

Testing in Social, Cultural, and Political Contexts

Gwan-Hyeok Im*

Im, Gwan-Hyeok. (2021). Testing in social, cultural, and political contexts. *English Teaching*, 76(3), 3-33.

Despite the popularity of the TOEIC in the Korean society for over 30 years, few studies have investigated the understanding and usage of TOEIC scores in the Korean context. This research gap needs to be filled to provide test users with useful information in the Korean context. Using an argument-based approach to validation, this study investigates the meanings and uses of the TOEIC scores in the Korean context, based on analyses of records publicly available at the YBM website (i.e., TOEIC administrator in Korea). One hundred ninety-four documents published between 2012 and 2017 were collected from the website and the data were analyzed by combining the content and the context analyses. Findings reveal that contextual factors affect the understanding and usage of scores by different TOEIC stakeholders. The document data reveal unintended meanings and uses of the scores.

Key words: YBM, TOEIC, an argument-based approach, document analysis, test validity and validation

*Author: Gwan-Hyeok Im, Researcher, Wonju University Innovation Support Project Team, Yonsei University; 1 Yonseidae-gil, Wonju, Gangwon-do 26493, Korea; Email: gwan.im@queensu.ca

Received 20 May 2021; Reviewed 29 April 2021; Accepted 16 September 2021



© 2021 The Korea Association of Teachers of English (KATE)

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits anyone to copy, redistribute, remix, transmit and adapt the work, provided the original work and source is appropriately cited.

1. INTRODUCTION

Ever since Messick's (1989) seminal work on validity was introduced, the concept of validity in testing has been expanded to include main interpreters' understanding and uses of test scores. To describe this particular aspect of validity Zumbo (2015) applied biological concepts of *in vitro* and *in vivo* to testing, the former referring to a virus in a glass at the laboratory and the latter to one in living organisms. That is, *in vitro* corresponds to meanings and uses of test scores intended by testing organizations and this can be understood and used in different ways *in vivo* corresponding to perceived meanings and uses of the test scores when the test is fully administered in a particular context. With these concepts adopted in testing, research on *in vitro* and *in vivo* aspects of validity has been widely conducted (e.g., Im & McNamara, 2017; Schmidgall, 2017a, 2017b).

In this study, the concept of "*in situ*" is adopted to be located between *in vitro* and *in vivo* to investigate how stakeholders perceive the score meanings, score uses, and consequences of the Test of English for International Communication (TOEIC) through YBM (i.e., the TOEIC administrator).

The TOEIC is the most widely used test in the Korean context for over 30 years for employment and promotion purposes and it has been recognized as the most high-stakes test in Korea (Im & Cheng, 2019). The test was developed by Educational Testing Service (ETS) in 1978 and its alternate forms (TOEIC Listening and Reading (TOEIC LR), TOEIC Speaking, and TOEIC Writing) have still been produced by the ETS; however, since Korea is the place that the TOEIC is most exclusively used, it has particular localized meanings, uses, and consequences. Specifically, despite the popularity of the TOEIC in the Korean society for over 30 years, there has been little research on how TOEIC scores are understood and used in the Korean context (e.g., Im, 2021) and how those have been localized in Korea. Using the operational validation model based on an argument-based approach (Bachman & Palmer, 2010; Chapelle, Enright, & Jamieson, 2008; Kane, 1992, 2006, 2013), this study addresses one research question as follows:

1. What are stakeholders' understanding and uses of TOEIC scores for selection decisions and consequences of such uses in Korean context based on the specified inferences for this study?

Seven inferences are looked at and discussed within the operational validation model: domain description, evaluation, generalization, explanation, extrapolation, utilization, and ramification (Chapelle et al., 2008; Kane, 1992, 2006, 2013).

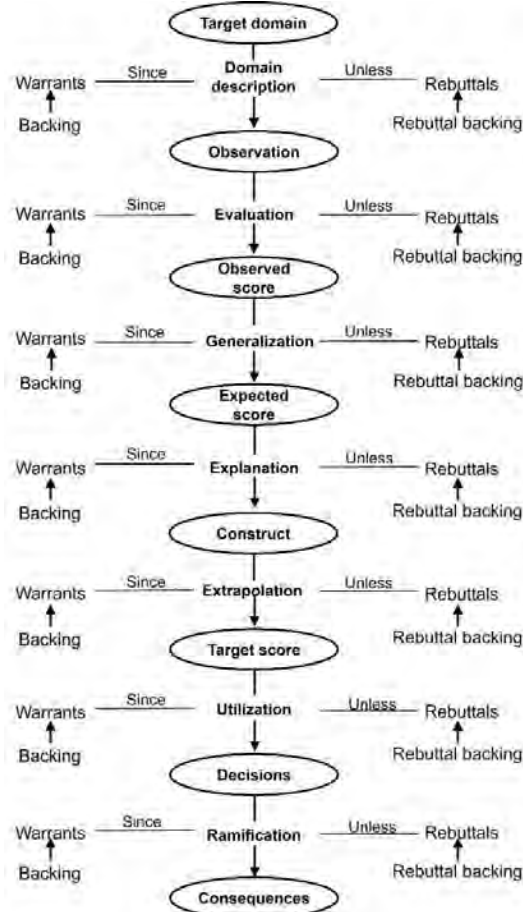
2. VALIDATION FRAMEWORK

Stakeholders' interpretations and uses of the TOEIC through YBM's perspectives are investigated based on validation model of an argument-based approach, which provides a systematic process for test validation. Bachman and Palmer (2010), Chapelle et al. (2008), Kane (1992, 2006, 2013) and Mislevy, Steinberg, and Almond (2003) proposed validation frameworks based on an argument-based approach in different names with different variations (Im, Shin, & Cheng, 2019; Im, 2021). Kane (1992, 2006) first proposed interpretive argument that is comprised of four inferences: scoring, generalization, extrapolation and decision-making inferences, and these inferences should be evaluated in terms of its coherence and of whether claims are supportive or challenged with supportive or challenging evidence called backing and rebuttal backing, respectively. Chapelle et al. (2008) and Bachman and Palmer (2010) and Mislevy et al. (2003) (see Im et al., 2019) applied this argument structure to their own validation frameworks. Validation researchers collect validity evidence to support or challenge the claim about the test framing warrants and rebuttals for the claim using the argument-based approach to validation.

In this paper, seven inferences were specified as illustrated in Figure 1 below: domain description, evaluation, generalization, explanation, extrapolation, utilization, and ramification based on Chapelle et al.'s (2008) and Kane's (2013) validation models.

The domain description inference relates to ensuring that valued knowledge, skills and abilities are appropriately identified and that those attributes are reflected on a test items and tasks. The evaluation inference pertains to ensuring that the performance on the test (i.e., observation) is adequately marked or scored based on scoring rubric and scoring procedures, which leads to test scores (i.e., observed score). The generalization inference is in relation to the reliability of test scores (i.e., expected score). The explanation inference is about whether the reliable scores are consistent with the theoretical constructs, that is, whether the scores are confirmed with the constructs using factor analysis. The extrapolation inference is about to what extent the constructs predict the performance in real life settings (i.e., target score). The utilization inference pertains to the use of the test scores (i.e., target score) for decision making (i.e., decisions) in terms of whether decisions are made appropriately based on the test scores. Lastly, the ramification inference relates to overall consequences including whether intended outcomes have been achieved, how value implications affected and what consequences and unintended consequences have been brought about. "These inferences are the processes that evaluate whether the claims are justified or challenged with warrants or rebuttals" (Im, 2021, p. 5). Through the processes, the validity of score interpretations and uses are evaluated.

FIGURE 1
Operational Validation Model for This Study



Based on the operational validation model for this study, localized meanings, uses and consequences are investigated and discussed in this study, as it provides necessary steps for evaluating testing program linking validity evidence for test design and test use.

