ITSC | International Testing Services Center

G-TELP™ | General Tests of English Language Proficiency

Stakeholders' Beliefs about the G-TELP Level 2 Test as a Measure of English Proficiency Required for Professional Purposes

2021

Prepared by
Yunjung Nam

Conference
American Association For Applied Linguistics

Abstract

This study aims to explore stakeholder beliefs about the G-TELP (General Tests of English Language Proficiency), which is designed to assess general English proficiency of non-native speakers (ITSC, 2015). The G-TELP Level 2 test is accepted at government agencies and private sectors in South Korea (Choi, 2008) for professional purposes, such as employment or certification process. However, the test has not been the focus of scrutiny from language testing researchers. This study aims to explore the stakeholders' beliefs about the G-TELP Level 2 and to provide valuable insights on what should be further researched for test quality and valid score use.

The current study examines three stakeholder groups: test takers, test preparation course instructors, and test score users. The data include responses to online surveys (125 test takers, seven instructors, six score users) and semi-structured interviews (10 test takers, two instructors, one score user). The survey and interview questions were adapted from previous studies (Malone & Montee, 2014; Murray et al., 2014). This study reports the results from descriptive statistics of the survey data in connection with qualitative analysis of the interview data. The findings suggest that the grammar section is perceived to be much easier to receive a high score on than other sections of the test, while the listening section is so challenging that test takers with lower English proficiency choose the strategy of giving up. The test takers perceived the G-TELP as "a rite of passage" rather than a measure of English proficiency. Implications and recommendations for test developers are also provided.

Stakeholders' Beliefs About the G-TELP Level 2 Test

as a Measure of English Proficiency Required for Professional Purposes

In the field of language testing and assessment, it is widely agreed that a test validation process should include various stakeholders to provide evidence for test score justification and use (Messik, 1996; Moss et al., 2006). To provide evidence for validity arguments, previous studies have collected evidence for validity issues of the TOEFL iBT or IELTS from diverse stakeholder groups, such as test takers (Cheng & DeLuca, 2010), language instructors (Llosa & Malone, 2017) or test score users (Hyatt, 2013). Particularly, Malone and Montee (2014) explored three different stakeholder groups: administrators, teachers, and students in academic settings, focusing on the TOEFL iBT. Based on the findings, they provided several recommendations, including educating administrators and doing more research on the speaking section. Their research certainly demonstrated how investigation of stakeholders' beliefs could provide valuable insights about what should be further explored for better tests and valid scores use. Inspired by the importance of test validation and the previous research on stakeholder beliefs in the field of language testing and assessment, this study aims to investigate stakeholder beliefs about the G-TELP (General Tests of English Language Proficiency), which has been well accepted as an English proficiency test for high-stakes decisions in South Korea.

The G-TELP is designed by International Testing Services Center (ITSC) to assess general English proficiency of non-native speakers (ITSC, 2015). There are 5 levels of G-TELP, with Level 1 targeting the highest proficiency level. The Level 2 test is designed to measure skills for authentic English in general communication (ITSC, 2015) and is accepted at government agencies and private sectors in South Korea (Choi, 2008). Despite the acceptance of the test scores for professional purposes, the test has not been the focus of scrutiny from

language testing researchers. Existing research has used the G-TELP merely as a research instrument to measure participants' reading ability (Taki, 2017) or to make comparisons with other proficiency tests, such as the TOEIC (Lee, 2018; Mihara, 2015). This study focuses on the G-TELP Level 2 as a measure of English proficiency for professional purposes, such as employment, certification, or policy decisions, to explore the stakeholders' beliefs about the test. The current study aims to examine three stakeholder groups: test takers, test preparation course instructors, and score users.

Literature Review

Stakeholder Beliefs in Language Assessment

Language proficiency test scores are used for high-stakes decisions, with the assumptions that the test tasks reflect the skills they purport to measure and that the meaning of test scores is clearly interpretable by score users (Chapelle et al., 2008). It is important that test results provide valid scores that reflect the language ability of test takers, since test performance has consequences on high-stakes decisions for educational promotion or employment. Extending Messick's (1996) discussion on evidential and consequential bases of score interpretation and use, Haladyna and Downing (2004) argued that evidence should be collected to examine construct representation (e.g., whether the errors in test performance are attributable to measurement of test construct) and construct-irrelevant variance in test performance (e.g., factors influencing test performance, such as social factors, that are not directly linked to test construct) (Cheng & DeLuca, 2011). To gain evidence for validity issues regarding construct underrepresentation and sources of irrelevant variance, the evidence should be collected from multiple stakeholders (Moss, 1996). Language testing researchers agree that the test validation process should include various stakeholders, such as test developers, test takers, administrators,

scorers, or test score users (Haladyna & Downing, 2004; Kane, 2002; Messik, 1996; Moss et al., 2006). Researchers have investigated the validity issues, but as Bachman (2000, 2007) pointed out, previous research has been conducted mostly from the perspective of test designers or test developers. Recently, there has been growing interest and awareness about collecting evidence for validity issues from other stakeholder groups such as test takers or test score users (e.g., Fox & Cheng, 2007). However, more research on diverse stakeholder groups is needed in the field to collect, report, and document validity evidence of language proficiency tests.

Test Taker Beliefs: Perceptions and Strategies

Among the different stakeholder groups in language assessment, test takers of standardized English tests have been investigated in terms of their attitudes and perceptions about tests and strategies. Test takers' performance on standardized English proficiency tests has intended or unintended consequences on their educational path, employment opportunities, career development, even social-emotional health (Bachman, 2000; Cheng & DeLuca, 2011). Given the consequences of tests for test takers, it is critical to understand the perspectives of test takers as stakeholders of language assessment. Previous research examined the attitudes and perceptions of test takers about English tests focusing on only one specific skill (e.g., writing) (Kim, 2017) or only using quantitative methods (e.g., survey) (Ata, 2015; Eom et al., 2010). Perceptions of experienced and prospective test takers toward test preparation courses also have been investigated, although the topic has been under-researched compared to general perceptions of the tests themselves. There have been a few studies that used qualitative methods to investigate the test takers of the TOEFL iBT or IELTS (Cheng & DeLuca, 2011; Hamp-Lyons, 2000; Malone & Montee, 2014).

The TOEFL Internet-based Test (TOEFL iBT) has been administered to millions of test takers since 2005 by the Educational Testing Service (ETS), with the purpose of evaluating ability to use and understand academic English. The TOEFL iBT is designed to effectively reflect the target domain of undergraduate university and aims to operationalize the construct of English for Academic Purposes (EAP). Mirroring academic tasks in university classrooms (Chapelle et al., 2008; ETS, 2007, DeLuca et al., 2013), the TOEFL iBT has four sections, Listening, Reading, Speaking and Writing, including test tasks assessing integrated skills. Several studies on TOEFL iBT test takers, for example, have reported how they perceive the validity of the tests, what section they find the most challenging, or how they prepare for the test (to name a few, Cheng & DeLuca, 2010; Malone & Montee, 2014; Stricker & Attali, 2010; Yu et al., 2017).

Research on test takers of the TOEFL iBT has investigated various topics. For example, there have been several research studies on test taker attitudes toward computer-based testing (Jamieson et al., 1999; Stricker et al., 2004), and keyboarding (typing while writing on computer) skills (Barkaoui, 2014, 2015). Previous research findings revealed test takers' positive attitudes toward the computer-based TOEFL, and Barkaoui (2014, 2015) found the effect of keyboarding skills on test task scores to be significant but mainly impacting the independent writing task.

Test takers' beliefs or perceptions about the TOEFL iBT have been investigated with different methods, scales, and foci. Stricker and Attali (2010) conducted an extensive study of test taker attitudes by administering questionnaires to more than 3,500 participants from four different countries. They found that students from China, Columbia, and Egypt held positive attitudes toward the TOEFL iBT while Germans did not. They also reported that test takers were less positive toward the speaking section than other sections. In a more recent extensive study of

stakeholders' beliefs about the TOEFL iBT, Malone and Montee (2014) used qualitative

methods to understand how different stakeholders including students, instructors, and

administrators perceive the TOEFL iBT as a measure of academic language ability. They found

that German students were generally positive about the test, in addition to students from Saudi

Arabia, South Korea, and the US. However, the students held the least positive attitudes toward

the speaking sections, as also found by Stricker and Attali (2010). There have been more

regional studies on test taker perceptions or attitudes toward the TOEFL. Eom et al. (2017)

examined how students in China and Korea (N=357) perceive test quality, cost, difficulty, and

test constructs of proficiency tests. They found that the TOEFL was perceived as a better test

than others in terms of test constructs and test quality. Unlike other studies, Kim (2017) only

focused on the writing section, analyzing data of 476 postings on a Korean online forum. The

research found that Korean test takers tend to resort to "templates" in compensation for their lack

of self-belief in their writing competency.

More qualitatively oriented test taker research studies have been recently conducted with

an emphasis on the importance of voices from the test takers. For example, Cheng and DeLuca

(2011) found a complex and interwoven relationship between test-taking and its consequences,

using data collected from the writings of 57 participants. Malone and Montee (2014) used focus

group interviews first to generate survey questions and conducted stimulated recalls. Based on

their findings that test takers were least positive toward the speaking section, they called for

more research on the speaking section.

As for another commonly used English proficiency test for English learners, stakeholder

beliefs about the International English Language Testing System (IELTS) also have been

explored. There have been several regional studies that investigated the attitudes of test takers or

candidates of the IELTS. Ata (2015) reported how much international students in Australia (N=200) know about the test and how they perceive each section of the test. The students were found to be positive about the test quality of the IELTS. Students from China and Arabic-speaking countries were the least positive toward the speaking section. The influence of topic familiarity on task performance was indicated as a reason for dislike, and the difficulties experienced in the speaking section were more varied and pronounced than those in the writing section. Zahari and Dhayaalan (2016) explored how Malaysian students changed their attitude toward the test before and after taking the test. They found more favorable attitudes among test takers after they actually took the test compared to their attitudes during the preparation stage.

There have been some studies focusing on a specific section of the IELTS. For example, Silahi (2014) focused on the writing section of the test to investigate attitudes of two IELTS teachers and 16 test takers in Indonesia. The participants were found to hold positive attitudes toward the IELTS writing section because it was helpful for cultivating critical thinking skills and it can reliably assess English proficiency. On the other hand, the listening section has been the focus of scrutiny in Tran and Nguyen (2018). They found that Vietnamese students generally had negative affective and cognitive attitudes to the listening section, which means they have low interest and confidence in the listening test and most of the tasks were perceived as challenging to them.

As discussed above, triangulated data from test takers using both quantitative and qualitative methods seems to contribute to more comprehensive understanding of test taker voices. Millions of English learners around the world take proficiency tests either to apply to college or graduate school admissions, to exit programs, to achieve scholarship or certification, or to obtain a working visa. Test takers' test performance on standardized English proficiency

tests has intended or unintended consequences on their educational path, employment

opportunities, career development, and even the social-emotional health of the test takers

(Bachman, 2000; Cheng & DeLuca, 2011). Considering the consequences of standardized

English proficiency tests on test takers, it is critical to understand perspectives of test takers as

stakeholders of language assessment. The review of literature underscores how important it is to

include test takers as stakeholders to better understand what tests are measuring and how tests

can be improved.

