnostic evaluation information. In this respect, the New TEPS is considered to be superior in terms of construct validity and diagnostic evaluation purposes, which is the ultimate objective of testing.

In addition to the number of questions and classification of the items by category, a qualitative analysis of the content of the test also shows that the New TEPS is superior in terms of content validity, in that it deals with various aspects of vocabulary and grammatical language elements, including syntax. More specifically, the items in the New TEPS vocabulary and grammar sections are given with both colloquial and written language passages, but all the TOEIC passages are given in written language. Also, as shown in the results of classification by category, the New TEPS encompasses a wider range of grammatical categories, whereas the TOEIC has a smaller distribution of items.

4.4.2. Authenticity

Next, the New TEPS and TOEIC tests were examined in terms of authenticity. Authenticity is the degree to which the verbal abilities that are used depending on the test items and the actual language use correspond, and it should be taken into

account since it can justify the generalization of the test scores. The characteristic that enhance the authenticity of the New TEPS that Bachman (1996) mentioned is that all of the items in the listening area are given in an oral manner. Since the script, direction, and options of the question item of the New TEPS listening section are all given in the form of listening, it contributes to the authenticity of the area measuring the listening ability. On the other hand, the TOEIC provides different types of questions, depending on the part. Part 1, which selects the appropriate explanation for the picture, is visual and auditory, while Part 2, which selects the appropriate response to the dialogue, hears both the script and the option. For Parts 3 and 4 however, which gives dialogues and monologues and solve problems, the passage is heard, but the problems and choices are in printed form. The authenticity can be observed to be degraded in this process of measuring the listening abilities since reading the questions and options that are necessary to solve the test items requires both reading and listening abilities. On the other hand, the fact that the TOEIC reading passages, which aims to focus on practical English during the evaluation, is given in the form of text messages, emails, or invoices contributes to authenticity.

The lack of separation of the grammar and vocabulary sections, which is considered as a disadvantage of the TOEIC test, negatively affects the authenticity. There is also a view that these two areas are unnecessary because the grammar usage and vocabulary abilities are not directly visible in an actual language environment. On the contrary to this perspective, the reason to measure the grammar and vocabulary, even though they are not a function of language but language elements, is that English tests have diagnostic evaluation purposes. If the examinee's grammar and vocabulary skills are not measured, it would not be possible to provide diagnostic feedback on their listening and reading scores. For example, if the examinee's reading score is low, it is difficult to know whether he or she lacks detailed knowledge or inferencing skills in the language function, or grammar or vocabulary knowledge. If the vocabulary or grammar score is low, diagnostic feedback should be given to work on their language element; and if the vocabulary or grammar score is relatively high, diagnostic feedback should be given to develop their comprehension skills. Therefore, it is essential to include vocabulary and grammar tests for diagnostic evaluation in EFL situations, where a very large number of students lack basic vocabulary or grammar skills.

4.4.3. Interactiveness

Finally, we examined the interactiveness, which is defined as the extent to which

examinees incorporate their communicative competence into a cognitive process as required by given test method facets. In other words, the interactiveness is a natural function between Communicative Language Ability (CLA), or language competence used in a natural communicative setting, and the Test Method Facet (TMF) (Bachman 1990; Test Task Characteristics: Bachman & Palmer 1996), which includes five essential facets such as the test setting, the rubric, the input, the expected response, and the relationship between input and response). Interactiveness is enhanced when it reflects the natural cognitive process required in actual communicative situations. Because of the difficulty of finding observable evidence to demonstrate the interactiveness of the test, only the potential interactiveness can be inferred (Bachman 1996). An example of the New TEPS test interactivity is the construct of the listening area. In the same manner as the original TEPS, the last part of the listening of the New TEPS, which is the type of listening and replying to long passages, also follows the method of listening to the passage and problem, and then listening to the passage and problem again together with the choices. In this test method, the interaction that reflects the cognitive processes required in listening activities occurs when the examinees listen to passages and questions again, to grasp the content of the passages to be heard accurately.

The test type that shows interactiveness in the test method of the TOEIC is the newly introduced multiple-passage multiple-question type in the reading test. However, this is a test type that is distant from the TOEIC's purpose of evaluating the BICS focusing on practical situations. The test method of multiple-passage multiple-question is evaluated as a more appropriate test method to measure CALP, and the required language ability in high-level academic language use situations requiring complex cognitive processes. In order to understand how important this complex cognitive process is in actual business English situation, and whether it raises unnecessary difficulty, it is necessary to analyze the needs of test takers and verify the validity.