3. METHOD

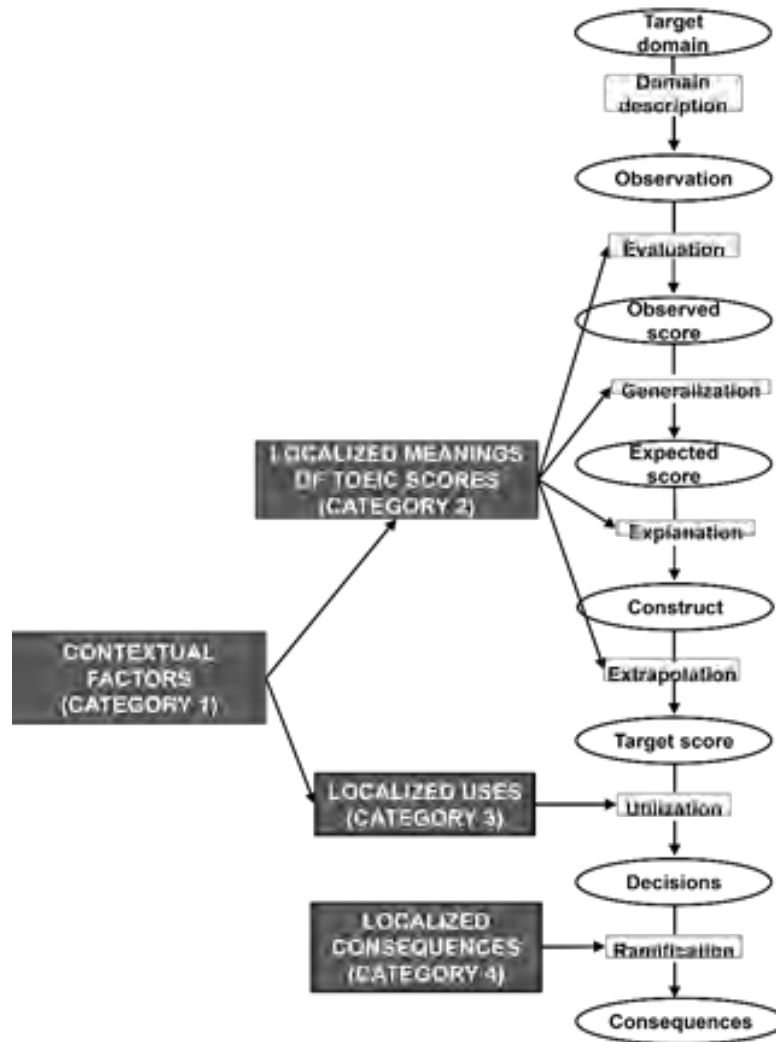
Data collection for YBM's documents was executed by searching YBM's website. Notably, YBM has recently started to run its own web blog, the TOEIC Story, and has published TOEIC Newsletters since 1997 (generally, every three months until 2002 and

quarterly since 2003 up to present). A total of 194 documents were collected from YBM's website, which included (a) 95 TOEIC newsletters published between August 1997 and October 2017; (b) 48 interviews with HR managers published between October 14, 2008, and July 17, 2017; and (c) 51 posts of the TOEIC Story published between August 14, 2015, and December 20, 2017. The source documents were analyzed using document analysis: content analysis (Bowen, 2009) and context analysis (Miller & Alvarado, 2005). By repeatedly visiting, emerging categories from the documents in relation to seven inferences were noted and then similar categories were grouped into superordinate categories under the inferences. For the whole coding processes, the relevant stakeholders (e.g., public or private enterprises, units within the enterprises, and student or employee test takers) were identified whenever a category or a subcategory was recognized.

4. FINDINGS

With regard to the research question, *What are localized meanings and uses of TOEIC scores for selection decisions and consequences of such uses in Korean context based on the specified inferences for this study?*, findings from four main categories—contextual factors, localized meanings of TOEIC scores, localized uses, and localized consequences—are reported as shown in Figure 2 below.

FIGURE 2
An Overview of the Findings from YBM's Documents and Their Corresponding Inferences
in the Operational Validation Model for This Study



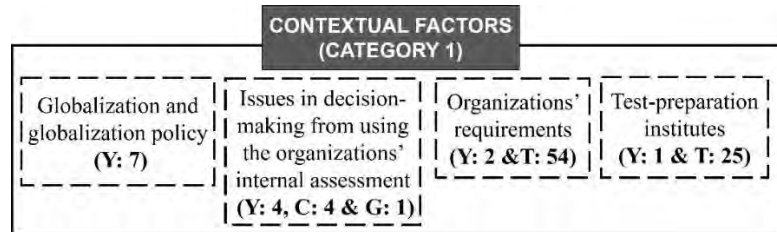
Note. Black rectangles = categories; ovals = data or claims

4.1 Contextual Factors

Contextual factors refer to factors that affect stakeholders' interpretations and uses of TOEIC scores. Four contextual factors were drawn from YBM's documents as shown in Figure 3 below: (a) *globalization and globalization policy*, (b) *issues in decision making, from using organizations' internal assessments*, (c) *organizations' requirements*, and (d)

test-preparation institutes. These factors are not independent, but interrelated, which shows testing’s associations with social, political, and cultural factors.

FIGURE 3
Category and Codes about Contextual Factors



Note. Dotted squares = codes; black rectangle = category, Y = YBM; C = companies; G = government agencies; T = test takers
 The numbers after the initials of stakeholders with colons indicate the frequency of codes mentioned by each stakeholder from three sources.

In the early 1980s when it was introduced, the TOEIC was not that popular. Because of the increasing number of international events held in Korea and the national, social, political, and economic factors the TOEIC became recognized as “a social phenomenon” everyone experiences in Korea (Korea TOEIC Committee, 2004, October, p. 13).

Additionally, YBM’s chairman further described the historical background of TOEIC’s popular use in Korea in terms of the government’s and companies’ needs.

Korea adopted the open market economy policy after the financial crisis (in 1997) and many foreign-affiliated companies invested in Korea. Because it was necessary to have English communication skills to get hired by the companies, there was a radically increased number of TOEIC test takers. In addition, Korean companies felt that they should make more emphasis on employees’ English language proficiency to be competitive in the global market, which is another reason TOEIC became popular in Korea. (Korea TOEIC Committee, 2013, February, p. 5)

To emphasize English language proficiency, companies used English interviews that they developed and administered for hiring and promotional decisions. However, the companies experienced issues with decision making using their own assessments to measure candidates’ English language proficiency. They usually used English interviews during hiring processes and encountered problems regarding fair scoring and costs. One company clearly described their challenges: “*When using English interviews during hiring*

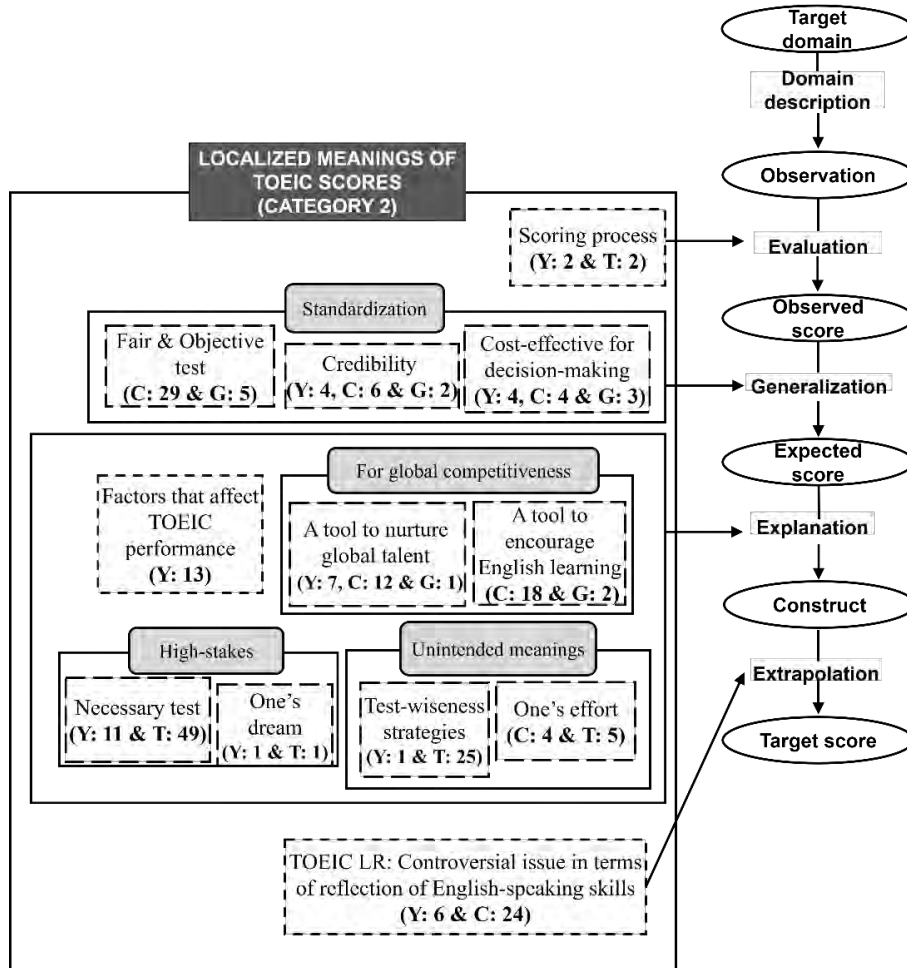
processes, discrepancies in scoring between interviewers happened a lot, and it took long for applicants to move to the interview sites and to wait for the interviews.” (Korea TOEIC Committee, 2011, October, p. 21). Alternatively, more companies started using the TOEIC because they felt the test was more effective and credible for decision making.

