

Legitimate or Illegitimate Uses of Test Scores in Contexts Unrelated to Test Purposes*

Gwan-Hyeok Im**

(Queen's University)

Tim McNamara

(The University of Melbourne)

Im, Gwan-Hyeok & McNamara, Tim. (2017). Legitimate or illegitimate uses of test scores in contexts unrelated to test purposes. *English Teaching*, 72(2), 71-99.

Scores on the Test of English for International Communication Listening and Reading (TOEIC LR) have been used for rolling-based university admissions in South Korea. Because this test is not designed for that purpose by the testing organization, there may be a need to explore the validity of TOEIC LR score use for admission purposes. Accordingly, this study investigated how admissions officers in South Korea perceived TOEIC LR scores and their uses and why they used the scores for admission purposes. We adopted a sequential mixed-method design (Creswell & Clark, 2011) involving three phases of data collection: (1) analysis of application handbooks to identify the questionnaire survey target, (2) a survey of faculty ($n = 20$) involved in admissions, and (3) semi-structured interviews with selected participants ($n = 5$) to delve into their questionnaire responses. The findings were that (a) almost half of the universities in Korea were implementing special admissions and accepted the scores on the recognized English tests; (b) use of the test lacks validity because most of the admissions officers did not view the test scores as an indicator of English proficiency and academic aptitude; (c) approximately half of the admissions officers viewed the scores as a reflection of an applicant's effort, test-taking skills learned at language institutes, and former residence in English-speaking countries; and (d) the TOEIC LR scores were used due to a lack of other English tests as well as social pressure such as test fees and testing site availability.

Key words: the TOEIC LR, score interpretations and uses, admissions officer system, value implications, English proficiency, academic aptitude

* This paper is built on research conducted by the first author for his master's thesis. It was presented at the 2014 International Conference of the Korea English Language Testing Association, which was held at the Seoul National University of Education, Seoul, Korea.

** Gwan-Hyeok Im: First author; Tim McNamara: Corresponding author

1. RESEARCH CONTEXT AND RATIONALE

Since 1999, Korean universities¹ have used an admissions officer system that involves an evaluation of university applicants' scholastic aptitude and their potential academic success at the university by collecting various types of evidence such as academic achievements in school and leadership roles. The applicants' efforts are reviewed based on a portfolio of school experience. This admission system is rolling based and was introduced to complement the existing university admission system, which focuses on scores on the College Scholastic Aptitude Test (CSAT) designed and annually administered during the second week of November by the Korea Institute for Curriculum and Evaluation (KICE). One admission system is *special admissions* for students with excellent language skills and has been independently implemented under various labels such as *Global Talent*, *Global Leader*, and *English (Foreign Language) Special Talent*. This type of admission was still implemented across Korean universities as of 2016.²

This special admission system is not limited to languages related to university departments such as the English, Chinese, Japanese, or German departments. Even admissions to the arts, engineering, and economics departments require English language test scores. Although each university has its own admission procedures, most universities use internationally or nationally authorized English test scores as part of the first stage of the admission process. To pass this stage, a score on a widely recognized standardized proficiency test is generally required, as in, the Test of English as a Foreign Language (TOEFL) and the Test of English for International Communication Listening and Reading (TOEIC LR) designed by the Educational Testing Service, the Test of English Proficiency (TEPS) developed by Seoul National University, and the International English Language Testing System (IELTS, academic) launched by the Cambridge English Language Assessment and the British Council. However, the required English test and cut-off scores vary depending on the university, department, and college entrance type. Although a variety of English test scores are used for university admissions, the most widely used test is the TOEIC LR since the TOEFL and IELTS may not be easy for high school students: the tests include speaking and writing sections, which are not generally taught in high schools in Korea.

When multiple admission procedures are used, applicants are not eligible to proceed to the next admission stages if they do not obtain the required eligibility scores on the English tests. In the admission process, academic staff are temporarily directly or indirectly

¹ This study was conducted in South Korea. Therefore, Korea in this paper refers only to the country of South Korea.

² For example, Korea University, Kookmin University, Dongguk University, Konyang University, Sangji University, and Keimyung University.

involved as admissions officers during the admission season. The cut-off score for the English tests is generally set by the university and reflects the opinions of the academic staff. The academic staff is then able to evaluate the applicant's performance using interviews, essays or other exams.³ The administrative staff collects the scores, determines the ranking, and then notifies the applicants of the result (i.e., accepted or declined). Therefore, the role of admissions officers (i.e., professors at their university) is critical in the selection of applicants in the university admission process.

However, there may be a problem in the admissions officers' use of TOEIC LR scores: the test was not designed for academic purposes but rather for business purposes (i.e., it assesses business English proficiency in the workplace). When test scores are used for unrelated purposes, the validity of the score may be threatened. These validity issues of score (mis)use have been discussed by researchers (e.g., McNamara & Roever, 2006; Menken, 2008; Shohamy, 2001) because tests were being used to achieve the political or other goals of stakeholders (e.g., policy makers in schools, companies, the nation, or multilateral institutions such as the Organization for Economic Cooperation and Development and the European Union).

As the notion of validity has expanded (Kane, 2006, 2013; Messick, 1989), the way in which score interpreters interpret and use test scores needs to be investigated. However, there has been limited research into the interpreters' beliefs and perceptions in a higher-education context (see Ginther & Elder, 2014; O'Loughlin, 2011, 2013). Thus, we examined how university admissions officers understand scores on the TOEIC LR in university admissions in Korea. Specifically, because the construct of the TOEIC LR does not endorse the use of test scores by admissions officers in a context (i.e., university admissions) unrelated to its intended purpose (i.e., to assess business English proficiency), we therefore tackled the issue of test (mis)use and construct validity directly by evaluating how TOEIC LR scores are perceived by university admissions officers throughout Korean higher education.

2. TOEIC LISTENING AND READING

The Test of English for International Communication Listening and Reading (TOEIC LR) component was developed by the Educational Testing Service (ETS) in the United States. It was introduced in Japan in 1979 following a request from the Japanese Ministry of International Trade and Industry (Booth, 2012; Powers & Powers, 2015) and was then

³ As of 2016, these additional admission procedures were used in addition to the authorized English test scores.

introduced in Korea in 1982. The test was administered without any changes to the format for more than 20 years (Booth, 2012), but it was redesigned in 2006 to “better align test questions with everyday workplace language scenarios and to provide test takers with more information about their listening and reading proficiency levels” (Powers, Kim, & Weng, 2008, p. 1). For example, while only standard American English accents were used in the TOEIC Listening section, this revised version included various English accents such as British, Australian, and Canadian English accents in listening stimuli, to improve the test’s authenticity (Schedl, 2010). Text messages and online dialogue were included in the reading section of the updated version⁴ of the TOEIC LR as of 2016 to reflect current communication methods in an international business context (ETS, n.d.). However, in both versions, the testing time, method of administration (i.e., pencil and paper), range of difficulty, and score scales were the same as those of the previous form of the test (ETS, 2013; ETS, n.d.), although there were additions and deletions of some question items in the TOEIC LR in both years.