Instructor and Test Score User Beliefs

While test takers' perceptions can be useful for test validation, experts have suggested

that another group of stakeholders, instructors, can provide valuable input for improving test

validity (e.g., O'Sullivan et al., 2002). Especially, in terms of test design, instructors' beliefs

might provide useful information (Meisels et al., 2001) because they can provide constructive

suggestions for more authentic test tasks. One teacher-verification study (Cumming et al., 2004)

investigated instructors' attitudes toward prototype tasks for the new TOEFL. In the study,

instructors were asked to evaluate whether the content of prototype tasks reflected the domain of

academic English at North American university settings and whether the prototype tasks

reflected the authentic language use in the classes. The study collected data from individual

interviews with seven ESL instructors at three universities. The results indicated that the

instructors were found to positively evaluate the prototype tasks and provided critical feedback

for task design. This study demonstrated the benefits of collecting instructor input for test task

and validity despite being a small-scale exploratory study. Since the TOEFL iBT was introduced,

several studies have investigated university instructors' perceptions. Malone and Montee (2014),

in their stakeholder beliefs study, found that the US university instructors generally agree that

performance on the TOEFL iBT integrated speaking and writing tasks reflects performance in academic classes, which provides evidence for construct validity. However, integrated writing tasks alone were not considered representative of the writing that TOEFL iBT test takers are expected to do in writing courses after college admission (Llosa & Malone, 2017). Llosa and Malone explored student and instructor perceptions of writing tasks and performance on the TOEFL iBT. In their study, 18 instructors completed a questionnaire, and six instructors were interviewed. The findings indicate that the combination of both independent and integrated tasks reflected many characteristics of writing course assignments. The instructors also believed that their grading rubrics were similar to the criteria of TOEFL iBT writing rubrics. Their perceptions provide additional evidence to supporting "the extrapolation inference" (Chapelle et al., 2008, p.11). in the TOEFL validity argument.

While the perceptions of instructors teaching in the target domain have been explored, another group of instructors has also received attention. The group of test preparation course instructors who are believed to be familiar with the target test and test takers' preparation process have also been investigated in the frame of washback studies. In the language testing field, washback refers to "the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (Messick, 1996, p. 241). While washback can be either positive or negative (Taylor, 2005), it is considered complex and context dependent (Green, 2013). With the primary focus on the washback effects, the previous studies on preparation course instructors have investigated what materials instructors use and what methods teachers use to increase the test scores. Alderson and Hamp-Lyons (1996) found that TOEFL preparation course teachers included more test-taking strategies and fewer interactive activities when compared to non-TOEFL courses. In

another study on TOEFL test preparation, Hamp-Lyons (1998) found that instructors were not effectively using TOEFL test preparation textbooks as intended. When the new TOEFL iBT was introduced, instructors were also influenced by the change. In the earlier stage of the change, instructors were found to have limited knowledge about the new version (Wall & Horák, 2008). As for instructional strategies the instructors used, Barnes (2016) explored Vietnamese teachers who were teaching TOEFL preparation courses and found that the majority of teachers focused on lecturing and test-taking skills rather than involving students in more interactive pair or group work. Similarly, Wang (2019) also found that TOEFL ITP course instructors were mostly relying on practice of test-taking skills, using more teacher-centered activities and individual practice rather than peer collaborative work. The study also found that students wanted more interactive and fun classes even though it was a test preparation course, which was one of the recommendations for the instructors. As discussed so far, test preparation course instructors' perceptions and practice were investigated only in the framework of washback. However, as they should have a high level of understanding about the target test, preparation instructors also can offer some insights that are useful for test validation. They are also the direct observers of how test takers prepare for the target test, so it seems necessary to explore their beliefs about the target test.

In the discussion of test validation, it is argued that including multiple stakeholders is crucial for validity and validation studies (Moss, 1994, 1996; Moss et al., 2006). Different groups of stakeholders might have different interpretations of the same scores (Koch & DeLuca, 2012; Macqueen et al., 2016; Zumbo, 2015). The involvement of multiple stakeholders would provide more evidence either supporting or challenging the intended use of test scores. To this end, language testing researchers have explored test score users of English proficiency tests such

as the TOEFL or IELTS for admission purposes. More recently, with the efforts to define, conceptualize, and promote the language assessment literacy of stakeholders such as test score users and policy makers, there is an increasing interest among language testing researchers on the score user population. In the current literature, undergraduate/graduate admissions officers and university faculty have been the focus of several empirical studies (Baker et al., 2014; Ginther & Elder, 2014). Collectively, the researchers call for more education for the test score users for valid use of the test as intended by the developers. For the context other than North American educational institutions, Im and McNamara (2017) investigated the use of Test of English for International Communication (TOEIC) scores for college admission purposes in Korea, arguing that use of the TOEIC scores for college admission lacks validity. Their participants included 20 university faculty members involved in undergraduate admission. They viewed TOEIC scores as a reflection of candidates' effort or test-taking skills rather than an indicator of English proficiency level or academic aptitudes for their undergraduate programs. This would pose a validity threat because the TOEIC is intended to measure English proficiency for the context of international communication. Despite the importance, however, there are still very few empirical studies that involved the score users of language proficiency tests to gather evidence for construct representation and valid score use.

English Proficiency Tests and Score Use in South Korea

In South Korea, English proficiency tests have been used to make consequential high-stakes decisions such as hiring and promotion (Choi, 2008). With the increasing importance of international business communication and collaborations in various areas, the English language is considered an essential tool for improving self-competitiveness for Korea (Jeon, 2012), and job seekers are required to submit evidence of their English language ability. In Korean labor

markets, English tests are mostly used to represent individuals' readiness for workplaces and improve job seekers' *specs*—an abbreviation of specifications—which is widely used to refer to individual's capabilities for a certain job (Kim et al., 2019). Job seekers in Korea, thus, take standardized English tests such as the G-TELP, TOEIC, TOEIC Speaking, Test of English Proficiency developed by Seoul National University (TEPS), the Oral Proficiency Interview - Computer (OPIc), or the Foreign Language Efficiency Examination (FLEX) for their employment or promotion.

In addition to the use of English proficiency test scores for high-stakes decisions in private sectors, government agencies started to replace their own English test with standardized English test scores (Im, 2019; Lee, 2018). The replacement policy was in response to the negative attitudes toward the construct validity of English subject in national exams (Kang et al., 2011; Kim, 2007). The Ministry of Personnel Management (2018) introduced "Public Official Appointment Examination Decree [President Decree No. 26233, Enforcement 2015. 5. 6.]" to replace the English subject in national exams with standardized English test scores, including the G-TELP, TOEIC, TOEFL, TEPS, and FLEX. This amendment's purpose was to evaluate candidates' general English ability through more standardized testing and to alleviate burdens of preparing for the national exams, including the Foreign Service and Civil Service Exams (Grade 5 and Grade 7). Currently, the replacement of the English subject is also effective for the Bar exam, Certified Public Accountant (CPA) exam, Certified Tax Accountant exam, Certified Public Labor Attorney exam, Certified Tourist Guide exams, or even overseas assignment of public officials (Choi, 2008; Im, 2019).

With the use of English test scores for high-stakes purposes, researchers have investigated the topics around the policy and practices. One of the notable studies is an analysis

study of test usefulness of standardized English tests used for Civil Service English Exam (CSEE) by Lee (2018). The analysis study invited experts to evaluate five English proficiency tests used in Korea, including the TOEIC, TOEFL, TEPS, G-TELP and FLEX, in terms of their construct validity, authenticity, and impact. The results indicate that the use of the tests as alternatives to CSEE was not valid because the five tests lack construct validity. As for authenticity, experts found the TOEIC and TEPS to be more authentic than other tests. However, the study did not specifically operationalize the authenticity in the context of civil service workplace context. In addition, the study conducted a survey on the test takers of civil service exams. It was found that they believed that the use of the five tests did not bring a positive impact to civil service job candidates or society. The study suggests that an independent CSEE should be developed based on the target language use domain and include more test items that assess English language ability useful and relevant for civil servants' job responsibilities.

Other researchers specifically focus on the use of the TOEIC test in Korea. Im and McNamara (2017) found that the use of the TOEIC test was not valid for college admission decision-making purposes. This was due to the common perception among the participants that the TOEIC score is an indicator of test takers' efforts and test-taking skills. Similarly, JTBC News (2016) reported that job seekers believed that the TOEIC scores demonstrate test takers' diligence and perseverance rather than their English proficiency (as cited in Kim et al., 2019). Kim et al.'s study was based on the survey data collected from 252 Korean job seekers to investigate their perceptions about the TOEIC, TOEIC Speaking, and OPIc. Based on the analysis of their survey responses, it was found that the tests were only perceived merely as one of the requirements for job seeking rather than an indicator of English proficiency. Furthermore, findings suggest that motivation to learn English did not have positive influences on score

improvement while English test stress was found to be the only significant predictor of lower scores.

Based on the review of the literature, it seems that the TOEIC has been used for multiple purposes in different contexts in Korea, and the G-TELP is also used for similar purposes. According to ITSC and G-TELP Korea, the G-TELP has been used as an alternative to English tests for recruiting competent civil employees, fulfilling university graduation requirements, completing certification programs, or hiring employees for private sectors. Nevertheless, unlike the TOEIC, there have been very few studies to investigate the validity of the G-TELP scores use.

As discussed above, in Korea English proficiency tests have been used with additional meanings and interpretations. In test validation process, it is important to consider the defensibility of intended score meaning and scores use (AERA, APA, & NCME, 2014; Messick, 1989). As the G-TELP scores are used for multiple purposes in Korea, sometimes even in an unintended way, there is an urgent need for more empirical research involving multiple stakeholders to investigate the score meaning and use of the tests. However, as Kane (1992) pointed out, the interpretations of scores might be different among the stakeholders, so this study aims to investigate multiple stakeholder groups in terms of their beliefs toward the G-TELP as a measure of English proficiency required for professional purposes.

The Current Study

As discussed in the literature review, it is critical to investigate stakeholders' beliefs to gain useful insights for test quality control and valid scores use. While the English proficiency tests administered in Korea, such as the TOEIC or TOEFL, have been explored by language testing researchers locally and internationally, the G-TELP has not been explored in terms of

stakeholder beliefs. For this reason, this study aims to investigate stakeholder beliefs about the G-TELP Level 2, which has been used for high-stakes decisions in South Korea.

The current study aims to answer the following research questions:

1. What strategies do test takers and instructors perceive to be effective for preparing for the G-TELP Level 2 test?

2. What are stakeholders' beliefs about what each section of the G-TELP Level 2 measures?

3. How do different stakeholders perceive the English language ability required for professional purposes?

Methods

The G-TELP Level 2

General Information about the G-TELP

The G-TELP is designed to assess general English proficiency of non-native speakers (ITSC, 2015). There are 5 levels of G-TELP, with Level 1 being the highest. According to ITSC, the G-TELP is different from other English language proficiency tests in several ways. First, G-TELP tests are criterion-referenced. The G-TELP is developed based on universally recognized criteria for performances in skill areas of grammar, listening, and reading/vocabulary. Second, G-TELP tests provide detailed task-referenced information on the examinee's performance in the form of a profile report. ITSC claims that the information is "invaluable for decision-makers in the industry and for educational purposes." The third and the most important characteristic of the G-TELP would be that the test is designed as a measure of "general English language proficiency" instead of focusing on proficiency in academic (e.g., TOEFL) or business (e.g., TOEIC) contexts. Finally, unlike other proficiency tests that assess multiple levels with a single test, the G-TELP is designed to evaluate test takers at five different levels. Among the five levels,

the G-TELP Level 2 is designed to assess authentic and modified English in normal

communication, according to ITSC:

> This level assesses the ability of a test taker to use the language outside of
>
> classroom situations. This test taker is able to cope with some authentic
>
> English, has had contact with some authentic English, and has had contact
>
> with native speakers. Although his/her learning of the language has been
>
> classroom-based, the examinee is able to communicate with a native
>
> speaker within a wide range of tasks.