5. Conclusion & Implications

By analyzing the New TEPS and TOEIC, we were able to examine the compatibility between the two tests by classifying items by category, analyzing test time and WPM, utilizing corpus analysis, and comparing Test Usefulness. The results of test item classification by part and item type was discussed, and the effect of the test time

limit for the sections of each test was examined. In addition, after analyzing the results using the corpus indices, the Test Usefulness was compared and analyzed in terms of validity, authenticity, and interactiveness.

First, the New TEPS and TOEIC passages were analyzed and compared, and then the difference in the distribution according to the composition and type of each part was analyzed. The New TEPS and TOEIC items showed large differences in the number of questions and part composition. The New TEPS showed that the number of questions was evenly distributed across the four areas of listening comprehension, grammar, vocabulary, and reading, whereas the TOEIC showed the number of questions concentrated on listening and reading. The results of analyzing the frequency distribution by classifying the items of the two tests also showed more differences than similarities.

The most significant difference found from the result of categorizing the items of the listening section into main expressions, topic, detail, inference, and other items was that the New TEPS had items that were only solvable when knowing particular idioms or English expressions, and in the TOEIC, there were picture description items. Otherwise, topic, detailed information, and inferential items were common, but there was a difference in the frequency distribution of each item type. In the grammar area, the number of the TOEIC items was very small, so there were very few grammatical elements that were actually evaluated among the various grammatical categories. On the other hand, the New TEPS grammar test was evenly distributed over more categories than the TOEIC. In the vocabulary section, unlike the New TEPS, which also included verbal passages in the form of dialogue, the TOEIC had one type of item that was a single sentence in literary language. The classification result of the reading section also showed differences in the material of the passage and item composition. In the New TEPS, both academic and practical passages were given, but in the TOEIC, there were only passages of office English due to its specific purpose for evaluating business English. In fact, there were no passages of various materials required for communication also needed for the formation of relations in business situations, regardless of Eastern or Western culture. In addition, it is questionable whether the multi-passage multi-item method, which requires a rather complicated academic cognitive process, is suitable for business situations.

Next, the two tests were compared and analyzed by calculating the test time and the speech speed, which has a great effect on the test. As a result of calculating the WPM of the listening area, the WPM average of the whole area of the TOEIC was slightly higher than New TEPS, but the speech speed of the two tests was almost

similar, considering that the New TEPS presents choices in clear pronunciation. Both tests were at a level of 150 WPM, which is slightly slower than the speaking level of natural communication. This can be interpreted as an effort according to the test development principles of maximizing the degree of difficulty and discrimination through dialogue that is somewhat quicker in an EFL test situation and monologues that are slower.

The number of questions given in the test that must be solved within the time limit and the number of vocabularies that must be processed in the grammar, vocabulary, and reading sections have also been analyzed. There are difficulties in comparing the two tests because the time limit is not given for each area, but as a whole in the TOEIC. This contrasting point implies that how much time the examinees spend in each area cannot be controlled. Grammar and vocabulary sections should be conducted as a speeded test to assess subconscious acquisition, and for the proper interpretation of test scores, a separate time limit should be given in the TOEIC. The processing speed of the passage and options for the New TEPS was about 155 words for reading, and 81 words for grammar and vocabulary sections and 143 words for the TOEIC reading section altogether, and considering several studies on the speed of reading by native speakers and EFL learners, this is a desirable level.

A consistent pattern was found across the four sections when comparing the New TEPS and TOEIC tests using corpus indices. When examining the listening and reading area as a whole, the length of the TOEIC was several times longer than the New TEPS. However, as a result of examining the words and grammatical characteristics of the test, the difficulty or composition of the two tests showed differences not only in length, but also in various indices. In particular, when comparing the TTR indices or syntactic complexity indices, English used in the listening and reading sections of the New TEPS generally had higher difficulty in vocabulary and complex grammatical structures. In other words, in the listening and reading areas, although the TOEIC's length was longer, the difficulty of the New TEPS was a bit higher in terms of the propositional content.