These modifications were made based on evidence-centered design (Mislevy, Steinberg, & Almond, 2002, 2003) to provide more precise information about English ability (Schedl, 2010). Before specifying the test constructs, a domain analysis was conducted to identify the language abilities needed for everyday work in a business context (i.e., valued knowledge and skills and representative task features that employees may encounter in their workplaces) (ETS, 2013). Then, the claims about candidates based on their performance (i.e., evidence) on the test, score descriptors, and relevant tasks, including test items, were systematically formulated in the TOEIC LR in alignment with the test purposes, detailed descriptions of which will be presented in the following section.

When evaluating the adequacy of the interpretations and uses of TOEIC LR scores in contexts (un)related to the test’s purpose, it is necessary to examine the test specifications to explore whether the test score is relevant and useful in certain contexts (Messick, 1989). Thus, we briefly describe the test’s purposes, target population, content and corresponding tasks, and the English abilities that it measures.

2.1. Test Purposes and Target Population

The TOEIC LR was “designed to measure everyday English skills used in an international business environment” by non-native speakers of English (Powers & Powers, 2015, p. 152). In addition to this primary purpose, specific appropriate uses of the test scores are described, which range from hiring in the workplace to measurement of students’ workplace English proficiency in schools (see ETS, 2013, p. 15). Because the test

⁴ The updated version of the TOEIC LR was introduced to only Korea and Japan in May 2016.

has multiple recommended uses, it has come to meet the needs of various stakeholders (Booth, 2012), e.g., college students who want to demonstrate their practical skills in English in a workplace setting, employees in international business, and personnel who use English in their work settings. However, if the test is used for purposes other than those listed by the ETS (2013), it is important for the test user to provide a valid justification for that purpose (ETS, 2013). It is also important that separate validation arguments be developed for these independent uses to ensure that inferences based on the scores are valid (and fair) in all approved contexts (Kane, 2013).

2.2. General Descriptions of the TOEIC LR: Test Content, Format, Tasks and Measured English Abilities

Samples of spoken and written English in the global workplace were collected from numerous countries around the world (ETS, 2013). The range of the test content is generally broad, e.g., travel, entertainment, housing, health, general business, technical areas, and personnel (ETS, 2013), and includes everyday English use related to business that employees or employers may encounter in their workplaces (for detailed information on the test content, refer to ETS, 2013, p. 4). That is, the content is not limited to a specific business domain but instead spans the use of English in general business and other commonplace settings (see the examples listed above).

The updated version of the TOEIC LR released in 2016 consists of seven parts (four in the listening section and three in the reading section). The test contains 100 items, and there is a possible score of 495 points for each section. The administration times are 45 minutes and 75 minutes for the listening and reading sections, respectively.

The listening component of the TOEIC LR includes six visual items in Part 1, 25 question-response items in Part 2, 39 conversation items (13 conversations with three items per conversation) in Part 3, and 30 spoken items (10 spoken items with three questions per talk) in Part 4. A single speaker provides the prompts in Parts 1, 2 and 4, whereas more than three speakers with different English accents such as American, British, Canadian, and Australian accents are included in the conversations in Part 3. The English listening abilities to be measured are reasoning ability regarding the essence, purpose and context and understanding of details in short and extended spoken texts.

The reading section includes 30 incomplete sentences to be completed in Part 5, 16 text completion items in Part 6, and 29 single-passage items (10 single passages with 2–4 questions per passage) and 25 multiple-passage items (two set-based double passages and three set-based triple passages with five questions per set) in Part 7. The skills measured in the reading section include grammar, vocabulary, understanding specific information, and reasoning based on information stated in the text.

As described above, it is evident that the TOEIC LR was designed to assess business English listening and reading skills. Thus, use of the test scores in a business context may be valid because the test results are an indicator of test takers' English listening and reading abilities when working in an international business context. However, use of the scores in an academic context for university admissions may need to be validated because validity evidence for this use of scores has not been provided by the test users. To address this issue, therefore, we will discuss the value implications embedded in the use of the tests in the following section.

3. VALUE IMPLICATIONS

Messick (1989) posited the unitary concept of validity under which consequential validity and the three traditional aspects of validity (content validity, predictive/concurrent criterion-related validity, and construct validity) were subsumed. In Table 1 below, the first row of the matrix, *evidential basis*, refers not only to traditional validity but also to the test relevance and utility that represent psychometric traits. The second row, *consequential basis*, indicates the social dimensions of a test including the value implications of the test construct and the social consequences of test use arising from the value implications.

TABLE 1
Facets of Validity as a Progressive Matrix (Messick, 1989, p. 10)

	Test Interpretation	Test Use
<i>Evidential basis</i>	construct validity (CV)	CV + relevance/utility (R/U)
	CV + R/U +	CV + R/U + VI
<i>Consequential basis</i>	value implications (VI)	+ social consequences (SC)

In Messick's facets of validity, *construct validity* (cell 1) refers to claims we would like to make about test takers and both the reasoning and empirical evidence that support these claims. *Relevance/utility* (cell 2) pertains to the relevance of the claims to the decision regarding test takers and support for the appropriateness of these claims based on a theoretical rationale and evidence. *Value implications* (cell 3) refer to the values and assumptions attached to the test construct, and social consequences (cell 4) indicate the *impact/consequences* resulting from the test use derived from the values embedded in the test constructs from individuals and/or social settings.

Messick's validity framework has influenced current validity arguments in that it has prompted validity researchers to consider the social dimensions of validity: it has broadened the scope of validity from its former emphasis on psychometric properties to

one that includes the appropriateness of interpretations of test scores for proposed use of tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). This can be illustrated using the two biological concepts of *in vitro* and *in vivo* (Zumbo, 2015). *In vitro* refers to the proposed interpretations and uses of test scores, and *in vivo* is defined as the stakeholders' perceived interpretations and uses of the scores. As Messick (1989) and Kane (2013) explained in their discussions of validity, gathering evidence for inferences about score interpretations and uses is an integral part of validation. Therefore, to explore the validity of score interpretations and uses, it is necessary to investigate how stakeholders perceive test scores and their uses.

However, the validity of Messick (1989) including other researchers and theorists (Kane, 2013; Zumbo, 2015) is discussed in terms of legitimate uses of a test, i.e., they do not discuss the misuse of a test. The misuse of a test may result from the values that stakeholders apply when interpreting the scores. Despite this possible result, value implications have been addressed only rarely in the literature (Fulcher, 2015), perhaps because including values when assessing validity places a burden on researchers (O'Loughlin, 2011). Fulcher (2015) noted that some researchers were still reluctant to accept value implications; they argued that the consequential basis complicated validity discussions and that it needed to be discussed in a policy debate (Mehrens, 1997; Popham, 1997), although they acknowledged the importance of this consequential basis, while others restricted the scope of values in validity arguments (e.g., Kane, 2013). McNamara (2005) argued that all language testing may be politically value-laden in that it has been initiated and used to achieve political goals (Fulcher, 2009; McNamara & Roever, 2006; Shohamy, 2001). Value implications, therefore, cannot be separated from validity discussions (Messick, 1989).