As most of the companies in Korea were using TOEIC scores for high-stakes decisions such as making hiring and promotional decisions (Korea TOEIC Committee, 2017, February), job seekers and employees tried to achieve higher TOEIC scores to get hired and promoted by employers. Accordingly, job seekers and employees prepared for the TOEIC in test-preparation institutes. Helping test takers achieve high TOEIC scores, test-preparation institutes in Korea became popular and affected stakeholders’ understanding of TOEIC scores, which are reported in the following section.

4.2. Localized Meanings of TOEIC Scores

Four subcategories are reported in this section: 1) *standardization*, 2) *for global competitiveness*, 3) *high stakes*, and 4) *unintended meanings*) and three individual codes (i.e., *scoring processes*, *TOEIC LR: Controversial issue in terms of reflection of English-speaking skills*, and *factors that affect TOEIC performance*) as described in Figure 4 below.

FIGURE 4
Category, Subcategories, and Codes about the Localized Meanings of TOEIC Scores and Their Corresponding Inferences about the Operational Validation Model

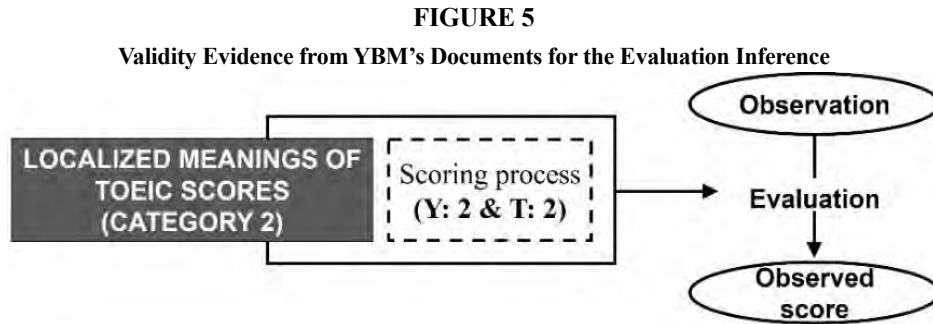


Note. Dotted squares = codes; gray squircles = subcategories; black rectangle = category, Y = YBM; C = companies; G = government agencies; T = test takers
 The numbers after the initials of stakeholders with colons indicate the frequency of the codes mentioned by each stakeholder from three sources.

4.2.1. Evaluation inference

Reporting the findings regarding the evaluation inference, the code *scoring process* within the localized meanings of TOEIC scores (category 2) in Figure 5 below, describes YBM's and test takers' contrasting views on scoring processes on the TOEIC Speaking

and YBM's effort to provide correct information about the scoring processes on the TOEIC Speaking.



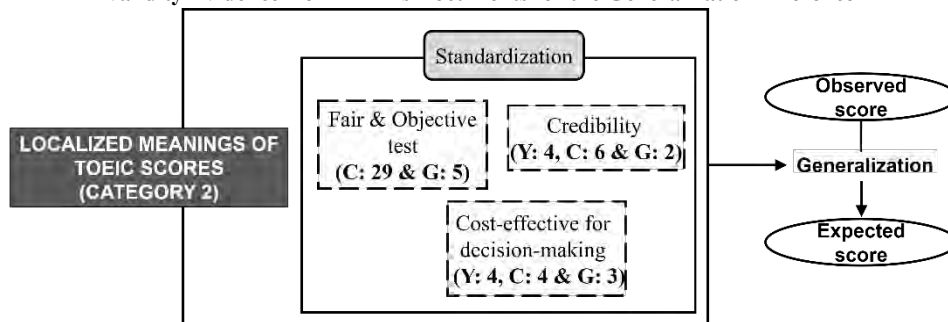
Note. Dotted square = code; black rectangle = category, Y = YBM; T = test takers
The numbers after the initials of stakeholders with colons indicate the frequency of the codes mentioned by each stakeholder from three sources.

YBM provided information about ETS's description of scoring processes, which is about administering the "Online Scoring Network" (ETS, 2010, p. 6) for the TOEIC SW (the most valuable method to enhance the accuracy and consistency of raters' scoring), while test takers' perceptions of scoring processes were quite different from ETS's. For example, a test taker stated, "It looks like raters are marking without listening to the whole responses of a test taker because the number of test takers from Korea is over the tens of thousands" (Korea TOEIC Committee, 2013, August, p. 20). Moreover, misinformation was circulating among test takers that they can get a certain number of points, even though their answers might be irrelevant to questions during the TOEIC Speaking test (Korea TOEIC Committee, 2013, August). To inform test takers about the correct information, YBM described the calibration test that raters must complete before they mark test takers' responses on the TOEIC Speaking test: "Before the raters start to mark, they have to pass the calibration test. The test is administered to examine whether the rater can mark in a right way on that day" (Korea TOEIC Committee, 2011, October, p. 29). Furthermore, the evaluation criteria for the TOEIC Speaking test were clearly provided. For example, information about the evaluation criteria for question 10 on the TOEIC Speaking test (task of Propose a solution) was provided: "The solution a test taker proposed should be well-delivered and clear. If the solution is incomplete or poorly delivered, the test taker may not be able to achieve a high score in the question" (Korea TOEIC Committee, 2011, October, p. 28). These findings exemplify YBM's efforts to provide accurate information about the TOEIC to TOEIC stakeholders.

4.2.2. Generalization inference

Standardization is important to ensure fair testing for all test takers, as “the test tasks and testing conditions are standardized in various ways” (Kane, 2013, p. 23). The standardization of the TOEIC has led its stakeholders to believe the TOEIC is fair and objective, more credible, and cost effective for decision making because the TOEIC is considered reliable, as shown in Figure 6 below.

FIGURE 6
Validity Evidence from YBM’s Documents for the Generalization Inference

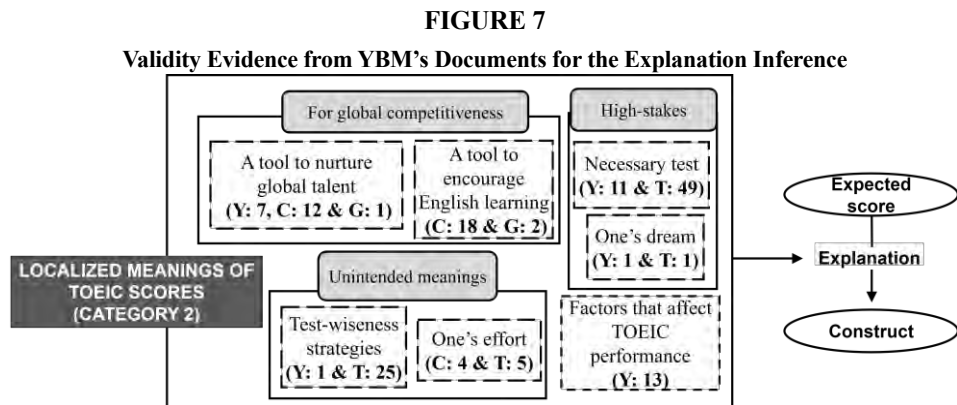


Note. Dotted squares = codes; gray squircles = subcategories; black rectangle = category, Y = YBM; C = companies; G = government agencies
The numbers after the initials of stakeholders with colons indicate the frequency of the codes mentioned by each stakeholder from three sources.

Because the TOEIC is a standardized test, organizations (i.e., companies and government agencies) understood TOEIC scores as fair, objective, and credible (codes: *fair & objective test* and *credibility* in subcategory of *standardization*) within the localized meanings of TOEIC scores (category 2): “*The reason we are using TOEIC Speaking is that the test is the fairest and most objective test*” (Korea TOEIC Committee, 2013, October, p. 8), and “*last year, we used our own assessment. However, it was less trustworthy. So we started to use TOEIC Speaking, which other major companies use in terms of credibility*” (Korea TOEIC Committee, 2009, December, p. 9). The use of TOEIC scores also helped them save costs for administering assessments they designed (code: *cost-effective for decision making*): “*I think it is more valid to use TOEIC in terms of cost-effectiveness rather than to design our own English test*” (Korea TOEIC Committee, 2009, October, p. 19). These codes in subcategory of standardization indicate stakeholders felt confident in using TOEIC scores for high-stakes decisions.