The G-TELP Level 2 is composed of three sections with 80 questions in total, and the test is 90

minutes. ITSC refers to the three sections as three "skills areas." However, this study uses

"sections" instead of "skills areas" to avoid confusion about what kind of preparation "skills" test

takers use for test preparation and what "skills" instructors focus on in their classes to help test

takers better prepare for the G-TELP. Table 1 shows the general description of the lengths and

numbers of questions for each section.

Grammar Section

The grammar section includes 26 questions of sentence completion type items. The

stimuli generally include two sentences with one blank. Test takers have 20 minutes to complete

this section. As described in Table 1, this section is designed to assess understanding of the basic

and more complex grammatical structures, such as tense, gerunds, conditionals, conjunctions,

subjunctives, and infinitives.

Listening Section

The listening section includes 26 questions with six or seven questions per passage. This

section is approximately 30 minutes, with four different passage types: casual conversation of

personal narratives, talk/presentation of product, informal conversation about pros and cons for

negotiation or decision-making, and talk/presentation explaining a non-technical

procedure/process or guidelines. The listening test questions are presented aurally, and only

multiple-choice options are presented on the test sheet. The audio is delivered at a normal rate of

speech.

Table 1

The G-TELP Level 2 Section Information

Section	# of questions	Length (Min.)	Description
Grammar	26	20	An examinee at this level is required to demonstrate understanding of the basic grammatical structures of English and some of the more complex grammatical structures.
Listening	26	30	An examinee at this level can understand authentic English which has been reworded, paraphrased and simplified, but which is delivered at a normal rate of speech. Tasks at this level include: - everyday conversations involving informal situations - formal monologues in which a particular product or course of action is recommended - discussions on pros and cons of certain situations - explanations of a non-technical procedure or process, or informative/useful/helpful tips for specific situations
Reading/ Vocab	28	40	An examinee at this level can read authentic and modified texts of a general nature. Modified texts are defined as texts in which low-frequency and idiomatic vocabulary is simplified. At this level, syntax is not modified. Tasks at this level include: - narratives that focus on historical or contemporary figures - magazine and newspaper articles that describe current social or technological subjects, or study/survey results - encyclopedic articles that are general and non-

> technical
> - commercial correspondence that is descriptive and persuasive
>
> A person at this level is also required to demonstrate understanding of selected vocabulary in the context of the reading tasks described above.
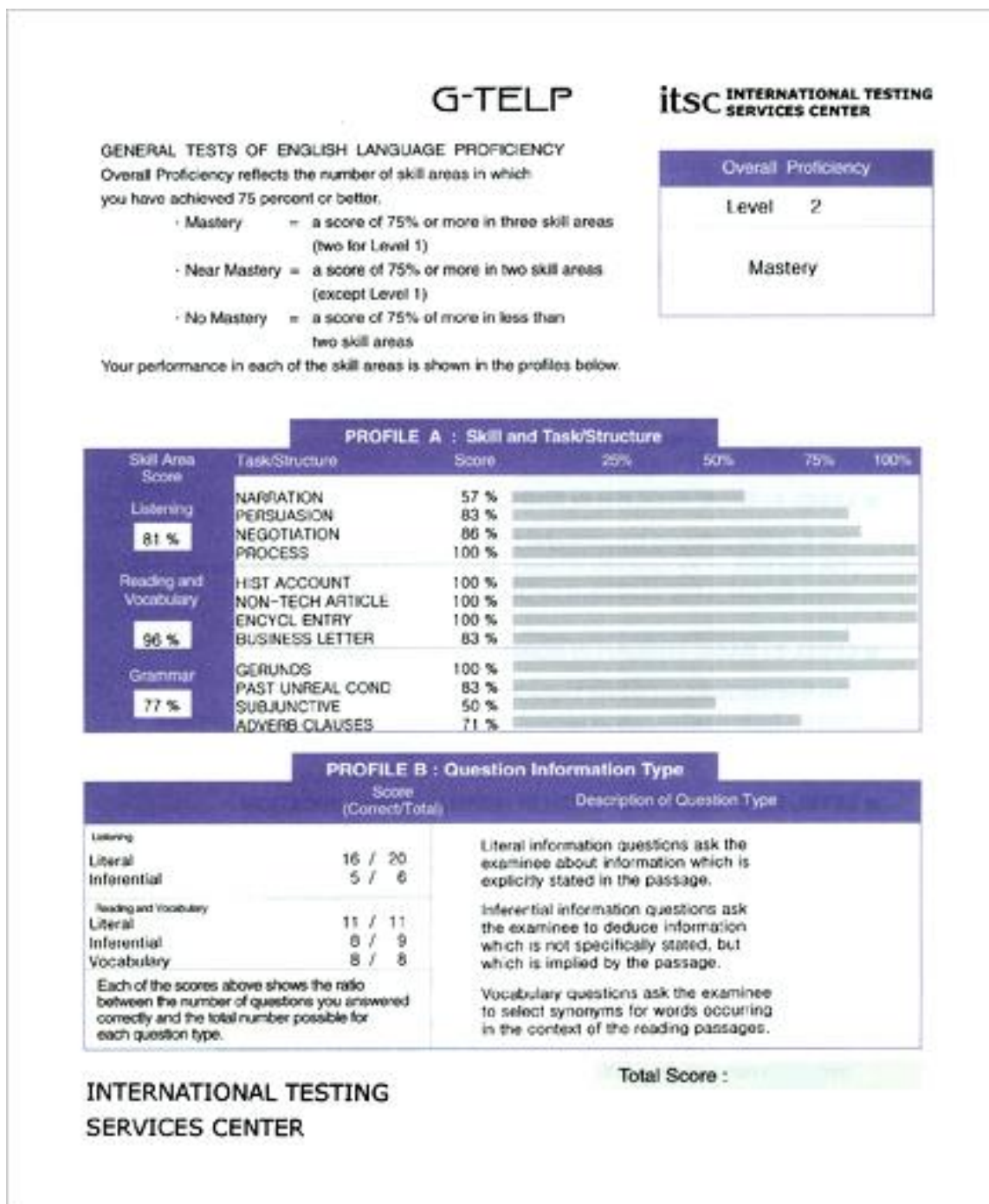
Reading and Vocabulary Section

The reading section includes 28 questions, with seven questions per passage. Each of the four passages comes from a different genre: biographical narrative, magazine or web article, encyclopedia article, and formal-style business or personal letter. Unlike the TOEIC, which focuses on business communication contexts, the G-TELP covers wide variety of topics and contents in reading. Among the seven questions, two are vocabulary questions asking about the contextual meaning of a word. The other five questions ask test takers to identify the main idea or the purpose, find detailed information, or make appropriate inferences. For the reading section, the recommended time for completion is 40 minutes. However, a recent change is worth noting to understand this study. In the past, test takers were not allowed to work on other sections outside of the designated time. Under the current testing administration policy, however, test takers can work on any sections at their disposal. For example, test takers can work on the reading section while the audio is being played for the listening section.

Score Report

The results of the G-TELP are released within five business days after the test date. This is a much faster score release policy compared to the TOEIC, which takes around 20 days in Korea. The test takers will receive the G-TELP score report, which provides a level mastery score and two profiles of the examinee's performance (see Figure 1).

Figure 1

The G-TELP Score Report

G-TELP

itsc INTERNATIONAL TESTING SERVICES CENTER

GENERAL TESTS OF ENGLISH LANGUAGE PROFICIENCY
Overall Proficiency reflects the number of skill areas in which you have achieved 75 percent or better.

- Mastery = a score of 75% or more in three skill areas (two for Level 1)
- Near Mastery = a score of 75% or more in two skill areas (except Level 1)
- No Mastery = a score of 75% of more in less than two skill areas

Your performance in each of the skill areas is shown in the profiles below.

Overall Proficiency
Level 2
Mastery

PROFILE A : Skill and Task/Structure

Skill Area Score	Task/Structure	Score	25%	50%	75%	100%
Listening 81 %	NARRATION	57 %				
	PERSUASION	83 %				
	NEGOTIATION	86 %				
	PROCESS	100 %				
Reading and Vocabulary 96 %	HIST ACCOUNT	100 %				
	NON-TECH ARTICLE	100 %				
	ENCYCL ENTRY	100 %				
	BUSINESS LETTER	83 %				
Grammar 77 %	GERUNDS	100 %				
	PAST UNREAL COND	83 %				
	SUBJUNCTIVE	50 %				
	ADVERB CLAUSES	71 %				

PROFILE B : Question Information Type

	Score (Correct/Total)	Description of Question Type
Listening		Literal information questions ask the examinee about information which is explicitly stated in the passage.
Literal	16 / 20	
Inferential	5 / 6	
Reading and Vocabulary		Inferential information questions ask the examinee to deduce information which is not specifically stated, but which is implied by the passage.
Literal	11 / 11	
Inferential	8 / 9	
Vocabulary	8 / 8	

Each of the scores above shows the ratio between the number of questions you answered correctly and the total number possible for each question type.

Vocabulary questions ask the examinee to select synonyms for words occurring in the context of the reading passages.

INTERNATIONAL TESTING SERVICES CENTER

Total Score :

Profile A shows the percentage of correct answers for all the questions in each section along with the percentage of questions that the examinee answered correctly for different types of tasks or structures. Profile B provides information about the examinee's performance in different

question types, such as understanding explicitly stated information, drawing inferences from

explicitly and implicitly stated information, or selecting synonyms for words occurring in the

context of the reading passages. When the examinee achieves skills area scores of 75% or greater

in all three sections, the overall proficiency will be considered "mastery." If the examinee

achieves 75% in only two skill areas, the proficiency level is "near-mastery." If the examinee's

scores are below 75% in all skills areas, the overall proficiency is "no mastery."

Instrument Development

The researcher prepared three online surveys, one each for test takers, test preparation

course instructors, and test score users. Based on the literature review about stakeholder beliefs

about a standardized language proficiency test, the questionnaires from Malone and Montee

(2014) were chosen because they also included different stakeholder groups in the research.

Their survey questions were adapted to suit the context of the G-TELP test administered in

Korea.

To make necessary adaptations to the online surveys, the researcher first analyzed the

basic information about the G-TELP, the test specifications provided by ITSC, and the G-TELP

preparation materials. In addition, to better design survey questions about test takers' strategies

and perceptions, blog and forum postings written by G-TELP test takers were reviewed by the

researcher. For example, unlike the TOEFL iBT, the G-TELP does not include writing and

speaking sections, so questions and options about writing and speaking sections were removed

from the survey. Also, as the G-TELP is designed for professional purposes in general, any parts

related to "English-speaking universities" were changed into "workplaces." Moreover, as the

review of test takers' blog and forum posts indicated that "studying grammar points" was one of

the major strategies, it was added as an option to the survey questions about preparation strategies and the grammar section of the G-TELP.