In their discussion of value implications of language tests, McNamara and Roever (2006) noted historical instances of distinguishing identity using such tests. They presented the narrative of the *shibboleth* test in the Book of Judges (Old Testament). The *shibboleth* test was used to distinguish the Ephraimites, whose dialect lacked the palate-alveolar fricative /ʃ/, from the Gileadites, whose dialect had the phoneme, by forcing the defeated enemy (i.e., the Ephraimites) to pronounce the word *shibboleth*. In the early 20th century, as McNamara and Roever (2006) noted, the Australian government used the Dictation Test to prevent unwanted applicants from immigrating to Australia: they administered a 50-word dictation in a European or Japanese language with which the applicants were not familiar.

In a discussion of educational contexts, McNamara and Roever (2006) noted the contrasting views of middle and high school teachers regarding the introduction of an English speaking test in the university entrance examination in Japan by referring to Akiyama's

(2004) study. Akiyama's study showed that middle-school teachers were in favor of implementing the oral test in the curriculum (stressing the importance of oral language skills) (Ministry of Education, Culture, Sports, Science and Technology, 1998, 1999), while high school teachers were opposed; the latter were not teaching English speaking skills in high schools because these skills were not assessed in university admissions. McNamara and Roever (2006) noted that these opposing views reflected the struggle over Japanese cultural values — diligence and hard work — because the high school teachers believed that English speaking skills could be acquired by extensive exposure to English, but this belief did not conform to the cultural values in the society (McNamara & Roever, 2006). This example shows how the introduction of a test can involve conflicting values and how different stakeholders can hold different beliefs about a test.

That diligence and hard work are valued in uses of test scores was made evident in a study by Elder (1997). She investigated university admissions authorities' score uses of the Year 12 Victorian Certificate Education (VCE) examinations in languages other than English (LOTE) in Australia for decisions about selection. She illustrated that universities appeared to disregard communicative ability in language as they saw the scores of those with a native-language background as an indicator of the benefit of mere natural exposure to the language. They compensated foreign language learners using scores on the English test as an external criterion as they placed the most value on students' academic efforts or aptitude rather than their language proficiency (Elder, 1997).

As we have seen, values are added to the meaning of test scores and the uses of test scores are affected by the values. Depending on how test users perceive test scores, decisions regarding test takers may cause unintended consequences, regardless of efforts to ensure the fairness of a test. As the notion of validity has been extended (Messick, 1989; Kane, 2006, 2013), some empirical studies of test users' perceptions of test scores and their uses were performed in the context of university admissions (Ginther & Elder, 2014; O'Loughlin, 2011, 2013).

O'Loughlin (2013) investigated admissions officers' understanding of the use of scores on the IELTS (academic module) in admissions to two Australian universities. He collected data from an online survey of 50 participants who were working in admissions, marketing, and academics and as pre-course and in-course language support staff at these universities. Follow-up interviews were conducted with 15 respondents to the survey. Two of the findings from the study were that the admissions officers focused primarily on the minimum score for enrollment in a particular program, and that universities in the United States used more than one piece of evidence related to English proficiency, whereas Australian universities generally used only a standardized English test. He argued that admissions officers needed to know what the test scores meant when they made decisions about test takers and that they needed to use multiple sources of evidence for their

decisions by considering the limitations of predicting an applicant's academic achievements based only on English test scores.

In other contexts, Ginther and Elder (2014) conducted a case study of institutional interpretations and the uses of scores on the three English tests in university admissions of international graduate students: the Test of English as a Foreign Language (TOEFL) internet-based test, the IELTS (academic module), and the Pearson Test of English (PTE). The cases included Purdue University in the United States and the University of Melbourne in Australia. The researchers investigated the admissions officers' understanding of the test scores and their uses in international graduate student admissions. Some findings were that (1) language proficiency test scores affected admission decisions to a small extent since other aspects of the applicants' file such as their academic record, oral interviews, research proposal, and letters of recommendation were considered relatively important; (2) admissions officers were not satisfied with enrolled students' English proficiency; (3) there was tension among respondents in terms of whether the cutoffs of English test scores represented adequate or minimal English proficiency; and (4) most admissions officers lacked familiarity with the English language tests.

The studies of O'Loughlin (2013) and Ginther and Elder (2014) are important because there has been little empirical research into the main views held by test score interpreters. Investigating how test users understand test scores and use them in their context is necessary in that the test users' decisions about an applicant may have significant consequences depending on their interpretations. The constructs of the TOEFL and the IELTS in Ginther and Elder's (2014) and O'Loughlin's (2013) studies are relevant to the score use contexts. However, the construct of the TOEIC LR in this study does not endorse this unrelated use of the test by the admissions officers (i.e., university admissions). In this case, there is a need to investigate the validity of score use that differs from the proposed score use by the test developer (AERA, APA, & NCME, 2014).

Therefore, we investigated the perceptions that the admissions officers have about the TOEIC LR score, an applicant's English proficiency, the selection process, and the values and intentions that are embedded in this example of test use in the context of Korean university admissions. The following research questions were addressed in this study:

- 1) How widespread is the use of TOEIC LR scores in admission to Korean universities?
- 2) How do admissions officers view TOEIC LR scores?
 - a) Do they think that TOEIC LR scores provide accurate evidence of an applicant's English language proficiency? If not, why?
 - b) Do they think that TOEIC LR scores are an accurate indicator of an applicant's academic skills? If not, why?
- 3) What values are embedded in the scores and use of the TOEIC LR?

- a) What do they think that TOEIC LR scores represent in an applicant?
- b) What do they think regarding the use of TOEIC LR scores in university admissions, and why do they use these scores?

4. METHODS

In this study, we adopted a sequential mixed-method design (Creswell & Clark, 2011) with one source (admissions officers from Korean universities) and involving three phases of data collection: (1) document analysis to provide information on how many and which Korean universities use TOEIC LR scores in their university admissions and to identify the questionnaire survey target; (2) a survey of the faculty involved in the admission procedures to elicit opinions regarding the issues considered in the research questions; and (3) semi-structured interviews with selected participants to delve into their questionnaire responses. In the following sections, we describe the methods involved including data collection procedures, the participants, instruments, and data analysis.

4.1. Data Collection 1: Information about the Admission Process in Four-Year Korean Universities

To identify the questionnaire survey target, we examined how many and which four-year Korean universities used the TOEIC LR in their admissions. In this data collection process, information about the types of English tests used, admission types, cut-off scores, enrollment quotas, and additional admission procedures was investigated by visiting 220 university websites. Application handbooks were downloaded from each of the 220 websites and were thoroughly examined between August 15th and September 10th, 2013.

4.2. Data Collection 2: Participants and Instruments

The second data collection phase spanned a month-long period from September to October 2013, and the third phase was performed during one week in October 2013. The activities involved in each phase are shown in Table 2 below.

TABLE 2
Overview of Research Activities During the Two Phases

Phase	Time	Sources and Methods
Two	September to October 2013	Questionnaires sent to admissions officers at several universities
Three	One week in early October 2013	Follow-up semi-structured interviews with 5 admissions officers who responded to the questionnaire

We administered the questionnaire via email after asking for written consent, and we collected the completed questionnaires via email. To conduct the semi-structured interviews, the first author visited Korea and conducted the individual interviews with the five participants between September 29th and October 4th, 2013.