4.2.3. Explanation inference

Along with organizations' perceptions of the TOEIC's fairness, they used TOEIC scores as a tool to train their employees and to motivate them to keep learning English through the TOEIC. Companies used the scores for promotional decisions, as English is perceived as an important skill for international businesses.



Note. Dotted squares = codes; gray squircles = subcategories; black rectangle = category, Y = YBM; C = companies; G = government agencies; T = test takers
The numbers after the initials of stakeholders with colons indicate the frequency of the codes mentioned by each stakeholder from three sources.

The subcategory of *for global competitiveness* within the localized meanings of TOEIC scores (category 2) in the middle of Figure 7 (codes: *a tool to nurture global talent* and *a tool to encourage English learning*) clearly demonstrates how organizations understand the TOEIC. One company identified from YBM's documents has used TOEIC scores to train its employees as global talents and to motivate them to learn English, as follows:

By setting a certain level or cut-off score of TOEIC as an objective for employees' performance appraisals, we would like employees to feel motivated in learning English and hope that they continue to set their own goals about TOEIC scores after they reach the level or score. (Korea TOEIC Committee, 2004, December, p. 13)

Perceiving the TOEIC as a useful, fair test for training employees and for making decisions about employees, the TOEIC became widely used across Korea and considered a necessary test, even a tool to make one's dreams come true (codes in subcategory of *high-stakes*) in Figure 7. YBM's chairman stated in the TOEIC newsletter: "*Business people*

who do not speak English are treated as second- or third-class people in Korea” (Korea TOEIC Committee, 2004, October, p. 12). The emphasis was on the importance of English language proficiency: *“Without English language proficiency, you cannot survive in this competitive society”* (Korea TOEIC Committee, 2013, February, p. 7). Furthermore, *“TOEIC has become a social phenomenon”* (Korea TOEIC Committee, 2004, October, p. 13; code: *necessary test* in subcategory of *high-stakes* in Figure 7). Indeed, TOEIC became identified with university students, meaning every student should take the TOEIC in Korea (Korea TOEIC Committee, 2008, February). Going beyond the intended meaning of the TOEIC (i.e., representation of English language proficiency in one’s daily life and international environments), YBM advertised TOEIC additionally as *“one’s passion and future”* (Korea TOEIC Committee, 2008, August, p. 2), the *“TOEIC SW for one’s dream”* (Korea TOEIC Committee, 2007, June, p. 32), and a *“different tomorrow with the TOEIC”* (Korea TOEIC Committee, 2017, February, p. 20; code: *one’s dream* in subcategory of *high-stakes* in Figure 7). These ads attach additional values to the TOEIC and represent the status of English and the popularity of the TOEIC in Korea.

As reported in the section about contextual factors, the high-stakes nature of the TOEIC led test takers to only try to achieve a high TOEIC score by going to test-preparation institutes. This contextual factor also affected stakeholders’ understanding of TOEIC scores as a reflection of test-wiseness strategies (code: *test-wiseness strategies* in subcategory of *unintended meanings* in Figure 7). Indeed, concerns about the TOEIC LR test have been raised respecting high TOEIC LR scores achieved by using test-wiseness strategies regarding TOEIC preparation in test preparation institutes.

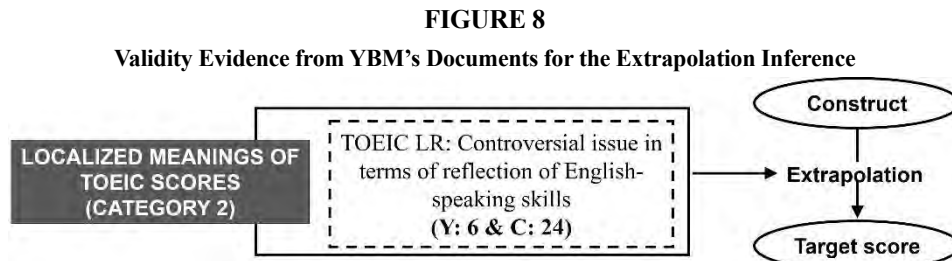
One of the interesting comments from test takers was that they commonly reported that high TOEIC scores could be achieved by time and effort, taking TOEIC classes at the institute, or studying by themselves (code: *one’s effort* in subcategory of *unintended meanings* in Figure 7). One test taker remarked, *“I think anyone can achieve what they are aiming for in the TOEIC if they highly concentrate on [it]. TOEIC scores depend on how much effort you make”* (Korea TOEIC Committee, 2017, September 13). Another test taker explicitly commented a high TOEIC LR score represents one’s effort and diligence: *“TOEIC LR scores are perceived as a barometer for one’s effort and diligence instead of English skills”* (Korea TOEIC Committee, 2017, October 12). These attributes are highly valued in the Korean context. This view of TOEIC scores was also evident in employers. One employer clearly reported that TOEIC scores represent *“the candidate’s diligence”* (Korea TOEIC Committee, 2016, September, p. 13). Moreover, YBM conducted a survey to ask employers why they use TOEIC scores for hiring, and 37 out of 100 major companies in Korea responded that they were using TOEIC scores to evaluate whether the candidate is diligent and ready to work (Korea TOEIC Committee, 2014, December). This understanding of TOEIC scores may be based on Korean cultural values regarding Im and

McNamara's (2017) assertion.

Additionally, YBM regularly published reports about the relationships between the TOEIC scores of Korean test takers and other factors that may be related to their TOEIC performance (code: *factors that affect TOEIC performance* in Figure 7). The factors identified were (a) the length of residence in English-speaking countries, (b) the number of TOEIC attempts, (c) education, and (d) the frequency of English use at the workplace. The results showed significantly high correlations between TOEIC scores and the following factors: higher TOEIC scores for test takers who stayed longer in English-speaking countries and made a greater number of TOEIC attempts, higher education, and more frequent English use in the workplace. However, reporting these factors are not associated with ETS's intended meanings of TOEIC scores, but provide information on contextual factors that may affect Korean test takers.

4.2.4. Extrapolation inference

Despite the wide use of the TOEIC in Korea, there has been much criticism of the TOEIC LR due to its under-representation of English communicative skills (code: *TOEIC LR: Controversial issue in terms of reflection of English-speaking skills* within the localized meanings of TOEIC scores [category 2] in Figure 8 below).



Note. Dotted square = code; black rectangle = category, Y = YBM; C = companies
The numbers after the initials of stakeholders with colons indicate the frequency of the codes mentioned by each stakeholder from three sources.

Some HR managers stated in the TOEIC Newsletter that they cannot assess an applicant's English-speaking skills with TOEIC LR scores. Others further commented the TOEIC LR is just for basic English skills, while the TOEIC SW is for English skills at work. To react to this criticism, YBM's chairman further expressed his thoughts about the meanings of TOEIC scores in the TOEIC Newsletter, stating, "*The score of 900 in the TOEIC LR indicates that the test taker is proficient in English as a non-native English speaker, but it does not mean that he/she speaks like a native English speaker*" (Korea

TOEIC Committee, 2006, December, p. 12). However, a YBM executive showed a different perspective of the TOEIC LR in the TOEIC Newsletter: *“Although he/she gets TOEIC LR scores above 900, it may not be possible to say that a test taker can speak English fluently”* (Korea TOEIC Committee, 2007, June, p. 7). In addition, the executive commented in the TOEIC Newsletter: *“The TOEIC LR is not a test that assesses all aspects of English language proficiency, and therefore, the LR scores should be used for assessing basic English proficiency”* (Korea TOEIC Committee, 2007, June, p. 7).

4.3. Localized Uses of TOEIC Scores (Utilization Inference)

Two stakeholder groups (i.e., companies and government agencies) as illustrated at the top left of Figure 9 were administering English programs for training employees. Additionally, they offered incentives and waived the TOEIC fee for motivation. English language programs were administered in various forms, such as in-house education, camp training, online programs, and financial support. One company administered three-month camp training, in-house English conversation classes twice or three times a week, and TOEIC online courses (Korea TOEIC Committee, 2005, June).