Meanwhile, the grammar and vocabulary sections of the two tests showed opposite results. When comparing the simple length of the test, the length of the New TEPS was much longer because the TOEIC did not distinguish grammar and vocabulary questions. However, vocabulary variation indices and vocabulary difficulty were found to be more difficult for the TOEIC. Also, comparing the syntactic indices showed that most of them were more difficult for the TOEIC. In particular, when comparing readability indices, including F-minus and FOG indices, the TOEIC's

grammar and vocabulary section passages were more difficult. In other words, the grammar and vocabulary areas of the New TEPS were longer, but the TOEIC was more difficult in terms of vocabulary and syntactic aspects. This can be interpreted as a phenomenon caused by changes in the principle of test development of the TOEIC to raise the difficulty level in order to overcome the problem of the high gain score.

Next, when comparing the passages of the listening and reading sections, the most noticeable difference was the ratio of academic and practical passages. The listening section parts with longer passages of the New TEPS and TOEIC consist of dialogue and monologue. The dialogues in the listening area of the New TEPS are of practical material, and the monologues are classified into practical and academic materials, according to the content of the passage. On the other hand, the TOEIC's listening area gave a practical passage for all questions. In addition, for the reading section, the New TEPS consisted of more academic passages, along with practical passages including advertising, email, and letters; whereas, the TOEIC passages were confined to business English (office English, excluding field English). These results suggest that the New TEPS, which includes academic passages and practical passages in the listening and reading areas, can be used to evaluate a broad proficiency with various materials and communication functions.

A review of the content of the New TEPS and TOEIC through category classification and corpus indices by item revealed that the New TEPS covers BICS and CALP as a whole and in a balanced manner. On the other hand, the TOEIC focuses on the local level of work English, BICS, required by companies or workplaces in accordance to its test objective. It should be noted, however, considering that the ability to communicate outside the office (i.e., in field English) is also required to deepen interpersonal relationships for successful business, it would be desirable to include materials used in various communicative settings, such as newspaper and magazine articles covering topics of everyday events.

Finally, the Test Usefulness of the two tests from the aspects of construct validity, authenticity, and interactiveness was examined. In terms of construct validity, the fact that the TOEIC gives the grammar and vocabulary sections in one part without separation is a major disadvantage. In addition, the lack of a number of items to measure each language element and the biased distribution of grammatical language elements hinders the validity of the TOEIC and negatively affects the authenticity of the test. On the other hand, the New TEPS is superior in terms of construct validity because listening comprehension, grammar, vocabulary, and reading are

distributed evenly and are presented as separate sections. In addition, the composition of items in Part 4 and Part 5 of the New TEPS listening area serves as an advantage in authenticity and interactiveness. Authenticity is enhanced by the well-balanced coverage of the TLU and the oral mode in which the test input (including the choices) are presented. Interactiveness is enhanced by reflecting the natural cognitive process of listening (i.e., listening to the passage (macro-listening), listening to the question, and listening to the passages again (micro-listening) with question and choices). As a result of the analysis of Test Usefulness, the New TEPS was relatively better overall.

In conclusion, the New TEPS and TOEIC showed that they are significantly different tests from the aspects of the items compared by category, the test time and the speed of speech, the passages compared by corpus indices, and Test Usefulness. Based on the results of the internal language analysis, BICS was found to be a similarly measured construct in the two tests, but the other constructs measured for each test were CALP and Business English for the New TEPS and TOEIC, respectively.

Overall, the research findings reveal that the differences outweigh the similarities of the two tests. Therefore, when accepting the two exams for employment or qualification purposes, it is strongly recommended for more caution to be exercised in the decision-making process of interpreting the comparability of different test scores and setting cut-off scores based on a given conversion table. This study is not without limitations in that one set of each test was analyzed due to the limited availability of test materials for research. Further studies that analyze an increased number of tests are in need to enhance the generalizability and corroborate the current research findings. Nonetheless, it is hoped that this study will shed light on how to conduct a comparability study, including test equating, and will contribute to the theoretical framework required for sophisticated processes of the design, development, and validation of an EFL proficiency test intended to measure EFL communication skills.