4.2.1. Questionnaire and interview participants

The data were collected from academic staff members who were working as admissions officers at the 16 universities where TOEIC LR scores were being used as one of the admission criteria. In total, 62 admissions officers who were acquainted with the first author were asked to voluntarily participate in the questionnaire in an email that asked whether they had work experience as an admissions officer. This step resulted in 20 of the 62 admissions officers participating in the questionnaire (a 32% response rate). Twelve of the participants were male, with an average age of 50 years and 9 months (Min = 38, Max = 61, $SD = 6.42$), and 8 were female, with an average age of 44 years (Min = 31, Max = 49, $SD = 5.72$). On average, they had 3 years and 10 months of experience in admissions at their university (Min = 2, Max = 10, $SD = 2.47$). They were all English language-related academic department staff members. The academic staff were temporarily appointed as admissions officers during the admission season. They gave marks for the interview/essay portion of the students' applications and made the final decisions regarding admissions. The administrative staff collected the scores, determined the ranking, and then notified the applicants of the result (i.e., acceptance or rejection). The universities where the admissions officers were working were located across Korea.

Five questionnaire respondents from phase two of the data collection portion were recruited as interviewees because their university implemented additional procedures for admission such as interviews and essay writing in English.

To differentiate the participants, each pseudonym started with (a) capital letter(s) "Q" or "IQ," indicating a questionnaire or interview/questionnaire respondent, respectively; followed by (b) the letter "F" (female) or "M" (male) to identify gender; and finally, (c) an identification number (1 through 15 for only the questionnaire respondents and 1 through 5 for both the questionnaire and interview respondents).

4.2.2. Questionnaire design and semi-structured interviews

The questionnaire items in this study were designed based on the items in O'Loughlin's (2013) study and in previous studies on test use (Elder, 1997; McNamara & Roever, 2006). Because O'Loughlin's (2013) study pertained to admissions officers' understanding of the English test scores (i.e., the IELTS) and their use of the scores in university admissions, we

selectively adapted the potential items in relation to background information about gender, age, career history, and setting cut-off scores. The other items were created based on the previous studies on values that were embedded in test scores.

The questionnaire in the study originally consisted of 17 items consisting of four short-answer questions concerning gender, age, and career history (items 1 through 4), seven multiple-choice items (e.g., 1 = strongly disagree; 5 = strongly agree), and one open-ended item. Five questions were a mix of multiple-choice and open-ended questions asking the participants the reason for their answer choice.⁵ (Items 5 through 9⁶ are not included because they were not related to the purpose of this paper.) The items were grouped into five categories: background information (questions 1 through 4), TOEIC LR scores as an indicator of English proficiency (questions 10 through 12) and academic skills (questions 13 through 15), intelligence, diligence and effort (question 16), and overall opinions about test use in admissions (question 17). Table 3 below shows how the questionnaire items relate to each of the research questions.

TABLE 3
Research Questions and Their Corresponding Questionnaire Items

Research Question	Questionnaire Item(s)
2) How do admissions officers view TOEIC LR scores?	
a) Do they think that TOEIC LR scores provide accurate evidence of an applicant's English language proficiency? If not, why?	Q 10–12
b) Do they think that TOEIC LR scores are an accurate indicator of an applicant's academic skills? If not, why?	Q 13–15
3) What values are embedded in the scores and use of the TOEIC LR?	
a) What do they think that TOEIC LR scores represent in an applicant?	Q 16
b) What do they think regarding the use of TOEIC LR scores in university admissions, and why do they use these scores?	Q 17

Based on the questionnaire, semi-structured interviews were conducted to clarify and elaborate on the participants' questionnaire responses. Each interview took 20 to 45 minutes, with an average of 33 minutes depending on the participants' available time and enthusiasm for the discussion of the issues brought up during the interviews. The interviews were conducted in Korean, recorded with the participants' written consent, and then translated into English by the first author.

⁵ The questionnaire items were written in Korean. The first author translated the items into English.

⁶ These questions pertain to the types of accepted English tests and the role of the admissions officer in setting cut-off scores and assigning marks.

4.3. Data Analysis

The questionnaire responses were analyzed in terms of comparative frequencies and percentages of responses to the multiple-choice questions. For the analysis of responses to the open-ended questions and the interview data, categories were developed for coding each segment after the segmentation was completed. For this coding, we recorded themes that emerged from the data and grouped similar themes into superordinate themes. Salient quotations that reflected these themes were then transcribed and coded in terms of whether they were representative of all interviewees or a particular sub-group or were a unique comment.

5. RESULTS

Findings from the analysis of documents for RQ 1, and from the analysis of the questionnaire and interview data for RQ 2 and 3 are presented under each research question. Specifically, the headings of each section represent (sub)research questions are described as noun phrases for conciseness

5.1. The Extent of the Use of TOEIC LR Scores for Admission to Korean Universities (RQ 1)

One hundred nine of the 220⁷ universities in Korea implemented special admission procedures, and information about admissions at two of the universities could not be determined because it was not included in their application handbook. Four of the 107 universities did not accept English test scores in the admission process because they implemented their own screening procedures such as interviews and/or essay writing and did not consider English test scores at the first stage of the admission process (i.e., document review). Three of the remaining 103 universities did not provide specific information about the type of English test scores the applicants could submit (it appeared that all types of English test scores were accepted). The types of English tests used at the 103 universities, including these three universities, consisted of the TOEFL, the TOEIC, the TEPS, the IELTS, the G-TELP⁸, the SMU-MATE⁹, the PELT¹⁰ and the FLEX¹¹, and

⁷ This total number of universities in Korea comes from the Korean Council for University Education as of October 2013.

⁸ The General Tests of English Language Proficiency (G-TELP) was developed by the International Testing Services Center.

⁹ The Sookmyung University-Multimedia Assisted Test of English (SMU-MATE) was developed by Sookmyung Women's University, and this test was used only by this university for admissions. The

99 of the 103 universities commonly used the TOEIC test score in their admission process. It was also ascertained from the handbooks that 16 of the 99 universities had not set cut-off scores, whereas the scores set by the 83 remaining universities varied from 600 to 950 points on the TOEIC test, depending on the department and university. It appeared that the lower the reputation of the university, the lower they set their cut-off score.

Because special admissions were conducted on a rolling basis, the enrollment quota was small, but the total number of students to be admitted was approximately 4,600 among the 107 universities. Eighty-one universities used other admission criteria (interviews at 70, essay writing at nine, and aptitude tests at two universities) in addition to the English test scores. Based on a review of the application handbooks of all the universities, 99 of the 109 universities used special admissions but also used TOEIC scores in their admissions, with 81 of these 99 universities using additional admission procedures.