One major addition to the survey for test takers (see Appendix A) was the section for English proficiency level. This section was added in the hopes of better understanding the participants' profiles. The researcher expected that this information would indicate if those who chose to take the G-TELP instead of other tests came from a certain proficiency group or not. The information gathered would also allow further analysis on any differences among proficiency levels in test preparation strategies and their perceptions. For practical challenges, it was not possible to measure the participants' language proficiency level with controlled instruments for the current study. Instead, the researcher decided to include questions to gather self-reported proficiency information. The participants were asked to indicate their proficiency levels by responding to a series of questions. One question asked them to indicate their proficiency for subskills of listening, reading, speaking, and writing, using the three levels of beginner, intermediate and advanced. However, to compensate for the possible issues of subjective interpretations of each level, the researcher decided to utilize self-assessment descriptors described in the Common European Framework of Reference for Languages (CEFR) created by the Council of Europe (CoE) in 2001.

Another major addition was a set of questions in the test score user survey (see Appendix C). The researcher made preliminary efforts to identify potential participants for surveys and individual interviews. Conversations with three professionals at public agencies (e.g., city government, public agency, police station) and three HR personnel at private companies indicated that the G-TELP was not widely known to English proficiency test score users for professional purposes. However, in the hopes of exploring test score users' beliefs about the G-

TELP, the researcher decided to include questions that present the sample test items and

reading/listening passages that were publicly available. The participants were asked to carefully

examine the sample items and passages and then rate whether it is necessary to include those test

items and text types for evaluating English proficiency required for their own workplaces.

Once the surveys were finalized (see Appendix A, B, and C for English versions), they

were translated into Korean by a native Korean speaker, who was a master's student in Applied

Linguistics at the time of translation. She had some knowledge about language testing and

assessment and suggested minored changes. After translation was completed, one middle school

Korean language arts teacher was invited to check the clarity of the language used in the surveys.

Individual interview questions for different stakeholder groups (see Appendix D for an

English version) were also adapted from Malone and Montee (2014) and translated into Korean

by the same translator.

Recruitment and Data Collection Process

Surveys were conducted online via Qualtrics from May to August in 2020. The link to the

surveys was distributed by posting the survey information on online forums (e.g.,

https://cafe.naver.com/G-TELPmaster). The surveys were also distributed through emails to

potential participants introduced by the researcher's acquaintances.

The G-TELP preparation course instructor participants were identified through an online

search or recommended by test takers who were interviewed. The instructors were contacted

using their public information disclosed for their marketing or professional purposes. Test score

users were not easily accessible because they are usually decision makers at their institutions

dealing with confidential information. Possibly for that reason, the initial method of participant

recruitment (i.e., online advertising) did not generate any responses. Previous studies such as

Alderson's (2010) also described a similar challenge in data collection, reporting that the final survey response rate was only 30%. In Alderson's research, two respondents gave confidentiality as a reason for not providing information, or they only provided vague answers. In addition, the researcher of the current study was not familiar with the process of decision-making for utilizing English proficiency test scores for hiring or promotion purposes. Therefore, the researcher decided to consult Korean professionals in public and private sectors, including two public officials working for a government agency and two office workers working for companies. They recommended potential participants. Using the information gained from these consultations, emails were sent to public email addresses of potential participants. Once the potential participants' email addresses were shared with permission, the researcher sent emails including information about the survey and the interview.

Among the survey participants, volunteers for interviews were contacted using the information they provided. The interviews were conducted in Korean through a telecommunication mobile application, KakaoTalk. Each interview ranged from 22 minutes to 70 minutes. The interviews were audio recorded and transcribed by the researcher. The participants were paid $10 for completing the online surveys and $30 for participating in the individual interviews.

Participants

The researcher initially received 198 responses to the survey. However, inspection of data revealed that there were 60 spam responses to the surveys. They were removed by the researcher. For the data analysis, the responses from 125 test takers, seven test prep instructors, and six test score users were included. A total of 13 participated in the interviews, including 10 test takers, two test prep course instructors, and one test score user.

Data Analysis

Survey

The data collected from the 138 participants (125 test takers, seven instructors, and six

score users) were descriptively analyzed to answer the research questions. Cronbach's alpha

reliability was checked for the surveys. The results show that the reliability of responses to all

surveys were in the acceptable range, with 0.91 for test takers, 0.86 for test preparation course

instructors, and 0.90 for test score users. The survey responses were first organized in Microsoft

Excel sheets and descriptively analyzed using R (R Core Team, 2019). The data were plotted

using Excel and R to create more visually intuitive plots using different types of data. In addition,

error bars were added to the plots wherever possible. For example, when the data were based on

a Likert-scale question, error bars were added to make more rigorous comparisons for

respondents' beliefs. In other cases, the responses were analyzed in terms of comparative

frequencies and percentages of responses.

Interview Data Coding Process

The transcribed interview data were coded by the researcher and a second coder who is a

Korean native speaker PhD student in applied linguistics. The major themes in the initial coding

scheme were adapted from Malone and Montee (2014), but the coding scheme was further

revised and refined through iterative coding process. After three rounds of initial coding process,

the first coder (i.e., the researcher) had a 60-minute meeting with the second coder for a training,

where they reviewed the coding schemes and co-coded 10% of data. They also resolved any

discrepancies once they arose, and revisions were made to the coding scheme based on the

discussion. After the meeting, the coders also had a discussion through online communication to

revise the coding scheme by adding more themes or subcategories and removing unnecessary or

irrelevant codes. The final version of the coding scheme is displayed in Table 2. Among the entire data set, 20% of data were coded for checking intercoder reliability. Using exact agreement percentage, the intercoder reliability reached 0.89, which was within the acceptable range. Discrepancies were resolved through discussion.

Table 2

Coding Scheme and Illustrative Quotes

Themes	Sub-codes	Illustrative Quotes
Experiences of English Learning and Use	English learning/testing history or experiences	I have studied abroad. I have taken the TOEIC before.
	Self-evaluation of English ability	I can ask for directions in English. My listening ability is not that great.
	Use of English at workplaces	I rarely use English at workplaces. It depends on professions. If it is necessary, you should practice speaking while on the job.
Test-taking experiences	Reasons for choosing the G-TELP over other tests	You can prepare for the G-TELP within a shorter period of time than the TOEIC. The G-TELP has a smaller number of test items than the TOEIC.
	Familiarity with the test	I have never heard of this test before. It is my second time to take the test. I had to take it again because the score has expired.
Preparation for the test	General Preparation	I read the online forum reviews, and I bought a practice test book. People should use different preparation strategies depending on their proficiency levels.
	Grammar	I only memorized the grammar points that are covered in the test. I watched freely available online lectures for grammar section.
	Listening	I learned about the question formats, but I did not study to prepare for listening section. Even if you study, it does not help increase the

		score.		
	Reading & Vocab	I learned about the reading passage types. I practiced with some sample questions.		
Test-taking Strategies & Perceptions of the G-TELP	Overall perception	Overall, it was easier than the TOEIC.		
	Grammar	The test range is narrower than the TOEIC. Question stems of the G-TELP are generally longer than those of the TOEIC.		
	Listening	Listening was extremely difficult because questions were not printed on the test sheet. I gave up on the listening section and worked on other sections.		
	Reading & Vocab	I had enough time for the reading section compared to the TOEIC. I enjoyed reading the passages because they covered diverse topics and contents.		
Suggestions for Change	Changes in test items or formats (Construct-relevant)	Grammar	Questions might be a little bit more difficult since the test covers a narrow range.	
		Listening	I hope the questions are printed on the test sheet.	
		Reading & Vocab	I hope the reading passages cover more diverse topics.	
	Other aspects (Construct-irrelevant)	I hope there are more testing sites. I hope the registration fee is cheaper.		
Impact of the test	Score interpretation	The score reflects my ability very well because it increases as I make more effort.		
	Gatekeeping purposes in job applications or certification processes	Taking the G-TELP is only to satisfy the requirements. I think someone who can receive a higher score on the G-TELP is more likely to be successful in completing tasks at workplaces.		
	Positive/Negative washback on English	Preparing for the G-TELP does not necessarily help improve my English ability.		

learning and English use on the job	Studying for the Reading & Vocab section would help in comprehending some documents in English on the job.

Stakeholder Background

Table 3 summarizes descriptive statistics for basic background information of participants. In the following, more detailed information about each stakeholder group is discussed.

Table 3

Participants' Background Information

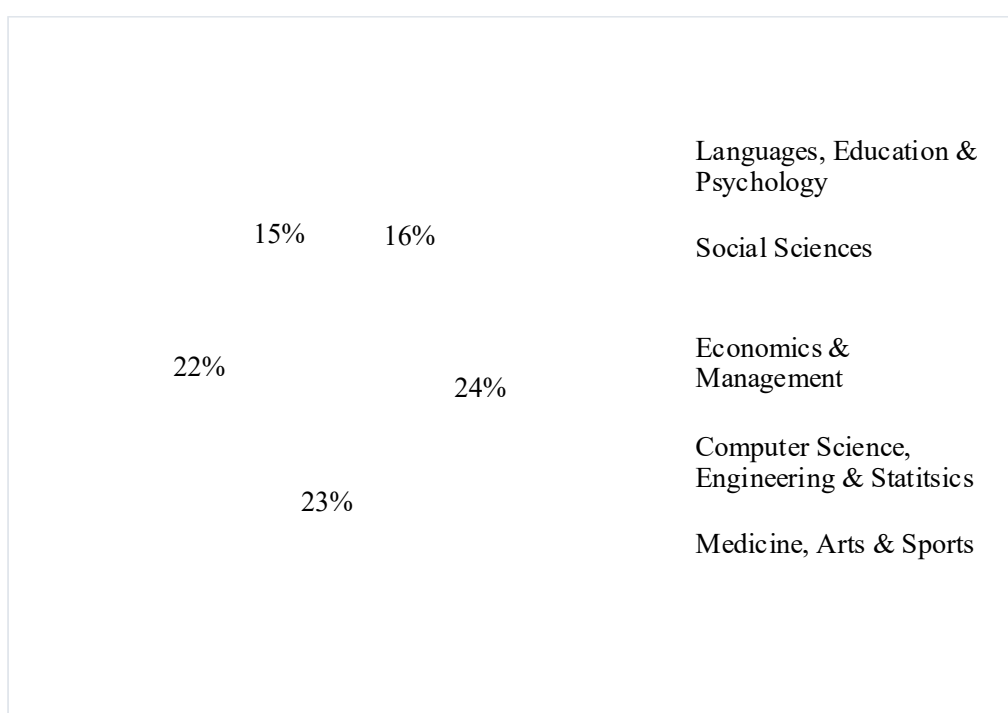| Groups	Survey (N = 138)				Interview (N= 13)			
	Total	Gender		Mean Age (SD)	Total	Gender		Mean Age (SD)
		M	F			M	F	
Test takers	125	41	84	28.94 (5.49)	10	2	8	27.50 (4.34)
Test Preparation Course Instructors	7	3	4	34.71 (5.79)	2	2	0	41 (-)
Test score users	6	5	1	34.00 (10.23)	1	1	0	33 (-)

Test Taker Background

Among the total of 125 test takers, the majority of them (62%) had previously taken the G-TELP (52% 1-2 times, 7% 3-5 times, and 3% more than 5 times) while the rest reported that they had registered for the G-TELP at the time of data collection. There were 41 male (32.80%) and 84 female (67.20%) respondents, with a mean age of 28.94 (SD = 5.49). The youngest participant, who was a college freshman, was 18 years old, while the oldest participant was a 45-year-old job seeker. As for their current or final degree, the majority of the respondents (73%)

were college graduates with a bachelor's degree, followed by high school diploma holders (14%).