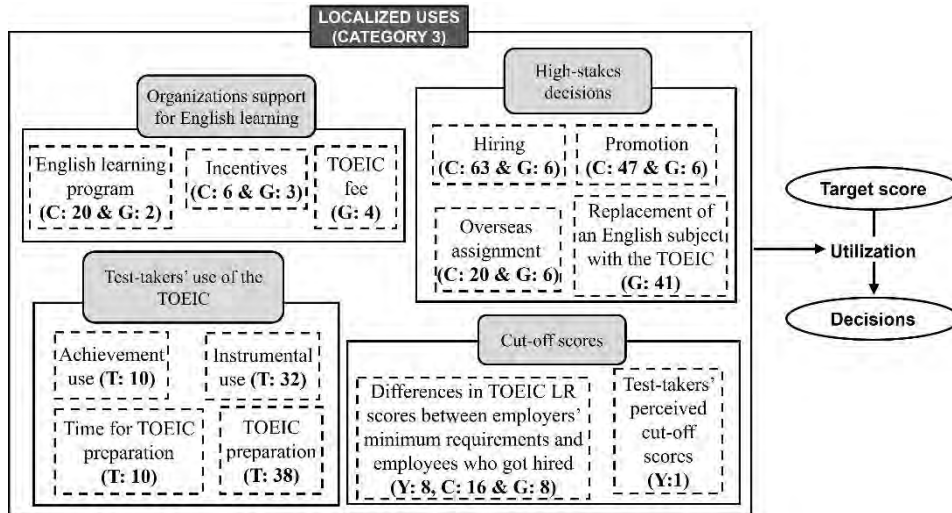
These supports were ways to encourage employees to learn English. The other way to do so was to use TOEIC scores for high-stakes decisions (Korea TOEIC Committee, 1997, August), such as hiring, promotion, and overseas assignments. One manager in a major company in Korea clearly stated, *“English skills are the most important quality to be competitive in the global market, and the company is using TOEIC scores for hiring, promotion, and overseas assignments to emphasize English language proficiency, as English language proficiency is important to be competitive”* (Korea TOEIC Committee, 2016, August, p. 15). However, some companies changed their personnel policy of using TOEIC LR scores with those of TOEIC Speaking for promotional decisions and overseas assignments.

With the use of TOEIC scores for high-stakes decisions in companies, government agencies also replaced the English subject on national examinations with TOEIC scores. The Korean government announced the replacement of the English subject on the Bar, Foreign Service, and Civil Service examinations by submitting TOEIC scores in 2003. Regarding this announcement, the YBM executive stated the purpose of the replacement was to improve the government officials’ performance in the globalized world and to help those preparing for the national examinations to choose another career (Korea TOEIC Committee, 2003, June) if they fail the national exams. Additionally, the Korean government started to use TOEIC scores for licensing and certifying such professions as a certified public accountant and a patent attorney. YBM also indicated the government agencies’ use of TOEIC scores can help test takers relieve the burdens on job-seeking

because they can use TOEIC scores for licensing, certifying, and job seeking for a company (Korea TOEIC Committee, 2009, October). These government agencies' uses of TOEIC scores show that social pressures have affected test users' utilization of TOEIC scores in the Korean context.

FIGURE 9

Validity Evidence from YBM's Documents for the Utilization Inference



Note. Dotted squares = codes; gray squircles = subcategories; black rectangle = category, Y = YBM; C = companies; G = government agencies; T = test takers
The numbers after the initials of stakeholders with colons indicate the frequency of codes mentioned by each stakeholder from three sources.

Despite the high-stakes nature of the TOEIC, not all test takers sit for the TOEIC to get hired or promoted. Because the TOEIC is an English test, test takers and employees undergo it to improve their English skills (code: *achievement use* in Figure 9). One employee in YBM's documents commented that he thought the TOEIC Speaking would be useful because he would have been involved in hiring processes as an interviewer in English and would have conference calls in English (Korea TOEIC Committee, 2017, July 27). Nevertheless, an interesting comment was identified from an undergraduate test taker. She shared the reason she started to take TOEIC courses at her university: "*I started to prepare for the TOEIC because I felt that I needed to study English because medical terminologies I learned are all in English*" (Korea TOEIC Committee, 2018, January 25). Although the TOEIC does not include medical terminology, the subject may have taken TOEIC courses to get familiar with English.

However, one of the major reasons for test takers to sit for the TOEIC was to get a job.

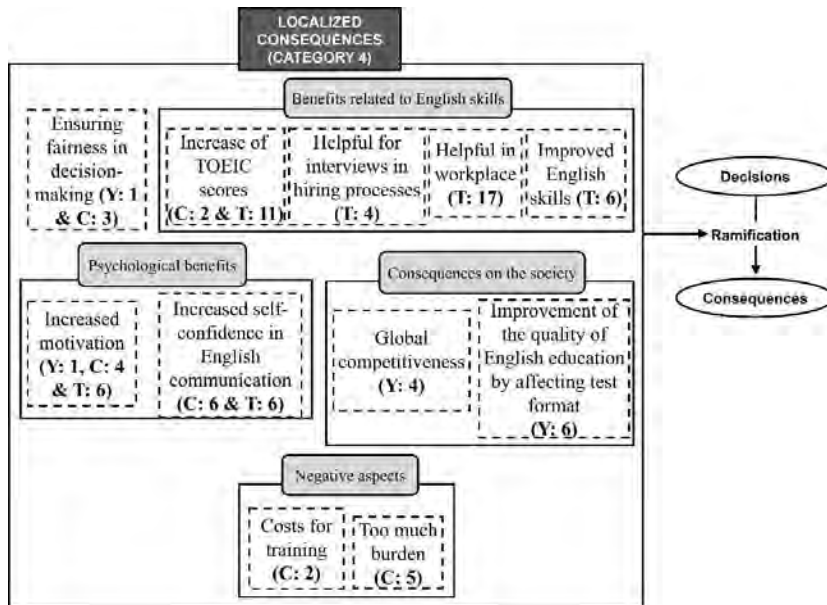
Indeed, organizations' wide uses of TOEIC scores for high-stakes decisions affected high-school and university students as job seekers preparing for the TOEIC. They registered for short- or long-term TOEIC courses run by their school and test preparation institutes to achieve high TOEIC scores and be competitive in hiring processes (code: *instrumental use* in Figure 9). For example, one university student who registered for TOEIC course remarked, "TOEIC scores are necessary for employment because it looks like almost all companies in Korea use TOEIC scores" (Korea TOEIC Committee, 2015, September 3, para 10). This represents test takers' instrumental use of TOEIC scores, mostly for getting hired. In addition, university students and even high-school students seeking a job after graduation tried to prepare for the TOEIC as early as possible for employment: "It is better to prepare for employment as early as possible" (Korea TOEIC Committee, 2010, April, p. 26; codes: *Time for TOEIC preparation* and *TOEIC preparation* in Figure 9). Because the TOEIC has been a dominant English test in Korea and its stakes have been high for stakeholders, the use of TOEIC has been localized in response to social and political pressures.

The other characteristics of the localized uses of TOEIC scores are that there are big differences between TOEIC scores in the minimum requirements for a job application and employees who get hired (code: *differences in TOEIC LR scores between employers' minimum requirements and employees who got hired* in Figure 9). In general, the cut-off score of the TOEIC LR for hiring and promotion in companies and government agencies was 700 out of 990 points. YBM's survey with 499 companies in Korea revealed that 188 provided cut-off scores as a requirement for a job application, and the cut-off score set by major companies was on average 724 points; by government agencies, 716 points; and by foreign-affiliated companies, 713 points on the TOEIC LR (Korea TOEIC Committee, 2009, August). However, one major company reported the TOEIC LR scores of recently hired employees ranged from 850 to 900 points (Korea TOEIC Committee, 2011, April). Consequently, test takers' perceived cut-off scores for hiring (code: *test takers' perceived cut-off scores*) were above 905 (32.5%), 855–900 (25%), and 805–850 points (25.1%; Korea TOEIC Committee, 2012, June, p. 6), according to YBM's survey of 1,146 test takers studying TOEIC courses at YBM's TOEIC preparation institutes. Moreover, 82.6% of test takers responded that they needed at least 805 points on the TOEIC LR to get hired. According to YBM, employees who get hired by the top-ten major companies in Korea on average earned 852 points on the TOEIC LR (Korea TOEIC Committee, 2013, August). These findings describe why test takers are trying to achieve a higher score to get hired in Korea by attending TOEIC preparation institutes.

4.4. Localized Consequences of TOEIC Uses (Ramification Inference)

The localized consequences of the TOEIC are shown by the beneficial effects of TOEIC use on TOEIC stakeholders and Korean society, alongside the negative consequences of TOEIC use. The category of localized consequences includes one individual code (i.e., ensuring fairness in decision making) on the far left and four subcategories and four, two, two, and two codes under each subcategory, respectively as described in Figure 10 below. While the individual codes and three subcategories of benefits related to English skills, psychological benefits, and societal consequences concern the positive aspects of TOEIC use, subcategory of *negative aspects* pertains to the negative unintended consequences on TOEIC stakeholders.