5.2. Admissions Officers' Views on TOEIC LR Scores (RQ 2)

5.2.1. Views on TOEIC LR scores regarding English proficiency (RQ 2a)

Responses to research question RQ 2a were elicited by questions 10 through 12, with related follow-up interview questions. Most of the respondents somewhat disagreed that the TOEIC LR scores provided accurate evidence of English proficiency, as shown in Table 4.

TABLE 4
Responses of Admissions Officers Regarding the TOEIC LR
as Accurate Evidence of English Proficiency

Response	Number of Responses (<i>n</i> = 20)
1. Strongly disagree	1 (5%)
2. Somewhat disagree	14 (70%)
3. Neither agree nor disagree	4 (20%)
4. Somewhat agree	1 (5%)
5. Strongly agree	0
Total	20

Fourteen participants indicated that they thought the test was not an exact indicator of

test assesses only English speaking and writing proficiency.

¹⁰ The Practical English Language Test (PELT) was developed by the Korea Foreign Language Evaluation Institute and consists of four sections: reading, listening, speaking, and writing.

¹¹ The Foreign Language Examination (FLEX) was developed by Hankuk University of Foreign Studies, and this test is used only by this university for admissions. The FLEX assesses proficiency in 7 languages: English, Chinese, Japanese, French, German, Russian, and Spanish; it assesses reading, listening, speaking and writing skills.

English proficiency because it only assessed English reading and listening proficiency. This belief was indicated by the responses to question 11: “*Do you think it is necessary for an applicant to submit other English test scores that assess English speaking and writing proficiency instead of a TOEIC score?*” As Table 5 below shows, aside from the six non-responses, most of the respondents thought that the other tests, which assessed practical skills (i.e., speaking and writing ability), were necessary.

TABLE 5
Responses of Admissions Officers Regarding the Necessity of Other Tests
Including Speaking and Writing Assessments

Response	Number of Responses (<i>n</i> = 20)
1. Completely unnecessary	2 (10%)
2. Unnecessary	0
3. Moderately necessary	0
4. Necessary	0
5. Very necessary	12 (60%)
No response	6 (40%)
Total	20

In the responses to the follow-up open-ended question asking why the participants expressed a need for other tests, it was commonly stated that there was a low correlation between the TOEIC LR scores and practical English language skills. One participant stated:

When looking at the real state of currently enrolled students’ academic ability, it is necessary to use other measures to examine their English proficiency via other English tests or other types of college exams because I’ve seen many students whose speaking proficiency is remarkably low; they are not able to engage in basic communication in English even though their TOEIC LR score is substantially high. (IQF1)

As indicated in this excerpt, most admissions officers did not consider TOEIC LR scores to be an accurate reflection of the applicants’ practical English language skills because they frequently encountered situations where high TOEIC LR-scoring students were not competent in English speaking and writing. In response to question 12, which asked whether other stages in the admission process were used to evaluate an applicant’s English proficiency, 16 of the 20 admissions officers reported that they administered other procedures such as English interviews or English essay writing. When thoroughly examining the 107 universities’ application handbooks (two universities did not publish application handbooks on their websites), it was found that 81 universities used their own tests, as mentioned in section 5.1. This was a result of the distrust in the TOEIC LR scores

as accurate evidence of an applicant's English proficiency. One interview participant whose response was representative of the 16 admissions officers stated:

We conduct English interviews along with considering the test score [...] To compensate, we assess the applicants' speaking proficiency because we run into applicants whose TOEIC LR scores are 950, but they are not good at speaking. Therefore, we evaluate the speaking ability [...] as a sort of supplement. (IQM4)

5.2.2. Views on TOEIC LR scores regarding academic skills (RQ 2b)

Research question RQ2b was addressed in questions 13, 14 and 15. Question 13 asked whether the respondents thought that TOEIC scores could be an indicator of an applicant's academic skills. Eighteen of the 20 respondents answered that the test did not reflect an applicant's academic skill (i.e., *No* to question 13), and 2 reported that the test did (i.e., *Yes* to question 13). However, 2 of the 18 who answered *No* changed their answer to *Yes* in the follow-up interviews.¹² Therefore, the total number of participants who answered *Yes* was four.

A representative response from three interview participants who answered *No* to this question came from participant IQM4, who reported that there were cases of students who were admitted to the university under the special admissions clause but performed poorly; some even failed.

[...] Because their TOEIC LR score is high enough, it can be predicted that their academic performance will be good. However, it is not like that. [...] It is likely to be good if we create a policy of supplementation because there are cases where those who were admitted with high TOEIC LR scores do not have good academic results and even dropped out. The TOEIC LR does not ensure academic achievement anyway. (IQM4)

In contrast, one questionnaire respondent out of the four who said *Yes* stated that he thought the score was an indicator of academic ability because in Korea, students did not learn English as a language but rather as a subject in school:

It has been our tradition that English is only a school subject like math, social

¹² They stated that they misunderstood the Korean word *Hakepnunglyek* as meaning "academic skill." To ensure the validity of the questionnaire responses, member checks were performed (i.e., having the participants check their responses).

studies, physical education and so on. That means that we have treated learning English as storing knowledge on the language. We are unaware that language learning is never confined to collecting grammar and vocabulary in the brain. Unlike other subjects, it needs constant practice. At any rate, the Korean style of language teaching is recognizable as a means of reflecting one's scholastic aptitude for future study. (QM4)

The participant thought that studying a foreign language was the same as studying other subjects at school and required the same scholastic aptitude. This is a common feature of English learning and education in English as a foreign language (EFL) in certain countries.

Based on the interview data from two participants (IQF1 and IQF5), they viewed TOEIC LR scores as a reflection of academic ability because they were teaching in the English department, and the test measured English proficiency. Thus, they viewed the test score as an indicator of academic skills in an English-related course. In the interview, participant IQF1 stated:

[...] As an example, in the English department, more than 80% of classes are conducted in English. Because students are studying with English textbooks, almost 100%, it would be difficult if they were not equipped to study at the college beforehand. Otherwise, they would feel difficulty in reading and understanding the classes. For their major, normally, a minimum English ability required should be present. [...] (IQF1)

Question 14 in the questionnaire asked, “*Do you think it is necessary to use other screening procedures to assess an applicant's academic skills?*” Seventeen of 18 respondents (there were two non-responses), including four who answered *Yes* to question 13, reported that there was a need to adopt other procedures to evaluate an applicant's academic ability. In response to question 15 asking whether “*they have other screening procedures in admission processes for assessing an applicant's academic ability,*” 19 admissions officers answered that they already had such procedures in place. The types of procedures used most often were *in-depth interviews about the academic major* (18%), *consideration of the College Scholastic Ability Test score* (27%), *consideration of high school records* (30%), *Korean essay writing (university's own written admission test)* (9%), *an award-winning career* (9%), and *other* (3%) (including one aptitude test and two oral interviews). In response to open-ended question 15–2 regarding the necessity of these procedures, all the respondents, apart from one non-response, stated that these processes were necessary and that TOEIC LR scores did not sufficiently reflect students' academic abilities. One respondent described his opinion in the questionnaire as follows:

I think it is desirable to involve a consideration of high school records, examination of basic knowledge or aptitude about the major as part of admissions to evaluate the applicant's cognitive ability, such as numerical ability and logic, because the TOEIC LR as the acknowledged English test does not assess this ability. (QM10)

Among the responses, three participants provided a different reason for adopting the procedures. They reported that the procedures were required because it was difficult to ascertain the academic skills of applicants who lived in English-speaking countries despite achieving high TOEIC scores. In other words, the admissions officers believed that the applicants' high scores were influenced by extensive exposure to English.