There were people with graduate-level education (6% master's and 2% PhD), while 5% did not

respond. Respondents came from diverse majors including economics, management, medicine,

arts, engineering, social sciences (e.g., politics, law), and language-related majors (e.g., English,

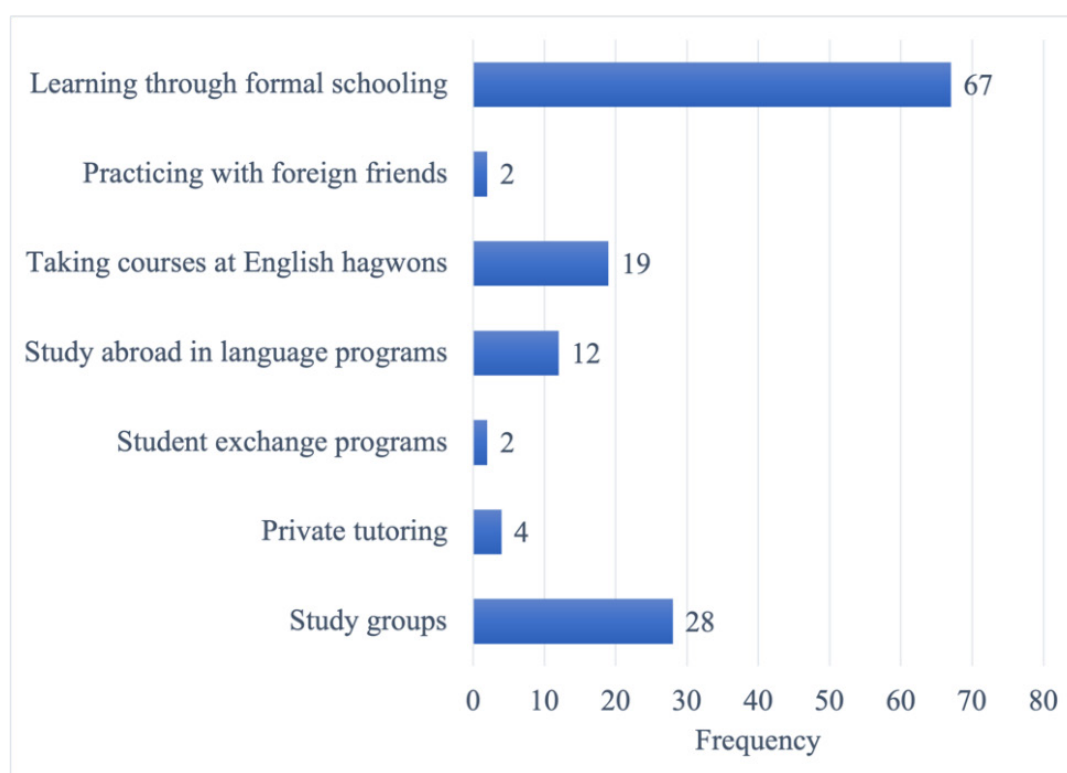Japanese, Korean) (see Figure 2 for proportion of each category).

Figure 2

Proportion of Major

For the current profession, 40% reported that they were job seekers, 26% of respondents

were students (20% college students, 4% MA students, 2% PhD students), and 23% were

professionals as employees in the private sector (16%) or as public officials (7%). Eleven percent

of respondents reported that they were either stay-at-home wives or part-timers looking for

potential job opportunities in the long run.

As for their English learning experiences, their responses to open-ended questions were thematically coded. Results show that the most common learning experience was English learning through their formal schooling (n = 67), while various other experiences such as practicing with foreign friends, taking courses at English *hagwons* (for-profit private institutes, academies, or cram schools prevalent in Korea), participating in study abroad programs or receiving private tutoring (see Figure 3 for detailed information).

Figure 3
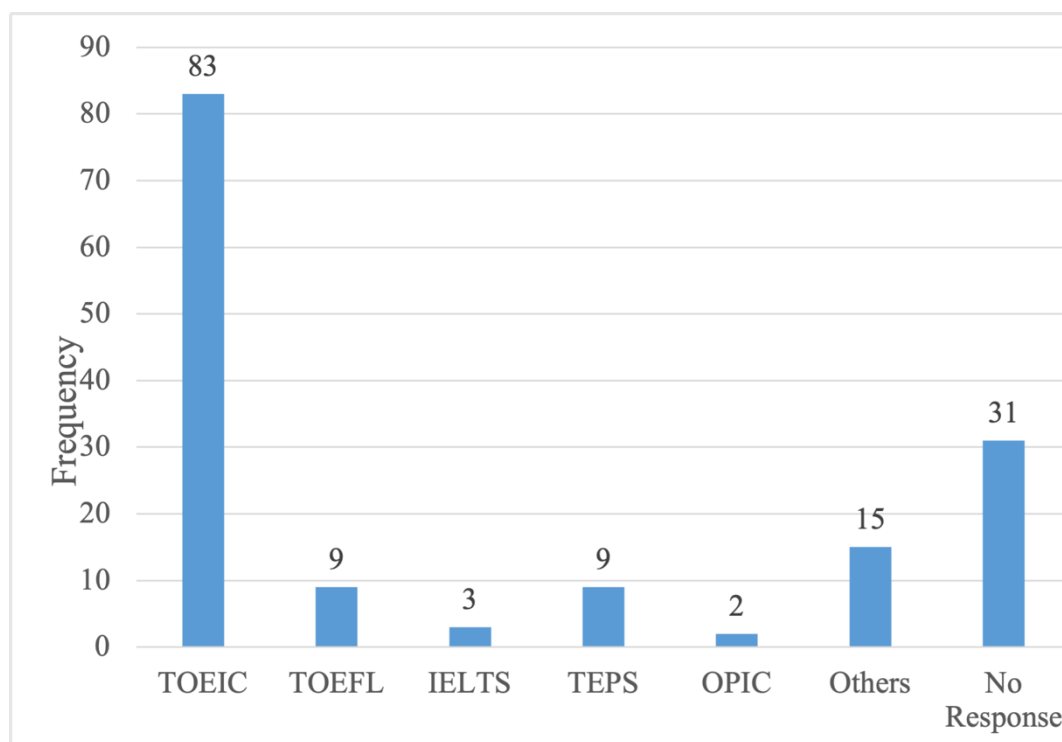
English Learning Methods

Note. Respondents were able to choose more than one answer; thus, the total exceeds 125.

The respondents also provided information about which English proficiency tests they have taken except the G-TELP. The results show that 83 of respondents (66%) have experiences of

taking the TOEIC while there were a relatively small number of respondents who have taken the TOEFL, IELTS, TEPS, OPIC or other tests. Thirty-one of them (24%) did not respond, which means they did not have any experiences of preparing for and/or taking tests, except the G-TELP (see Figure 4 for detailed information).

Figure 4

Proficiency Test-Taking Experience

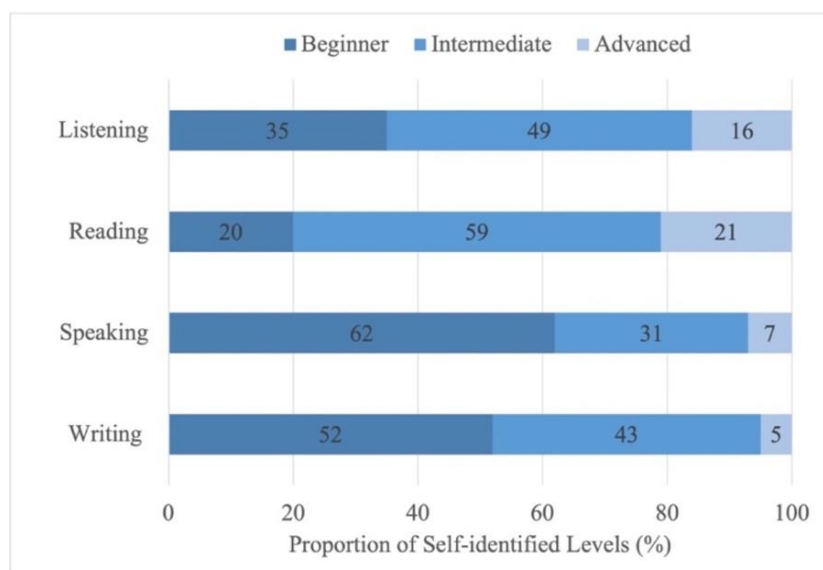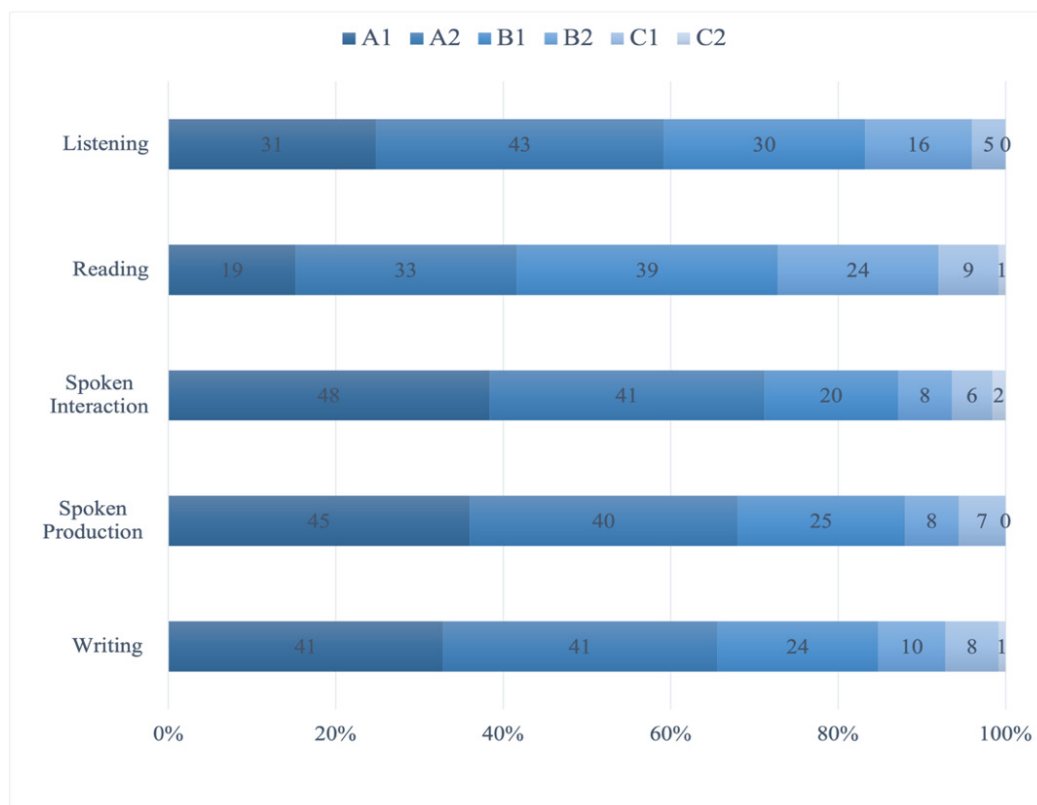*Note*. Respondents were able to choose more than one answer; thus, the total exceeds 125.

For the respondent's self-reported proficiency levels in four subskills and CEFR self-assessment responses, their responses were first tabulated into a table (see Table 4). A series of chi-square tests of independence were conducted to check if their self-reported levels appropriately corresponded to the CEFR levels. The test results showed that there was a significant association between their self-reported level and CEFR self-assessment measures in

all subskills, listening with $\chi^2 = 45.36$, $p < .001$, reading with $\chi^2 = 41.35$, $p < .001$, speaking

(interaction) with $\chi^2 = 58.53$, $p < .001$, speaking (presentation) with $\chi^2 = 53.10$, $p < .001$, and

writing with $\chi^2 = 69.81$, $p < .001$. Based on these results, while the majority self-identified as

intermediate-level English learners in listening (49%) and reading (59%) skills, there were more

self-reported beginners in speaking (62%) and writing (52%) skills (see Figure 5 and Figure 6 for

complete results).