FIGURE 10
Validity Evidence from YBM’s Documents for the Ramification Inference (Localized Consequences)



Note. Dotted squares = codes; gray squircles = subcategories; black rectangle = category. Y = YBM; C = companies; G = government agencies; T = test takers
The numbers after the initials of stakeholders with colons indicate the frequency of codes mentioned by each stakeholder from three sources.

Regarding the positive aspects of TOEIC use, YBM frequently reported that the use of TOEIC scores led to accurately ensuring fairness in decision making (code: *ensuring*

fairness in decision making, at the far left of Figure 10). An employer at a major Korean company stated: “Due to the use of TOEIC scores, complaints about fairness in decision making were gone” (Korea TOEIC Committee, 2008, April, p. 7). The employer even further said, “TOEIC use contributed to having clearer criteria for selection decisions. [...] As everyone recognizes the credibility of TOEIC, the image that the company is fair can be reinforced” (Korea TOEIC Committee, 2008, April, p. 7).

Benefits for test takers regarding benefits related to English skills also were found from the analysis of YBM’s documents. University students, job seekers, and employees registered for short- and long-term TOEIC courses in company or test preparation institutes. From the courses, test takers commonly stated in YBM’s documents that they attained higher TOEIC scores (code: *increase of TOEIC scores* in Figure 10). One test taker who took a TOEIC course run by YBM expressed his satisfaction with the course, as he could learn test-wiseness strategies and achieve a high score in the short-term (Korea TOEIC Committee, 2016, January 26).

Especially while test takers prepared for the TOEIC Speaking test, they commonly delineated that TOEIC preparation helped them during an English job interview and hiring processes (code: *helpful for interviews in hiring processes* in Figure 10) and in workplaces where English was used as a medium (code: *helpful in workplace*). In addition, some test takers reported that they had improved their English skills (code: *improved English skills* in Figure 10), although they started to prepare for the TOEIC only to meet companies’ requirements.

Besides these benefits for decision making and English skills, test takers commonly mentioned they felt increased motivation to learn English and greater confidence in English communication (codes: *increased motivation* and *increased self-confidence in English communication* in subcategory of *psychological benefits*) at the middle left of Figure 10. These psychological benefits were commonly expressed by those who registered for TOEIC courses run by YBM and universities. Those TOEIC courses were operated for about one month. Through the intensive course, test takers achieved higher TOEIC scores, could build self-confidence, and got motivated to learn English. Besides these individuals’ perceptions of the consequences of TOEIC use in the Korean context, YBM stated the Korean society has enjoyed positive consequences.

YBM executives and a manager commonly delineated that positive contextual consequences from the use of TOEIC scores helped Korean human resources to compete in the global market (code: *global competitiveness* in Figure 10). Moreover, Korea’s English education was improved when the test format was changed from measuring English grammar to assessing practical English skills (code: *global competitiveness* and *improvement of the quality of English education by affecting test format*) in subcategory of *consequences on the society* in Figure 10.

The emphasis on practical English skills even led to a change in test format for the English subject on the College Scholastic Ability Test (CSAT) in Korea, administered annually on Thursdays in the second week of November for senior secondary school students in Korea to take for university admissions. In 1994, English listening comprehension was introduced to the CSAT to measure practical English listening skills. In 2010, the number of listening items was increased from 17 to 20 items. YBM pointed out that these changes were led by the TOEIC, which measures English listening and reading skills (Korea TOEIC Committee, 2011, October).

However, negative aspects of consequences were found by employers, as displayed at the far right of Figure 10. One employer stated that he needed to train employees who submitted TOEIC scores, and the required training was a huge cost for international businesses (code: *costs for training* in subcategory of *negative aspects* in Figure 10).

This negative aspect of using TOEIC scores may represent the high-stakes nature of the TOEIC for Korean job seekers to achieve high TOEIC scores to get hired. The high stakes of the TOEIC were also evident in employers' comments on their employees' voices regarding the requirements of TOEIC scores for promotion. There were contrasting views on employers' policy between new and existing employees. One employer implemented a policy that all employees needed to submit TOEIC Speaking scores mandatorily. He commented on employees' different views of the policy: "*New employees had positive perceptions of taking the TOEIC Speaking and felt less burden and high confidence in the test because they already took TOEIC Speaking for a job application*" (Korea TOEIC Committee, 2013, December, p. 7). One employer in another company described his/her employees' burdens (code: *too much burden* in Figure 10):

All employees in this company were required to take the TOEIC Speaking mandatorily. At that time, those who had a score above 160 in the TOEIC Speaking were exempted from submitting TOEIC Speaking scores. [...] New employees who had opportunities to take the TOEIC Speaking or had TOEIC Speaking scores were in favor of the company's policy. However, existing employees felt a huge burden. In particular, engineers in the company did not like learning English. (Korea TOEIC Committee, 2014, February, p. 9)

The use of TOEIC scores led to both positive and negative aspects of consequences for individuals and companies. The negative aspects of TOEIC use are the unintended consequences of companies' English language testing policy respecting the side effects of ETS's intended uses of TOEIC scores. Kane (2013) noted that consequences are the main criteria when testing is used for policy practices and test users such as employers and

employees can best identify unintended consequences. Kane (2013) also commented that test developers are responsible for the consequences from the intended uses of test scores. Although the uses of TOEIC scores have been localized, ETS may need to be responsible for the negative unintended consequences of the companies' uses of TOEIC scores.

5. DISCUSSION AND CONCLUSIONS

5.1. Contextual Factors

One of the major findings in this study is the existence of contextual factors that play into validation, which I did not expect. The findings about contextual factors do not directly answer the research question addressed in this paper, but may explain how and why TOEIC scores were understood and used in Korea. This point is evident in figures 7 and 9 from YBM's documents in terms of contextual factors that affected TOEIC users' understanding (meanings) and uses of TOEIC scores and the consequences in Korea. The finding may be valuable and supports previous studies, in terms of social, cultural, and political aspects of language testing (Im, Cheng & Shin, 2020; Im & McNamara, 2017; McNamara & Roever, 2006), as well as direct future studies.

The findings in this study claim that contextual factors should be included in validation (Shepard, 2000). As the findings showed the relationships between contextual factors and TOEIC stakeholders' interpretations and uses of TOEIC scores, contextual factors affect stakeholders beliefs and their uses of test scores.

Validation requires evaluating to what extent the intended interpretations and uses of test scores have been achieved. This means that validation should be conducted to investigate different stakeholders' interpretations and uses of test scores, considering the contextual factors, as noted by Moss, Girard, and Haniford (2006): "Externally mandated tests are always interpreted and used in a particular local context, which shape and are shaped by them" (p. 145). What needs to follow would be to conceptualize a validation framework that provides systematic procedures of validation to include contextual factors in test validation in a particular context. The findings reported specifically in section 4 would be data for a validation framework that includes contextual factors in validation.

5.2. Meanings of TOEIC Scores (Domain Description to Extrapolation Inferences)

The meanings of TOEIC scores are relevant to the first five inferences (the domain

description to extrapolation inferences) in Figure 4. It was found that the findings from YBM's documents described the contentious issues regarding the evaluation, explanation, and extrapolation inferences (i.e., distinct localized meanings of TOEIC scores in the Korean context), while generally supporting the generalization, utilization, and ramification inferences.

5.2.1. Evaluation and generalization inferences

The evaluation inference in the operational validation model of this paper pertains to the adequacy of a scoring rubric and scoring procedures on the TOEIC. The other inference for ensuring the validity of TOEIC scores is the generalization inference, which refers to estimates of the consistency of test scores over test tasks and testing contexts (Kane, 2013).

To enhance the accuracy and consistency of scoring procedures for fairness, ETS is administering a very systematic scoring system (ETS, 2010), while the findings from YBM's documents (code: *scoring process* in Figure 5) may call for more efforts to provide more information about scoring processes to stakeholders. Test takers' perceptions of the scoring processes were quite different from what YBM provided (e.g., Korea TOEIC Committee, 2011, October): "*It looks like raters are marking without listening to the whole responses of a test taker because the number of test takers from Korea is over the tens of thousands*" (Korea TOEIC Committee, 2013, August, p. 20). In addition, test takers perceived that they could achieve more points, even if they kept talking anything irrelevant while answering questions during the TOEIC Speaking (Korea TOEIC Committee, 2013, August). Although test takers had varying perceptions of the scoring processes, it is ETS's and YBM's responsibility to provide more sufficient information to help them understand the meanings of TOEIC scores more accurately. Otherwise, test takers may have a misguided understanding and uses of TOEIC scores.