5.3. Admissions Officers' Values Embedded in Scores and Use of the TOEIC LR (RQ 3)

5.3.1. Admissions officers' values added to TOEIC LR scores (RQ 3a)

Research question RQ 3a was addressed in question 16 "*Do you think that TOEIC scores represent an applicant's intelligence, diligence, or effort?*" Nine admissions officers stated that the score represented an applicant's intelligence, diligence, and effort, whereas 11 answered that it did not. Five of the nine respondents reported that an applicant's effort could be revealed by the score because the applicant prepared for the test beyond only reviewing the curriculum. Furthermore, three of these five respondents stated that the score indicated intelligence as well as effort, as demonstrated in the excerpt below:

An applicant's effort, diligence, and intelligence should be considered favorably if the applicant obtains a high score on the TOEIC test in an EFL setting: English is so different from Korean, and there are many components for learners to do, such as memorization. (IQM3)

Five of the 11 participants who answered *No* (the test did not represent the attributes) viewed the score as an indicator of effort but not of a broader set of traits. One response that was representative of the five responses is as follows:

I think that if a student puts a lot of effort into getting a high score, [...] his/her diligence, effort, and intelligence need to be appreciated. Nevertheless, it may not be appropriate to view a TOEIC score as a barometer of the applicant's overall intelligence and diligence. (IQF2)

In contrast, the rest of the 11 respondents reported that the TOEIC LR scores reflected only English ability and that a high score can be achieved with residence in an English-speaking country and with test-taking tips learned at a private academy:

I do not think that an applicant's effort and intelligence played a significant role in achieving a high score if the applicant had lived in an English-speaking country because he/she could naturally acquire English [...]. I think that the stay in the English-speaking country and private education had a dominant effect on high scores when I meet the applicants. (QM12)

5.3.2. Admissions officers' views on the use of TOEIC LR scores in university admissions (RQ 3b)

Research question RQ3b was addressed in open-ended question 17, "*What do you think regarding the use of TOEIC scores in admission, and why do you use the score?*" Fourteen of the 20 respondents thought that it was not desirable to use the TOEIC LR in admissions, and 10 of the 14 suggested instead using the TOEFL or other tests, which assess academic English proficiency. Six out of these 10 reported that the reasons for using the TOEIC LR in admissions were that the fee for taking the TOEIC LR was significantly lower than that for the TOEFL, there were many more opportunities to take the TOEIC LR than the TOEFL, and many other institutions were using TOEIC LR scores. The following response by participant IQM3, which best represents the 14 admissions officers' responses, illustrates his opinions regarding the inappropriate use of the test:

[...] The TOEIC LR focuses only on business English proficiency in all respects. It is only for graduates and job seekers in society. Nonetheless, the test is used in admissions, which is not appropriate. Furthermore, if we review the reading section of the test and find that the test content is business oriented, the content appears to be strange for high school students and extremely different from the curriculum that they encounter in their classes. [...] In a sense, it is hard to say that the TOEIC LR scores are an accurate indicator of an applicant's English ability. Additionally, only office workers may need the test, and it is not appropriate to require high school students to submit their test scores. [...] (IQM3)

As is evident in the above extract, it was posited that it was not appropriate to use the test, because the original purpose, content, and target population of the test do not correspond to an academic setting. The excerpt below typifies the 10 respondents' views.

[...] I think that the TOEIC LR is used because the students and the university feel it is easy to take the test and use the test scores, respectively. Of course, the fact that there is no acknowledged English measurement to use is the most important reason. I feel that there is a need for alternatives to the TOEIC LR. (QF11)

These 10 respondents expressed the opinion that other English tests such as the TOEFL and the NEAT need to be adopted instead of the TOEIC LR. The following excerpt discusses the reason the TOEIC LR is used and is representative of six of the respondents' views.

In terms of the content of the TOEIC LR, it is not seen to be desirable to use the test as an indicator of an applicant's English proficiency compared to the TOEFL, but because it is useful for relieving the applicants' burden in terms of the test fee and testing opportunities, it appears that many universities are using the test in admissions. (QM10)

In contrast, six participants reported that there was a positive aspect to the test's use because the TOEIC LR also measures an applicant's general English proficiency. However, they also stated that it would be better if other tests were used:

I think that although the test does not assess English proficiency in an academic setting, the TOEIC LR can reflect an applicant's general reading and listening ability. [...] When it comes to test difficulty, because it is more difficult for students who have been taught English in the curriculum and take the TOEFL or the TEPS to score well, they cannot obtain a good result without private education, such as at an academy or with private tutoring. Therefore, I think that a TOEIC score reflects English ability to some extent because an authorized English test can be a practical means. (QF9)

Interestingly, one interview participant reported that the test use was related to post-graduation employment rates. As shown in the excerpt below, the participant expressed the view that a student could enjoy an advantage in employment opportunities if he/she receives a high TOEIC score. Furthermore, she stated that the employment rate played an important role as an indicator in a university evaluation by the government.

[...] In a sense, it (the use of the TOEIC score in admissions) is related to employment prospects after graduation. [...] In that case, a TOEIC score is

required. [...] However, students majoring in science and engineering have difficulties getting a job because their TOEIC scores are below 800 or 850. Therefore, I think it is not misleading to consider the correlation between the test score and employment. This applies to students whose major is English as well. [...] In the application review process, if the applicant's TOEIC score is below 800, an applicant is not eligible for hiring. Therefore, there are many universities that run TOEIC preparation classes in general education. [...] Thus, we do not need to care about employment if we select a student whose TOEIC score is high enough. [...] Therefore, it is not wrong to select students using the test score in the Korean context. [...] To not be weeded out from the university, students should not drop out, and the dropout rate should be low. [...] If students drop out of school, the university may be in financial trouble. Then, what is an admissions process that achieves a low dropout rate? Students who are admitted through rolling admissions do not drop out, while students admitted through regular admission do. Furthermore, as the Ministry of Education demands, we do care about future employment and students not being weeded out and must consider what type of admissions we need to [best] select students. (IQF5)

6. DISCUSSION

In this study, we investigated, as evidence of test validity, admissions officers' perceptions of TOEIC LR scores and their uses in special college admissions in Korea. One reservation of this study is that it may be difficult to generalize the findings to the entire population of admissions officers in Korea because the sample size was relatively small. However, this study tried to tackle and explore in depth the validity issues of test score use in an unrelated context and the test's purpose because there has been little empirical research into test users' perceptions in Korea. In the following sections, the findings from this study are discussed in light of the research questions.