Table 4

Self-Reported Levels and CEFR Self-Assessment Ratings (N =125)

Subskills	Proficiency Levels	A1	A2	B1	B2	C1	C2
Listening	Beginner	23	18	1	1	1	0
	Intermediate	8	18	23	10	2	0
	Advanced	0	7	6	5	2	0
Reading	Beginner	9	12	3	0	1	0
	Intermediate	8	15	34	13	4	0
	Advanced	3	6	2	11	4	0
Speaking (Interaction)	Beginner	45	25	7	0	1	0
	Intermediate	2	15	11	7	3	1
	Advanced	1	1	2	1	2	1
Speaking (Presentation)	Beginner	45	25	9	1	1	0
	Intermediate	3	14	13	6	3	0
	Advanced	0	1	3	1	3	0
Writing	Beginner	35	26	1	2	1	0
	Intermediate	6	15	21	7	4	1
	Advanced	0	0	2	1	3	0

Figure 5

Self-Identified Levels in Subskills

Figure 6

CEFR Self-Assessment Ratings

For the interview participants, Table 5 summarizes their background information. All names are pseudonyms and there were eight female and two male participants. Their ages ranged from 20 to 35. There were five job seekers with a bachelor's degree, three current college students, one office worker and one stay-at-home wife. They all had taken the G-TELP at least once, with the average score ranging from 40 to 92. Interviewees #3, #4, and #10 took the test, but the scores were not available at the time of data collection. Most of them reported that they do not study or use English on a regular basis except for the G-TELP preparation. However, there was some notable background information about two participants. Interviewee #7, Suji, lived in Canada when she was young, went to a foreign language high school, and participated in a study abroad program in college. Interviewee #9, Yoon, had lived abroad to accompany her husband and had past experiences of using English at workplaces on a daily basis.

Table 5

Test Taker Interview Participant Background Information (N = 10)

#	Pseudo nyms	Gender	Age	Current Job	G-TELP (score)	Reasons for taking the G-TELP & Major preparation methods/strategies
1	Mina	F	33	College student	Once (70)	- Certificate exam requirement - Online and offline courses
2	Yumi	F	25	Job seeker	Once (59)	- Public official exam requirement - Online course & focusing on memorizing grammar points
3	Hyun	M	27	Job seeker	Once (NA)	- For job application & exams - Used a practice test book
4	Minji	F	27	Office worker	Once (NA)	- Certificate exam requirement - Online course & practice test book

5	Eunji	F	25	Job seeker	Once (40)	- Public official exam requirement - Online course & focusing on memorizing grammar points
6	Nahee	F	27	Job seeker	Twice (68)	- Certificate exam requirement - Practice test & study listening
7	Sujin	F	24	College student	Once (92)	- CPA certificate exam requirement - Practiced sample test questions
8	Hanna	F	20	College student	Once (52)	- For job application in the future - Read other test takers' reviews online and practiced item types
9	Yoon	F	35	Stay-at-home wife	Once (80)	- Certificate exam requirement - Used a practice test book
10	Joon	M	32	Job seeker	Once (NA)	- Public official exam requirement - Online course & practice test book

Note. NA means the participant's score was not available at the time of data collection.

Test Preparation Course Instructor Background

As summarized in Table 6, the instructors who participated in the survey include three male and four female instructors with the mean age of 34.71. While four of them hold a bachelor's degree, three received a master's degree. Instructor #1 received her MA degree in an English-speaking country and others completed their education in Korea. Their majors include English language and literature (Instructor #3), German language and literature (Instructor #4 & 5), education (Instructors # 1 & 2), international relations (Instructor #6), and social welfare (Instructor #6). As for their teaching experiences, they were either early career teachers or mid-career teachers with around 10 years of teaching experience. Two of them were private tutors who help a group of English learners for test preparation. The other five reported that they are

teaching at *hagwons* (i.e., language institutions or test preparation cram schools in Korea). All of them reported that they are teaching courses for college students who are job seekers and those who prepare for public official exams. They started teaching a G-TELP preparation course relatively recently, with much shorter experience than their general teaching experience. Among the seven participants, Instructors #5 (Daeho) and #6 (Hyunwoo) volunteered for individual interviews.

Table 6

Instructor Background Information (N = 7)

#	Gender	Age	Teaching Experience	G-TELP Prep Course Teaching	Current Workplace
1	F	37	More than 10 years	3-5 years	Online/offline *Hagwon*
2	M	35	1-2 years	1-2 years	Private Tutoring
3	F	28	Less than 1 year	Less than 1 year	Offline *Hagwon*
4	F	29	1-2 years	Less than 1 year	Offline *Hagwon*
5	M	37	9-10 years	3-5 years	Online/offline *Hagwon*
6	M	45	More than 10 years	3-5 years	CEO/Instructor of *Hagwon*
7	F	32	9-10 years	Less than 1 year	Private Tutoring

Test Score User Background

Table 7 displays background information of test score users who responded to the survey. There were five male and one female respondents, with the mean age of 34. Their most common responsibilities include recruiting students or employees, reviewing applications, answering questions from applicants, and placing admitted/hired people. However, none of the respondents were responsible for building people's English language skills or giving input on policies for

language requirements. Score user #3 (Changmin) was the only professional who works for a

public institution, and he reported that his responsibilities are related to promotion and overseas

training of public officials. All the respondents who work at private companies reported that their

institutions utilize the TOEIC and OPIC scores. Score user #3 provided further explanation in the

survey that English proficiency test scores are required for public officials who apply for

overseas training. He was the only participant for an individual interview.

Table 7

Score User Background Information (N = 6)

#	Gender	Age	Workplace	Position	Working Experience	Use of the G-TELP Scores
1	M	36	Private Company	Manager	More than 5 years	No
2	M	50	Private Company	Vice President	More than 5 years	No
3	M	33	Public Institution	Public Official Manager	4 years	Yes
4	M	29	Private Company	HR Assistant Manager	2 years	No
5	F	28	Private Company	HR Regular Staff	4 years	No
6	M	30	Private Company	HR Senior Professional	Less than 1 year	No

Findings

In the findings section, the responses from surveys are presented using descriptive

statistics, such as mean, percentage, standard deviation, and maximum and minimum figures.

The results are discussed in connection with the interview data using quotes from the participants.
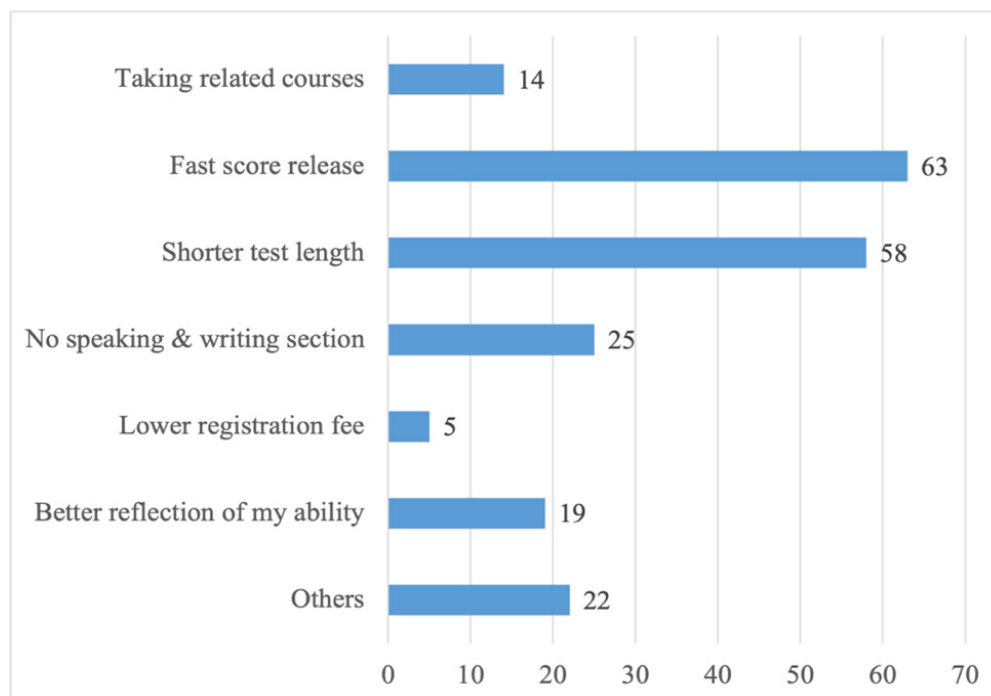
RQ1. What strategies do test takers and instructors perceive to be effective for preparing

for the G-TELP Level 2 test?

Test takers

Figure 7 shows why test takers chose to take the G-TELP. This is important because choosing the G-TELP itself is already a strategic decision to pass the requirement for public official exams or certification processes, which require a large amount of time and effort.

Figure 7

Reasons for Taking the G-TELP (N = 125)

Note. Respondents were able to choose more than one answer; thus, the total exceeds 125.

The two most selected reasons include fast score release and shorter test length. As reported above, most of the test takers were job seekers (40%) and students (26%) who will be looking for jobs during study or after graduation, so they are not only studying for the English proficiency test, but also preparing for other requirements for jobs applications, public official exams, or certificate exams. This necessitates a faster score release so that they can strategically plan their job seeking process. A test taker, Yoon, said, "I had to leave Korea to accompany my family, so I wanted to have results as soon as possible. Taking the G-TELP was a better choice

since it released the score so fast." Also, the shorter test length was associated with less time and

energy spent to prepare for the test. This tendency was observed in answers to the question about

the length of preparation period, as displayed in Table 8. Most of the respondents reported that

they spent less than 1 month (36.80%) or 1-2 months (35.20%). While there were respondents

who spent a longer time, such as 3-4 months, 5-6 months, or more than 6 months, 10.40% of the

participants said that they did not prepare for the test.

Table 8

Length of Preparation Period (N = 125)

Period of Time	N	%
No preparation	13	10.40
Less than 1 month	46	36.80
1-2 months	44	35.20
3-4 months	13	10.40
5-6 months	6	4.80
More than 6 months	3	2.40

The instructor Hyunwoo also said that "I think the G-TELP is a decent proficiency test

because it does not place a lot of burden on test takers. There are 80 questions in total. The test is

90 minutes long, which is 30 minutes less than the TOEIC. I believe this makes a huge

difference for test takers." The test taker, Minji, who is currently working for a company, had a

positive attitude toward the G-TELP. This was because preparing for the TOEIC took too much

energy and time as she experienced before. She agreed that choosing the G-TELP was a more

efficient strategy. As she was busy with work, preparing for the G-TELP with a smaller number

of questions required less of her time and effort.

The preparation activities that were perceived to be important were also related to the need to save time to study for other subjects or work while preparing for the test. For this question, test takers rated how important the given activities were in preparing for the G-TELP according to a 4-point scale as indicated in Table 9.