References

- Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

- Bachman, L. (1996). Review of Seoul National University Criterion-Referenced English Proficiency Test (SNUCREPT). *Language Research* 32(2), 373-383.
- Bachman, L., F. Davidson, K. Ryan, and I-C. Choi. (1995). *Studies in Language Testing 1: An Investigation into the Comparability of Two Tests of English as a Foreign Language*. Cambridge, UK: Cambridge University Press.
- Bachman, L. and A. Palmer. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Boudjella, A., M. Sharma, and D. Sharma. (2017). Non-native English speaker readability metric: Reading speed and comprehension. *Journal of Applied Mathematics and Physics* 5, 1257-1268.
- Choi, I-C. (1991). *Theoretical Studies in Second Language Acquisition: Application of Item Response Theory to Language Testing*. New York: Peter Lang Publishing Inc.
- Choi, I-C. (1992). Interplay of natural phonology and sociolinguistic variable of formality. *English Teaching* 44, 65-88.
- Choi, I-C. (1993). Construct validation study on SNUCREPT (Seoul National University Criterion-Referenced English Proficiency Test). *Language Research* 29(2), 243-275.
- Choi, I-C. (1994). Content and construct validation of a criterion-referenced English proficiency test. *English Teaching* 48, 311-348.
- Choi, I-C. (1995). A comparability study of SNUCREPT and TOEIC. *Language Research* 31(2), 357-386.
- Choi, I-C. (2002). Validation of computer-based/adaptive language tests focusing on response time. *Multimedia-Assisted Language Learning* 5(1), 9-31.
- Dubin, F. and D. Bycina. (1991). Academic reading and the ESL/EFL teacher. In *Teaching English as a Second or Foreign Language (2nd ed.)* Ed. Celce-Murcia, M. Boston: Heinle and Heinle.
- Dubin, F., D. Eskey, E. Grabe, and, S. Savignon. (1986). *Teaching Second Language Reading for Academic Purposes*. Addison-Wesley Publishing.
- Freedle, R., and I. Kostin. (1993). The prediction of TOEFL reading item difficulty: Implications for construct validity. *Language Testing* 10(2), 133-170.
- Goodman, S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist* 4, 126-135.
- Graesser, A. C. D. S., McNamara M. M., Louwerse, and Z. Cai. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods* 36(2), 193-202.
- Haiyang, A. and L. Xiaofei. (2010). A web-based system for automatic measurement of lexical complexity. *Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*. Amherst, MA. June 8-12.
- Hamada, A. (2015). Linguistic variables determining the difficulty of Eiken reading passages. *JLTA Journal* 18, 57-77.
- Higgins, J. and R. Wallace. (1989). Hopalong: A computer reader pacer. *System* 17(3), 389-399.
- Jensen, L. (1986). Advanced reading skills in a comprehensive course. In *Teaching Second Language Reading for Academic Purpose*. Eds. Dubin, F, Eskey, D, and Grabe, W. Reading,

- MA: Addison-Wesley Publishing.
- Krashen, D. (1985). *The Input Hypothesis*. London: Longman Inc.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal* 25(2), 21-33.
- Nuttall, C. (1996). *Teaching rEading Skills in a Foreign Language*. Oxford: Heinemann.
- Oller, J. (1979). *Languages Tests at School: A Pragmatic Approach*. London: Longman Inc.
- Rivers, W. (1980). *Teaching Foreign Language Skills*. MA: Harvard Univ. Press.
- Rivers, W. and M. Temperley. (1978). *A Practical Guide to the Teaching of English*. NY: Oxford Univ. Press.
- Rost, M. (1990). *Listening in Language*. New York: Longman Press.
- Sheehan, K. M., I. Kostin, Y. Futagi, and M. Flor. (2010). Generating automated text complexity classifications that are aligned with targeted text complexity standards. *ETS Research Report Series* 2010(2).
- Smith, F. (1982). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. (2nd ed.) NY: Holdt, Rinehart and Winston.
- Ur, P. (1984). *Teaching Listening Comprehension*. Cambridge: Cambridge Univ. Press.
- Xiaofei, L. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4), 474-496

Inn-Chull Choi (1st Author)

Professor

Dept. of English Language Education

Korea University

Anam-ro, Seongbuk-gu, Seoul 02841, Korea

E-mail: icchoi@korea.ac.kr

Youngsun Moon (2nd Author)

Researcher

Center for English Language Education

Korea University

Anam-ro, Seongbuk-gu, Seoul 02841, Korea

E-mail: youngsun1993@korea.ac.kr

Received: May 28, 2018

Revised: July 20, 2018

Accepted: July 23, 2018