6.1. Prevalence of Use of TOEIC LR Scores for Admission to Korean Universities (RQ 1)

In the special college admissions, various English tests were used including those that each university developed for its own admission process. However, almost half of the Korean universities (99 of 220) were using the TOEIC LR scores in their university admissions, although the test was designed to measure business English proficiency in the

workplace. This particular use may have arisen because the test was developed by a recognized testing enterprise, the ETS, because it was convenient in terms of timing and test fees and because few reliable, alternative English tests had been developed. Because the TOEIC LR has been enormously popular in Korea (it also accounts for the largest percentage of the test-taking population worldwide, approximately two million people [Munhwa Broadcasting Corporation, 2016]) since its introduction in the country in 1982, the test has been used in many contexts including academic contexts. According to the TOEIC council in Korea and the ETS website, the fee for taking the TOEIC LR was approximately \$40 U.S., whereas the fee for taking the TOEFL was \$170, i.e., more than 4 times as expensive. Furthermore, in Korea as of late 2013, the number of sites for taking the TOEIC LR was 286, whereas the sites for taking the TOEFL numbered only 27 (approximately one-tenth that of the TOEIC LR). It was thus easier for test takers to select the TOEIC LR based on its fee and number of testing sites.

6.2. TOEIC LR Scores Regarding English Proficiency and Academic Skills (RQ 2)

As indicated by the supporting documents for the TOEIC LR, including the discussions of test review and the test's purposes, content, and tasks for assessing English language abilities, this test pertains to business English proficiency. Based on Messick's matrix (i.e., cell 2), our findings indicate that the use of the test in an academic admission context lacks validity. This lack of validity is clearly exemplified by the participants' comments indicating that these admissions officers did not view the test score as an accurate indicator of English proficiency: "*We run into applicants whose TOEIC score is 950 but who are not good at speaking*" (IQM4), and "*I've seen many students whose speaking proficiency is remarkably low and who are unable to engage in basic communication in English, although their TOEIC scores are substantially high*" (IQF1). These results are consistent with those of Ginther and Elder (2014) and O'Loughlin (2013) in that the test users were not satisfied with their university students' English proficiency.

In response to question 13, four respondents reported that the test score reflected an applicant's academic ability. Due to their position in English-related departments, some of the respondents thought that English skills were necessary to successfully study the offered courses. Their view of TOEIC LR scores as an indication of academic ability stemmed from their conception of English ability as necessary for students to study their major and from their identification of ability with scholastic aptitude. One participant (QM4) reported that since Korean students learned English as a subject in school, the test score thereby indicated scholastic aptitude, reflecting the nature of English education in an EFL setting.

However, most of the admissions officers did not trust the test score as a gauge of

academic ability: one stated that *“those who were admitted with a high TOEIC score did not obtain a good academic result and even dropped out”* (IQM4). This view was also supported by the finding in section 5.3 in response to questions 14 and 15. All the respondents stated that other procedures for assessing academic ability were necessary in their responses to question 14, and 17 of the 18 respondents reported they used other procedures to evaluate an applicant’s academic ability, such as *“consideration of the College Scholastic Ability Test score (18%),” in-depth interviews about the academic major (18%), and consideration of high school records (30%)* in response to question 15-1.

There must have been a need to use additional procedures similar to universities in the U.S. and Australia (Ginther & Elder, 2014) to evaluate an applicant’s English and academic skills. However, implementation of such additional procedures may be because of distrust of the TOEIC LR scores as an indicator of individuals’ English speaking and writing proficiency (J. Kim, 2006) because there had been criticism of employees’ lack of English speaking and writing skills regardless of their high scores on the TOEIC LR (Hirai, 2009; Powers, 2010). Although the ETS introduced the TOEIC Speaking and Writing tests to address this issue, the TOEIC LR was still used most for university admissions. This is discussed in the following section.

6.3. Values Attached to Scores and Uses of the TOEIC LR (RQ 3)

The value implications of the respondents’ use of the test scores were also expressed in their responses to question 16 in the questionnaire. To examine the attached meaning to test constructs, the researchers questioned whether they regarded the test as a mirror of an applicant’s intelligence, efforts, or diligence based on the previous discussions of test use in a social context (e.g., McNamara & Roever, 2006). Although nine respondents thought that TOEIC LR scores reflected these attributes while 11 did not, five of the 11 reported that the test reflected an applicant’s efforts but not the entire set of attributes. Therefore, 14 of the 20 respondents thought that the test scores reflected an applicant’s efforts.

In educational settings in Korea, students have few opportunities to improve their English speaking and writing skills because they focus on preparing for the English component of the CSAT test, which assesses English reading and listening comprehension only. Therefore, preparation for English tests such as the TOEIC LR, TOEFL, or TEPS is left to the test takers themselves. In this way, it appeared that the 14 admissions officers thought that the TOEIC LR scores represented an applicant’s academic effort to some extent.

However, as described in section 5.3.1, six of the 11 participants thought that a high TOEIC LR score could be achieved by exposure to English by residing in an English-speaking country or learning test-taking skills at a language academy, which did

not conform to the cultural values in Korean society.

In the discussion of values, various social factors regarding test use were expressed in the responses to question 17. The admissions officers could not avoid using the TOEIC LR because of prevailing social conditions and requirements, although they thought the implementation of other tests was needed. First, the TOEIC LR was the test that test takers might feel most comfortable with because it was relatively inexpensive and was perceived as easier than other English tests since it only assessed English reading and listening proficiency.

Second, the TOEIC LR is accepted by most companies and institutions and therefore also plays an important role in employability after graduation; employment rates, in turn, play a significant role in university evaluations and accountability, which makes the TOEIC LR important for securing government support for the university. Although this second social factor was addressed in only one participant's (IQF5) response, it was expected that most universities in Korea were concerned with their post-graduate employment rates because they were evaluated by the government annually. The most important factors in these government evaluations are financial stability, the employment rate (Lee, 2013), dropout rate, and quota of English lectures. As indicated by participant IQF5's comments regarding the enrollment quota, the financial stability of the university is one of the most important elements in the evaluation by the government. When the enrollment quota exceeds the number of high school graduates, it is important for universities to have a high percentage of full-time students registered. This association of various social factors and conditions with test use, as shown in this study, represents a divergence from previous theoretical understandings of validity (Messick, 1989) and discussions of values (McNamara & Roever, 2006) in that those earlier authors discussed values as an aspect of the construct of a test. What emerged from this study is that we may also need to consider the values embedded in *test use* and not only those represented in the column of *test interpretation* in Messick's matrix.

7. CONCLUSIONS

This study explored admissions officers' perceptions of the TOEIC LR via questionnaires and interviews in the context of test use for university admissions. It also explored the values underlying the test's use in a context unrelated to the appropriate score uses recommended by ETS. It appears that admissions officers were in fact sensitive to the social pressures placed on applicants and accepted the use of the TOEIC LR as a just course of action in the circumstances even though they had a somewhat negative perspective of the test. This finding indicated that the test use was indeed interwoven with

social demands and factors, which confirmed current perspectives of test validity (Bachman, 2005; Bachman & Palmer, 1996, 2010; Fulcher, 2015; Kane, 2013; Lynch, 2001; Messick, 1989) regarding the importance of the social dimensions of testing.