Table 9

Test Takers' Perceived Importance of Preparation Activities (N =125)

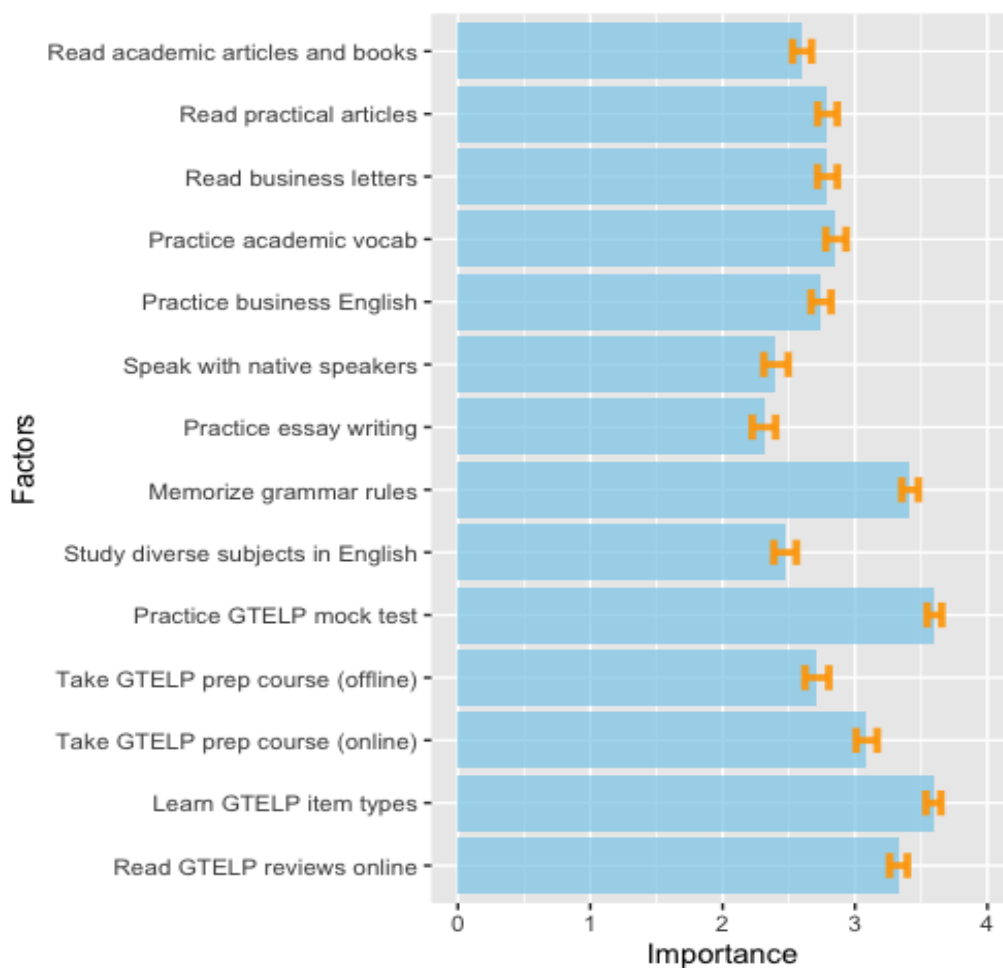| Please indicate the importance of the following activities.	Mean	SD	Min.	Max.
Read academic articles and books	2.60	0.80	1	4
Read practical articles	2.79	0.83	1	4
Read business letters	2.79	0.84	1	4
Practice academic vocab	2.86	0.86	1	4
Practice business English	2.74	0.84	1	4
Speak with native speakers	2.40	1.03	1	4
Practice essay writing	2.31	0.99	1	4
Memorize grammar rules	3.42	0.70	1	4
Study diverse subjects in English	2.47	0.96	1	4
Practice G-TELP mock test	3.60	0.62	1	4
Take G-TELP prep course (offline)	2.71	1.01	1	4
Take G-TELP prep course (online)	3.09	0.89	1	4
Learn G-TELP item types	3.59	0.64	2	4
Read G-TELP reviews online	3.33	0.75	1	4

Note. 1 =Not important at all, 2 = Not important, 3 = Important, 4 = Very important

Clearly, a mean of 3.0 or higher indicates agreement, and a mean of 2.0 or lower indicates disagreement. There is no definitive standard on how to determine the midpoint for data of this kind (McIver & Carmines, 1981) as pointed out by Malone and Montee (2014). Therefore, this study follows the previous study and determined that agreement of 2.75 or higher

indicates that most respondents think the activity is important; below that number means that respondents generally do not think it is important. Figure 8 shows the frequency of general preparation strategies that test takers reported to have used with error bars.

Figure 8

Test Takers' Perceived Importance of Preparation Strategies (N =125)

The activities respondents considered most important—mean rating above 3—include learning G-TELP item types, practicing G-TELP mock tests, memorizing grammar rules, reading G-TELP reviews online, and taking G-TELP prep courses online. The high importance of learning what kind of item types the G-TELP includes was corroborated by interview data. All of the 10 test takers and two instructors who participated in interviews mentioned the importance of

understanding the item types. A test taker, Yumi, reported that "I think learning about item types was a great help. I originally thought the G-TELP has the same types of questions as the TOEIC, but it was not the case." Another test taker, Sujin, emphasized the importance of understanding item types by saying "the G-TELP definitely has some question types that are different from the Korean CSAT English section that most Korean college students are very familiar with. I don't think Korean people would be very familiar with the G-TELP items. I told my friend that they need to make sure to learn what kinds of question types they will see on the G-TELP." An instructor, Daeho, said, "Even the test takers with a higher English proficiency level, without an understanding of G-TELP item types, would miss a number of questions." This is closely related to the second useful activity, practicing with G-TELP mock tests, as all of the participants mentioned the importance of mock tests during the interview. One test taker, Nahee, said that the most helpful activity was to practice with mock tests within the enforced time limits. Memorizing grammar rules was considered important as six test takers reported that they memorized only the grammar rules that are essential for the G-TELP Level 2. They took online or offline preparation courses, and the instructors told them to do the memorization activities. This was probably recommended because it is considered an effective and efficient strategy to increase the G-TELP score within a short time. As the instructor Hyunwoo said, "If you memorize grammar rules, practice with mock tests, and review them, then your score will definitely be improved." The interview data do not provide much evidence for the importance of reading the information about the G-TELP online or taking a course online. However, a test taker Hanna said that "there is not much information about the G-TELP compared to other tests, so I searched for some reviews on NAVER (a Korean search portal) written by experienced test takers." Test instructor Daeho said that he always gets invited to do an online lecture on the test
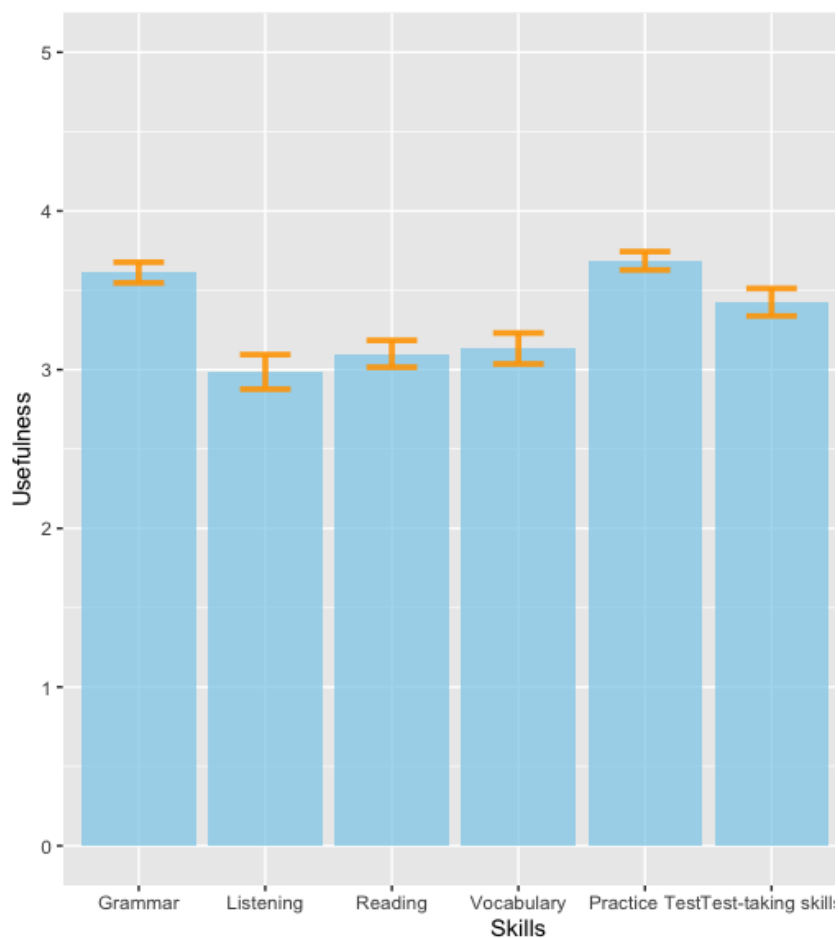
review after he himself takes the test. As there are several freely available online videos like his,

test takers who cannot take offline courses would benefit from those online lectures for the G-

TELP preparation.

The next question was asked to examine the perceived usefulness of skills they practiced

in the G-TELP preparation courses. Respondents were first asked if they had taken any G-TELP

preparation courses. They were able to choose all that apply. Twenty-seven test takers (21.60%)

reported that they took an offline G-TELP course, and 79 of them (63.2%) said that they took an

online course. As more than half of test takers took an online course, this result once again

corroborates the tendency for test takers to consider taking an online course as an important

activity for the G-TELP preparation. Thirty-one (24.80%) took a general course for English

proficiency tests, but 13 (10.40%) did not take any courses. Thus, the result presented here is

based on the responses from 113 test takers who took an offline or online G-TELP course and/or

an English course. This set of questions asked the respondents to indicate their perceived

usefulness of the given skills on the scale of 4.

The practiced skill that is considered most useful among test takers is grammar (see

Figure 9 and Table 10). The usefulness of practicing grammar skills is evident in interview data

as well. All of the test takers and instructors who were interviewed mentioned that practicing for

the grammar section was important since the grammar section is the easiest and most accessible

section for improving one's score. For example, the test taker, Yumi, said that she found that

memorizing grammar points was very useful and helpful in achieving the target score of 32

points.

Figure 9

Test Takers' Perceived Usefulness of Practice (N =125)

Table 10

Test Takers' Perceived Usefulness of Practice (N =125)

Skills	Mean	SD	Min.	Max.	Did Not Practice
Grammar	3.61	0.60	2	4	5
Listening	2.99	0.92	1	4	17
Reading	3.10	0.76	1	4	11
Vocabulary	3.13	0.84	1	4	16
Practice tests	3.69	0.54	2	4	5
Test-taking skills	3.43	0.78	1	4	11

Notes. 1 =Not useful, 2 = Somewhat useful, 3 = Useful, 4 = Very useful

Instructors also considered practicing grammar to be quite useful (see Table 11). The

instructors Hyunwoo and Daeho reported that for those who need to receive more than 32 points

on the G-TELP to qualify for the army public official exam, the most useful skill is grammar.

Hyunwoo said, "I really hate to say this, but I cannot help but primarily focus on the grammar

section. I teach them only the grammar points and patterns that are covered in the G-TELP Level

2, and teach my students some tricks and test-taking strategies."