The generalization inference in the operational validation model of this paper is important to provide decision makers with a reliable measure to help them make valid decisions about test takers. Supporting validity evidence for the generalization inference was found from YBM's documents (codes in Figure 6). Organizations (i.e., companies and government agencies) showed positive perceptions of the TOEIC, as they could minimize, by using the standardized TOEIC, the decision errors resulting from using their internal assessments to evaluate candidates' English language proficiency.

The validity evidence from ETS and YBM that was provided for the evaluation and generalization inferences are necessary conditions for TOEIC scores to be valid for decision making. The findings from YBM's documents included subcategory, *standardization*. As Kane (2013) noted, standardization may threaten the extrapolation inference regarding the under-representation of the target domain. More important aspects

than internal consistency and reliability are the evidence for the explanation and extrapolation inferences that pertain to the extent to which TOEIC accurately measures the intended constructs and TOEIC scores predict the test taker's performance in the real-life settings. These inferences may be the most important qualities of the TOEIC to help decision makers feel confident when using TOEIC scores for decision making.

5.2.2. Explanation inference

The explanation inference pertains to whether TOEIC accurately measures its intended constructs, including the extent to which TOEIC scores means equally to all stakeholders. The explanation inference in this subsection is discussed in terms of theoretical constructs underlying the TOEIC and construct-irrelevant variances such as test-wiseness strategies and stakeholders' values associated with the TOEIC.

The findings from YBM's documents explicitly showed the construct-irrelevant variances, for example, test-wiseness strategies in the Korean context. However, "increasing familiarity with the TOEIC and test-wiseness strategies may help test takers use their knowledge optimally by reducing the psychological burden" (Im, 2021, p.25) (e.g., test anxiety; Ma & Cheng, 2018), which increases the validity of score interpretations of a test.

Besides the test-wiseness strategies, the view of TOEIC scores as a reflection of one's effort was common across YBM's documents, which is consistent with Im and McNamara (2017). Furthermore, additional unintended meanings of TOEIC scores in the Korean context, such as a necessary test, and mandatory to make one's dream come true due to the TOEIC's stakes were found from YBM's documents.

Although these views of TOEIC scores may be the unintended meanings, stakeholders' interpretations of TOEIC scores may be valid in the Korean context if positive results from decision making have been brought about based on the interpretations of TOEIC scores. For example, that employers evaluate a candidate's character quality (i.e., one's effort) through TOEIC scores may also be valid if they hire more diligent and more dedicated candidate for their companies, which will be discussed in the following sections. "What ETS may need to consider is that its intended meanings for TOEIC scores may be interpreted in a different way in a given context and there is a need to investigate unintended meanings of TOEIC scores, to ensure positive consequences" (Im, 2021, p. 26)

5.2.3. Extrapolation inference

The last inference regarding the meanings of TOEIC scores to be discussed is the extrapolation inference, "which pertains to the prediction of test takers' performance in

real-life settings, based on their TOEIC scores” (Im, 2021, p. 26). The extrapolation inference in this subsection is discussed in terms that the TOEIC reflect employees’ performance in international business workplaces.

Controversial issues regarding TOEIC LR’s reflection of English-speaking skills (code: *TOEIC LR: Controversial issue in terms of reflection of English-speaking skills*) were found in this study. There has been controversy regarding whether TOEIC LR scores can reflect English-speaking skills in the Korean context. This work discerned that views of the TOEIC LR as an indicator of English-speaking skills have not been supported.

This issue of the TOEIC LR’s reflection of English-speaking skills may have been raised because of the prevalence of TOEIC preparation institutes in Korea because test takers learn and use test-wiseness strategies to achieve a higher TOEIC score without improving their knowledge about English. This phenomenon was described by the YBM executive: “*The reason those who have TOEIC LR scores above 900 do not speak English fluently is that they only learned listening and reading skills with going to language schools*” (Korea TOEIC Committee, 2005, February, p. 3).

5.3. Uses of TOEIC Scores (Utilization Inference) Affected by Contextual Factors

Findings about YBM’s localized uses of TOEIC scores showed consistency with ETS’s intended uses of TOEIC scores, regarding companies’ uses of TOEIC scores for making selection decisions and their support for employees’ English learning and test takers’ uses of TOEIC in section 4. However, some uses of TOEIC scores had been affected by contextual factors and showed different uses of the scores from the intended uses in the Korean context. For example, the code, *replacement of an English subject with the TOEIC*, by government agencies showed the social pressures in Korea regarding helping test takers relieve their burdens because they can use TOEIC scores for job seeking at a company if they fail to gain licensing and certification from government agencies. This finding is unique in the Korean context and showed that test use is also affected by contextual factors in a particular context. This use of TOEIC scores may also need to be investigated further, as it is not specified by ETS.

In reality, based on the findings from YBM’s documents, employees who got hired had very high TOEIC LR scores, although the minimum requirements for the TOEIC LR scores were above about 700 points. This finding from YBM’s documents explains why test takers are trying to achieve a higher score to get hired by going to TOEIC preparation institutes and, furthermore, organizations’ uses of TOEIC as a gate-keeping tool.

5.4. Consequences of TOEIC Use (Ramification Inference)

Consequences of TOEIC uses correspond to the ramification inference in the operational validation model for this study, which links decisions to consequences. This inference may be one of the most important qualities of validity and the main criterion for evaluation (Kane, 2013).

Findings regarding consequences of the TOEIC use in this section are discussed in terms of the consequences of TOEIC and TOEIC uses. The findings in this study generally show both positive and negative aspects of consequences. This subsection discusses the findings in relation to consequences of the TOEIC relevant to English skills and consequences of TOEIC uses such as selection decisions.

YBM's documents provided positive validity evidence for the ramification inference in the Korean context regarding the benefits related to English skills and the benefits for society in Figure 10. Rather, negative aspects of TOEIC uses were reported in terms of the costs to train employees. For example, one employer from YBM's documents stated that employers had to spend a great deal to train their employees for international businesses. Furthermore, the employer from YBM's documents pointed out the test takers' fever to achieve high TOEIC scores rather than to improve their English language proficiency: "*We spend huge costs in training employees regarding business manners and English presentation skills. I think that companies, students, and schools need to pay attention to improving English communicative skills rather than achieving a high TOEIC score*" (Korea TOEIC Committee, 2012, April, p. 5). These findings do not support the intended consequences for business organizations: making "more informed decisions on hiring, training and promoting the best candidates to roles where English skills matter the most" (ETS, 2018, para 3). Rather, the TOEIC's consequences might happen because of the invalidity of the test itself and social pressures, as Messick (1989) pointed out.

Positive consequences regarding TOEIC uses and selection decisions were found in YBM's documents respecting ensuring fairness in decision making. Code of *ensuring fairness in decision making* may have resulted from the standardization of the TOEIC to minimize decision errors. As discussed in the previous section, standardization may threaten the extrapolation inference (Kane, 2013) in terms of under-representing the target domain.

The findings discussed thus far clearly show the effect of contextual factors on TOEIC score meanings and uses, and a washback effect on the employees, which can be either negative or positive on the employees in this study, and support previous research (e.g., Cheng & Curtis, 2004; Shohamy, Donitsa-Schmidt, & Ferman, 1996).

Overall, consequences of selection decisions were highly associated with the stakes of the TOEIC. Throughout the discussion section, contextual factors that affected the stakes

of the TOEIC and stakeholders' understanding and uses of TOEIC scores as well as TOEIC's under-representation of English constructs in international business workplaces have led to mostly negative consequences and some positive consequences. As discussed in this section, it may be necessary to discuss the validity in a test within sociopolitical systems and to conceptualize a validation model that includes social, cultural, and political factors in validation, as validity cannot be separated from the social aspects of testing (Im & McNamara, 2017).

Applicable levels: Tertiary

REFERENCES

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. London, UK: Routledge.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–18). Mahwah, NJ: Lawrence Erlbaum.
- Educational Testing Service. (2010). *User guide speaking & writing*. Retrieved on September 1, 2019 from Educational Testing Service website: https://www.ets.org/s/TOEIC/pdf/TOEIC_sw_score_user_guide.pdf
- Educational Testing Service. (2018). About the TOEIC® Listening and Reading test. Retrieved on October 18, 2018, from <https://www.ets.org/TOEIC/organizations/listening-reading/about/>
- Im, G.-H. (2021). What do international professionals say about the TOEIC? *Journal of the Korea English Education Society*, 20(1), 1–37.