We identified a use of the TOEIC LR that was unlikely to have been included in its validation because the test users were using it for purposes other than those listed by the ETS (2013) as described in section 2. Although they would argue that this use of the test can be justified by the limitations of time, money, and human resources in terms of practicality, the test users need to consider that this use may have negative consequences in educational settings and beyond and may encourage high school students (and middle-school students) to learn business English (which adults learn when looking for a job). In turn, these potential consequences may threaten the validity of the test scores in this particular use. Accordingly, we urge university authorities to reconsider uses of TOEIC score for university admissions, and to be responsible for providing evidence for the test consequences.

REFERENCES

- Akiyama, T. (2004). *Introducing speaking tests into a Japanese Senior High School entrance examination*. Unpublished doctoral dissertation, The University of Melbourne.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1–34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Booth, D. K. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity*. Unpublished doctoral thesis, The University of Auckland, New Zealand.
- Creswell, J., & Clark, V. (2011). *Designing and conducting mixed methods research* (2nd ed.). London, England: Sage.
- Educational Testing Service. (2013). *User guide listening and reading*. Retrieved from https://www.ets.org/Media/Tests/Test_of_English_for_International_Communicati

- on/TOEIC_User_Gd.pdf
- Educational Testing Service, (n. d.). *2016 Sin Thoik Silhayng* (New version of the TOEIC LR implemented). Retrieved from <http://exam.ybmnet.co.kr/toEIC/info/new/overview.html>
- Elder, C. (1997). What does test bias have to do with fairness? *Language Testing*, *14*(3), 261–277.
- Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics*, *29*, 3–20. doi:10.1017/S0267190509090023
- Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. New York: Routledge.
- Ginther, A., & Elder, C. (2014). *A comparative investigation into understandings and uses of the TOEFL iBT® Test, the International English Language Testing Service (Academic) Test, and the Pearson Test of English for Graduate Admissions in the United States and Australia: A case study of two university contexts* (TOEFL iBT Research Report No. 24, ETS Research Report No. RR-14-44). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12037
- Hirai, M. (2009). Correlation between STEP BULATS speaking and TOEIC® scores. In E. Skier & T. Newfields (Eds.) *Proceedings of the 8th Annual JALT Pan-SIG Conference*. Retrieved from <http://jalt.org/pansig/2009/HTML/Hirai.htm>.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: American Council on Education/Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73.
- Kim, J. (2006, July 28). Silmwu Yengewa Tongttelecin Thoik Silhyoseng nonlan ... Kiep Pwulsin Phayngpay (Companies' distrust of the TOEIC LR due to lack of authentic business English). *Kookminilbo*. Retrieved from <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&oid=005&aid=0000252479>
- Lee, B. (2013, October 18). Naynyenpwute tayhacengwuen kamchuk ... chwoyhawuykulwup 'thwoychwul' (University enrollment will be cut from next year, 'weeding out' lowest ranking universities). *Seoul Sinmun*. Retrieved from <http://www.seoul.co.kr/news/newsView.php?id=20131018010005>
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, *18*, 351–372.
- McNamara, T. (2005). 21st Century shibboleth: Language tests, identity and intergroup conflict. *Language Policy*, *4*(4), 351–370.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. London: Blackwell Publishing.

- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practices*, 16(2), 16–18.
- Menken, K. (2008). High-stakes tests as de facto language education policies. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education, 2nd edition, volume 7: Language testing and assessment* (pp. 401–414). New York: Springer.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education & Macmillan.
- Ministry of Education, Culture, Sports, Science and Technology. (1998). *Chugakuko Gakushu Shidoyoryo*. (The course of study for junior high schools). Tokyo: Kairyudo.
- Ministry of Education, Culture, Sports, Science and Technology. (1999). *Kotogako Gakushu Shidoyouyou-Kaisetu* (The course of study for senior high schools). Tokyo: Kairyudo.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477–496.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of assessment arguments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62.
- Munhwa Broadcasting Corporation. (2016, May 24). *Pakkwuyn Thoik Cekunghalya Sinsupheyk kacchwulya Swuhemsayng 'icwungko'* (Double whammy for test-takers as they adjust to new TOEIC LR). Seoul: Munhwa Broadcasting Corporation. Retrieved from http://imnews.imbc.com/replay/2016/nwdesk/article/3962233_19842.html
- O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8, 146–160.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380.
- Popham, W. J. (1997). Consequential validity: Right concern—Wrong concept. *Educational Measurement: Issues and Practices*, 16(2), 9–13.
- Powers, D. E. (2010). *The case for a comprehensive, four skills assessment of English language proficiency* (TOEIC Compendium TC-10-12). Retrieved from http://www.ets.org/Media/Research/pdf/RD_Connections14.pdf.
- Powers, D. E., & Powers, A. (2015). The incremental contribution of TOEIC® Listening, Reading, Speaking, and Writing tests to predicting performance on real-life English language tasks. *Language Testing*, 32(2), 151–167.
- Powers, D. E., Kim, H.-J., & Weng, V. X. (2008). *The redesigned TOEIC® (listening and reading) test: Relations to test-taker perceptions of proficiency in English* (ETS

- Research Report RR-08-56). Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2008.tb02142.x/epdf>
- Schedl, M. (2010). *Background and Goals of the TOEIC® Listening and Reading Test Redesign Project* (TOEIC Compendium TC-10-02). Retrieved from Educational Testing Service: <https://www.ets.org/Media/Research/pdf/TC-10-02.pdf>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London, UK: Pearson.
- Zumbo, B. D. (November, 2015). *Consequences, side effects and the ecology of testing: Keys to considering assessment 'in vivo'*. Keynote address, the annual meeting of the Association for Educational Assessment– Europe (AEA-Europe), Glasgow, Scotland.

APPENDIX A

Interview Items

Opening questions:

- How long have you worked as an admissions officer?
- Have you worked as an admissions officer on a temporary basis during the admission period?
- Did you work during every admissions season?
- How many English proficiency tests is the university using in admissions?
- What test scores do the applicants mainly submit?

Questions pertaining to an applicant's English proficiency:

- What do you think of the relationship between the TOEIC score and overall English proficiency?
- What do you think regarding the use of other English tests to assess practical skills such as speaking and writing? Why do you think such tests are necessary (or unnecessary)? What is your reason?

Questions pertaining to TOEIC as an indicator of an applicant's academic skills:

- What do you think regarding the relationship between the TOEIC score and academic skills?

Questions pertaining to the use of TOEIC test scores:

- Is there any reason to use TOEIC test scores in admissions?
- What do you think regarding the test use situation in general?

Applicable levels: Tertiary

Gwan-Hyeok Im
Faculty of Education,
Queen's University
Duncan McArthur Hall, Kingston, K7M 5R7, Canada
Phone: 1-613-533-600
Email: gwan.im@queensu.ca

Tim McNamara
School of Languages and Linguistics
Room 619, Babel Building
The University of Melbourne, Victoria 3010, Australia
Phone: 61-3-8344-4207
Email: tfmcna@unimelb.edu.au

Received on March 1, 2017

Reviewed on April 15, 2017

Revised version received on May 15, 2017