- Im, G.-H., & Cheng, L. (2019). The Test of English for International Communication. *Language Testing*, 36(2), 315–324.
- Im, G.-H., Cheng, L., & Shin, D. (2020). Sociopolitical factors surrounding the rise and demise of the National English Ability Test (NEAT) in South Korea. In S.-A. Mirhosseini & P. I. De Costa (Eds.), *Sociopolitics of English language testing* (pp. 147–164). London, UK: Bloomsbury Publishing.
- Im, G.-H., & McNamara, T. (2017). Legitimate or illegitimate uses of test scores in contexts unrelated to test purposes. *English Teaching*, 72(2), 71–99.
- Im, G.-H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: Their limitations and future directions for validation research. *Language Testing in Asia*, 9(14). <https://doi.org/10.1186/s40468-019-0089-4>
- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51–66.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535.
- 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.
- Korea TOEIC Committee. (1997, August). Newsletter. *YBM Newsletter*; 1. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/1.pdf
- Korea TOEIC Committee. (2003, June). Newsletter. *YBM Newsletter*; 18. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/18.pdf
- Korea TOEIC Committee. (2004, October). Newsletter. *YBM Newsletter*; 24. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/24.pdf
- Korea TOEIC Committee. (2004, December). Newsletter. *YBM Newsletter*; 25. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/25.pdf
- Korea TOEIC Committee. (2005, February). Newsletter. *YBM Newsletter*; 26. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/26.pdf
- Korea TOEIC Committee. (2005, June). Newsletter. *YBM Newsletter*; 28. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/28.pdf

- Korea TOEIC Committee. (2006, December). Newsletter. *YBM Newsletter*; 37. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/37.pdf
- Korea TOEIC Committee. (2007, June). Newsletter. *YBM Newsletter*; 40. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/40.pdf
- Korea TOEIC Committee. (2008, February). Newsletter. *YBM Newsletter*; 44. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/44.pdf
- Korea TOEIC Committee. (2008, April). Newsletter. *YBM Newsletter*; 45. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/45.pdf
- Korea TOEIC Committee. (2008, August). Newsletter. *YBM Newsletter*; 47. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/47.pdf
- Korea TOEIC Committee. (2009, August). Newsletter. *YBM Newsletter*; 53. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/53.pdf
- Korea TOEIC Committee. (2009, October). Newsletter. *YBM Newsletter*; 54. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/54.pdf
- Korea TOEIC Committee. (2009, December). Newsletter. *YBM Newsletter*; 55. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/55.pdf
- Korea TOEIC Committee. (2010, April). Newsletter. *YBM Newsletter*; 57. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/57.pdf
- Korea TOEIC Committee. (2011, April). Newsletter. *YBM Newsletter*; 63. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/63.pdf
- Korea TOEIC Committee. (2011, October). Newsletter. *YBM Newsletter*; 66. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/66.pdf
- Korea TOEIC Committee. (2012, April). YBM Korea TOEIC® committee newsletter. *YBM Newsletter*; 69. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdf/69.pdf

- Korea TOEIC Committee. (2012, June). YBM Korea TOEIC® committee newsletter. *YBM Newsletter*; 70. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/70.pdf
- Korea TOEIC Committee. (2013, February). YBM Korea TOEIC® committee newsletter. *YBM Newsletter*; 74. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/74.pdf
- Korea TOEIC Committee. (2013, August). YBM/Korea TOEIC® committee newsletter. *YBM Newsletter*; 77. Retrieved on November 11, 2018 from http://upsisa.ybmsisa.com/si/exam_ybmsisa_com/newsletter/pdffile/77.pdf
- Korea TOEIC Committee. (2013, October). YBM/Korea TOEIC® committee newsletter. *YBM Newsletter*; 78. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/78.pdf
- Korea TOEIC Committee. (2013, December). YBM/Korea TOEIC® committee newsletter. *YBM Newsletter*; 79. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/79.pdf
- Korea TOEIC Committee. (2014, February). YBM/Korea TOEIC® committee newsletter. *YBM Newsletter*; 80. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/80.pdf
- Korea TOEIC Committee. (2014, December). YBM/Korea TOEIC® committee newsletter. *YBM Newsletter*; 85. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/85.pdf
- Korea TOEIC Committee. (2015, September 3). Thoik? Cakyekcung? Chwuyepay kacang cwungyohan supheykun? [TOEIC? Certificates? The most important qualification for a job application?] [Web log post]. Retrieved on November 11, 2018 from <http://www.TOEICstory.co.kr/33>
- Korea TOEIC Committee. (2016, January 26). Kongkiep spheyk cunpiuy sicak! Thoikulo chetpal tteyki [TOEIC: Starting point for qualification to get a job at the government enterprises] [Web log post]. Retrieved on November 11, 2018 from <https://www.TOEICstory.co.kr/153>
- Korea TOEIC Committee. (2016, August). Newsletter: YBM Korea TOEIC® Committee. *YBM Newsletter*; 97. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/97.pdf
- Korea TOEIC Committee. (2016, September). Newsletter: YBM Korea TOEIC® Committee. *YBM Newsletter*; 98. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/98.pdf
- Korea TOEIC Committee. (2017, February). Newsletter: Expand your possibilities. *YBM Newsletter*; 102. Retrieved on November 11, 2018 from http://upsisa.ybmnet.co.kr/si/exam_ybmsisa_com/newsletter/pdffile/102.pdf

- Korea TOEIC Committee. (2017, July 27). Konginhwoykyeysa sihemul cwunpihandamyen? Samilhwoykyeypepin ipsacaka malhanun CPA kkwulthip! [Preparing for CPA? Tips from the employee at Samil accounting firm] [Weblog post]. Retrieved on November 11, 2018 from <https://www.TOEICstory.co.kr/696>
- Korea TOEIC Committee. (2017, September 13). Inthen chayyongey kwansimitnun salam cwumok! Hankwukkyengceysinmwun inthensawuenduluy inthencwunpi thip kongkay [Attention, those who are interested in internship! Tips for an intership application provided by intern staff at the Korea Economic Daily] [Web log post]. Retrieved on November 11, 2018 from <https://www.TOEICstory.co.kr/733>
- Korea TOEIC Committee. (2017, October 12). Hapanki myencep philswuyoken! Thoiksuphiking kyenghemcaka cikcepmalhanun thosuhyokwanun? [Requirement in English interviews in the second half of the year! The usefulness of the TOEIC Speaking from a test-taker?] [Web log post]. Retrieved on November 11, 2018 from <http://www.TOEICstory.co.kr/754>
- Korea TOEIC Committee. (2018, January 25). Chwuyepcwunpiey thoik sengcekun kiponicyo, halwu 10 sikanssik thoik sengcekul wuyhan thoikkhaymphu! Seymyengtayhakkyo thoikkaymphu hyencangsukheychi [TOEIC scores: Necessary for a job application. Ten hours of TOEIC preparation a day in the in-house TOEIC courses at Semyung University!] [Web log post]. Retrieved on November 11, 2018 from <https://www.TOEICstory.co.kr/838>
- Li, Z. (2015). *An argument-based validation study of the English placement test (EPT)—Focusing on the inferences of extrapolation and ramification*. Unpublished doctoral dissertation, Iowa State University, Ames, IA.
- Ma, J., & Cheng, L. (2018). Preparing students to take tests. In J. I. Liontas, M. DelliCarpini, & S. Abrar-ul-Hassan (Eds.), *TESOL Encyclopedia of English language teaching* (1st ed.). <https://doi.org/10.1002/9781118784235.eelt0321>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. London, UK: Blackwell Publishing.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). New York, NY: American Council on Education & Macmillan.
- Miller, F., & Alvarado, K. (2005). Incorporating documents into qualitative nursing research. *Journal of Nursing Scholarship*, 37(4), 348–353.
- 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.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30, 109–162.
- Schmidgall, J. E. (2017a). *Articulating and evaluating validity arguments for the TOEIC® tests* (ETS-RR-17-51). <https://doi.org/10.1002/ets2.12182>

- Schmidgall, J. E. (2017b). *The consistency of TOEIC® Speaking scores across ratings and tasks* (ETS-RR-17-46). <https://doi.org/10.1002/ets2.12178>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 299–317.
- Zumbo, B. D. (2015, November 5). *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. Retrieved on September 11, 2017 from <https://www.youtube.com/watch?v=0L6Lr2BzuSQ>