Table 11

Instructors' Perceived Usefulness of Practice (N = 7)

Skills	Mean	SD	Min.	Max.
Grammar	3.14	0.69	2	4
Listening	2.43	0.98	1	4
Reading	3.14	0.69	2	4
Vocabulary	3.00	0.82	2	4
Practice tests	3.43	0.79	2	4
Test-taking skills	3.57	0.53	3	4

Notes. 1 =Not useful, 2 = Somewhat useful, 3 = Useful, 4 = Very useful

However, the listening skill was perceived as the least useful by both test takers and

instructors. The reasons can be gathered from the interview data. One of the salient themes for

test takers' perception of the G-TELP listening was the excessive difficulty of the section. That is

why a certain group of test takers choose the strategy of giving up on the listening section. All of

the interviewees, including test takers and instructors, expressed concerns that the listening

section was too difficult. The test taker, Joon, expressed frustration with his listening ability and

even asked for tips during the interview with the researcher. The instructor Hyunwoo commented

that "I enjoy the G-TELP listening section. However, the passages are too long and extremely

challenging to beginner-level learners. If the questions were easier, maybe people would try the

first couple of questions that are less challenging, but the listening passages are too long, so they

would just give up." On this issue, the other instructor, Daeho, gladly shared his experiences of

taking the G-TELP recently, saying that "nobody except me was taking notes and marking answers during the listening section. I feel that this is unfortunate because test takers learn that the listening section is very challenging, and they already decide to give up on the section in the preparation stage. The listening questions are not very accessible for beginner-level learners." The two instructors' opinions and observations were in line with results from the test taker survey. Among the participants, 35% of them self-evaluated their listening ability as beginner-level while only 20% of them said their reading is at beginner-level. To recap the common strategies around the listening section, it is too difficult, and it does not have any items that are accessible and manageable for beginner-level learners. Those learners decide to give up on the listening section based on the advice from experienced test takers, friends, or instructors. This "giving-up" strategy also takes us back to the discussion of test takers' limited time and energy. Listening ability might be possible to improve at least a little in a shorter period of time with intensive practice. However, the test takers need scores to satisfy the requirements for the job applications or certificate exams as soon as possible with minimum effort. This justifies their decision to give up on the listening section and focus on the grammar section since it covers a certain scope of grammar rules that test takers can easily study before the test.

The survey data also show that taking a practice test is perceived as more useful by both test takers and instructors. This is related to the importance of practicing with mock tests, which has already been discussed above. It seems that test takers and instructors think it is important because it is useful for taking the test. In addition, test-taking skills were considered useful by test takers and instructors. Several test-taking skills were mentioned by test takers during the interviews. The test takers, Mina and Yumi, reported that they gave up on the listening section and use the time for solving the grammar and reading sections. This was already mentioned

above as a popular strategy recommended by instructors and utilized by test takers with lower

listening proficiency. However, test takers with higher listening proficiency, such as Sujin and

Hanna, reported that on the listening section, they used note-taking skills selectively because

sometimes understanding the content while listening to the passages facilitated the problem-

solving process. This was possible because, as Sujin pointed out, the questions were in the order

of information presented in the listening passage, which was quite straightforward to test takers
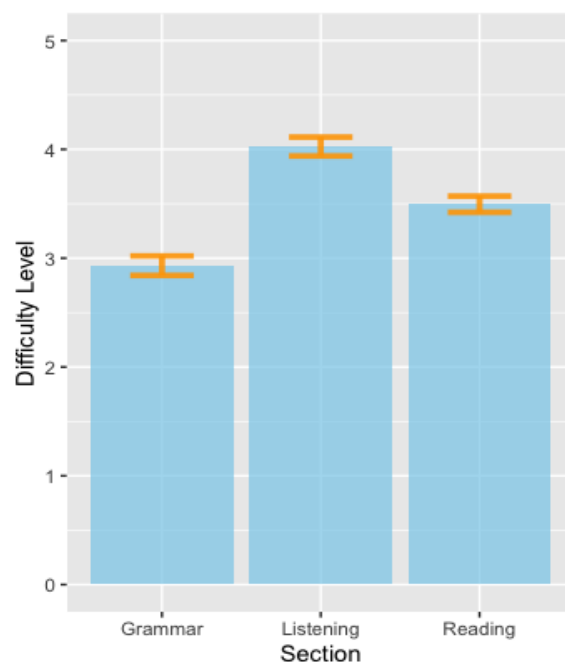
with higher proficiency level.

Summary: Research Question 1

For test takers, understanding question types, using practice tests, or utilizing relevant

test-taking strategies were most effective for achieving the target scores. However, the strategies

perceived as effective by test takers might be different for different proficiency levels. For

example, test takers with a lower listening ability just give up on the listening section and focus

on grammar section, which is easier and more accessible. This strategy is quite prevalent among

the test takers who only need 32 points or above to qualify for their exam requirements. On the

other hand, more proficient test takers tend to develop their own test-taking strategies after they

practice with mock tests.

RQ2. What are stakeholders' beliefs about what each section of the G-TELP Level 2

measures?

Difficulty Level of Different Sections of the G-TELP

Before discussing what test takers believe each section measures, the perceived difficulty

of each section is discussed. The results are used to discuss how test takers' perceptions of the

difficulty level have influenced their beliefs about each section. Figure 10 and Table 12 display

the responses from the test takers.

Figure 10

Test Takers' Perceived Difficulty Level by Sections (N =125)

Table 12

Test Takers' Perceived Difficulty Level by Sections (N =125)

Skills	Mean	SD	Min.	Max.
Grammar	2.93	0.97	1	5
Listening	4.03	0.94	2	5
Reading & Vocab	3.50	0.81	2	5

Note. 1 = Very easy, 2 = Easy, 3 = Moderate, 4 = Difficult, 5 = Very difficult

To the respondents, the grammar section was perceived as the easiest among the three sections, while the listening section was perceived to be the most difficult. This seems to be closely related to the strategies test takers utilized: giving up on the listening section but focusing primarily on the grammar section. This tendency was evident in the interview data. All 10 of the test takers who were interviewed reported that they learned the information that the listening

section is difficult after reading reviews or watching online lectures. Then, when they took the test themselves, the listening section posed the greatest challenge for them. By contrast, the grammar section was perceived as easy. Either from fellow test takers or from the instructors, test takers learned the information that the grammar section is more manageable since it only covers a certain range of grammar rules. Then, they used the strategy of studying those rules through utilizing resources such as online lectures or practice tests. The test taker, Yoon, said:

> The G-TELP only covers a set of certain grammar rules. I did not study as much as I used to do when preparing for the TOEIC. I only studied using a practice book. I feel like if you can read a-b-c, then you can study the grammar rules and take the test. The grammar section was that easy.

Table 13 presents the test takers' perceptions about the three sections organized by self-reported proficiency levels of the participants: Beginner, Intermediate, and Advanced in each of the four skills. For the grammar section, because respondents were not asked to indicate their levels in grammar skill, writing skill was used as a proxy. For listening and reading sections, respondents' self-reported levels in the listening and reading skills were used for analysis. Eight respondents did not respond to this set of questions, so only valid responses from 117 test takers were analyzed.

Initial inspection of data indicated that the data deviated from the assumption of normal distribution and the group sample sizes were different and small. A series of non-parametric Kruskal-Wallis H tests were conducted to compare how different proficiency groups rated the difficulty of each section differently. A Kruskal-Wallis H test indicated no significant differences in difficulty rating for the grammar section among different proficiency levels (H (2) = 0.54, p = 0.764). However, Kruskal-Wallis H tests showed that there was a statistically significant

difference in perceived difficulty levels in the listening and reading sections. A significant difference was found for the listening section with H (2) = 29.396, $p < 0.001$, and the reading section with H (2) = 20.687, $p < 0.001$. The Dunn's post-hoc tests indicated significant differences in perceived difficulty of the listening and reading sections between beginner and advanced levels as well as between beginner and intermediate levels. No significant results were found between intermediate and advanced levels. A series of statistical test results showed that those who self-reported as beginners in English listening ability found the listening section significantly more difficult compared to those who self-identified as intermediate or advanced in the skill. This was also true for the reading section. However, the grammar section was generally perceived as easy or moderate in all proficiency groups.

Table 13

Test Takers' Perceived Difficulty Level by Section and Self-Reported Proficiency Level (N = 117)

Section	Level	Mean	SD	Min	Max
Grammar	Beginner (41)	2.90	0.96	1	5
	Intermediate (50)	2.93	0.95	1	5
	Advanced (26)	3.00	1.08	1	5
Listening	Beginner (41)	4.59	0.67	3	5
	Intermediate (50)	3.75	0.90	2	5
	Advanced (26)	3.65	1.04	2	5
Reading/Vocab	Beginner (41)	3.73	0.78	2	5
	Intermediate (50)	3.41	0.80	2	5
	Advanced (26)	3.25	0.79	2	5

Note. 1 = Very easy, 2 = Easy, 3 = Moderate, 4 = Difficult, 5 = Very difficult

This tendency was also found in the interview data. For example, the test taker Joon, expressed a high level of frustration with his low listening ability, commenting that the listening section is overwhelmingly difficult for him. On the contrary, the test taker Sujin, who achieved

92% in the listening section, commented, "The listening section must be very challenging for

those who are struggling, but I don't think the listening section should be troublesome for

proficient test takers."

Test takers were also asked to rate the extent of agreement with the statements about time

pressure of each section (see Table 14). The mean ratings for the grammar and listening and

sections were lower than 2.75, indicating that on average respondents did not agree that they

have enough time for the grammar and listening sections. This might be because most test takers

only focus on the grammar section to achieve a higher score, which requires a longer time and

more effort to ensure that their answers are right.

Table 14

Test Takers' Perceived Time Pressure by Section

Statement	Mean	SD	Min.	Max.
I had enough time to answer the questions on the grammar section.	2.71	0.94	1	4
I had enough time to answer the questions on the listening section.	2.34	0.90	1	4
I had enough time to answer the questions on the reading & vocabulary section.	3.13	0.81	1	4

Note. 1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree

On a slightly different note, the listening section is perceived as too difficult, and test

takers believe that the listening audio speed is quite fast. That might have contributed to the

respondents' perception that they do not have enough time to answer the questions. However,

with the mean rating of 3.13, test takers agree that they have enough time for the reading and

vocabulary section. This might be because test takers are allowed to work on any section during

the entire test time. If a test taker decides to give up on the listening section, the time can be

spent on the reading/vocab section. Thus, some test takers perceive that this is an effective

strategy when they do not have a higher level of listening ability yet need to achieve a certain

score within a short time. With this test administration policy, it seems that the reading section

does not measure the test takers' ability to comprehend the given texts within a highly controlled

time limit.

As for the instructors' general perceptions about each section of the test, they were asked

to rank the importance of test preparation for each section in the survey. Five out of seven

instructors ranked the grammar section as the most important section to focus on in preparation

for the following reasons:

- Even first-time test takers can improve scores within a short time.

- Anyone can get higher scores because the item types have fixed patterns.

- As long as test takers are prepared, they can achieve higher scores. Definitely worth
  focusing on the grammar section.

This perception was also evident in the interviews with the instructors. Hyunwoo said "Except

one or two questions, the grammar section is pretty easy and accessible to all proficiency levels."

Daeho also commented that the section is not supposed to be easy, but test takers can use the

strategy of being familiarized with item types, studying question patterns, and memorizing

grammar rules. With this strategy, the grammar questions are relatively easy to solve.

Based on the discussion so far, it seems that some G-TELP test takers strategically focus

on preparing for the easiest section, the grammar section, instead of making more effort to

improve English abilities to achieve higher scores in the reading and listening sections. This

strategy is recommended and encouraged by the instructors as well. Test takers choose this

strategy especially if they need to achieve the target minimum score within a short time to

qualify for certification processes or public official exams.

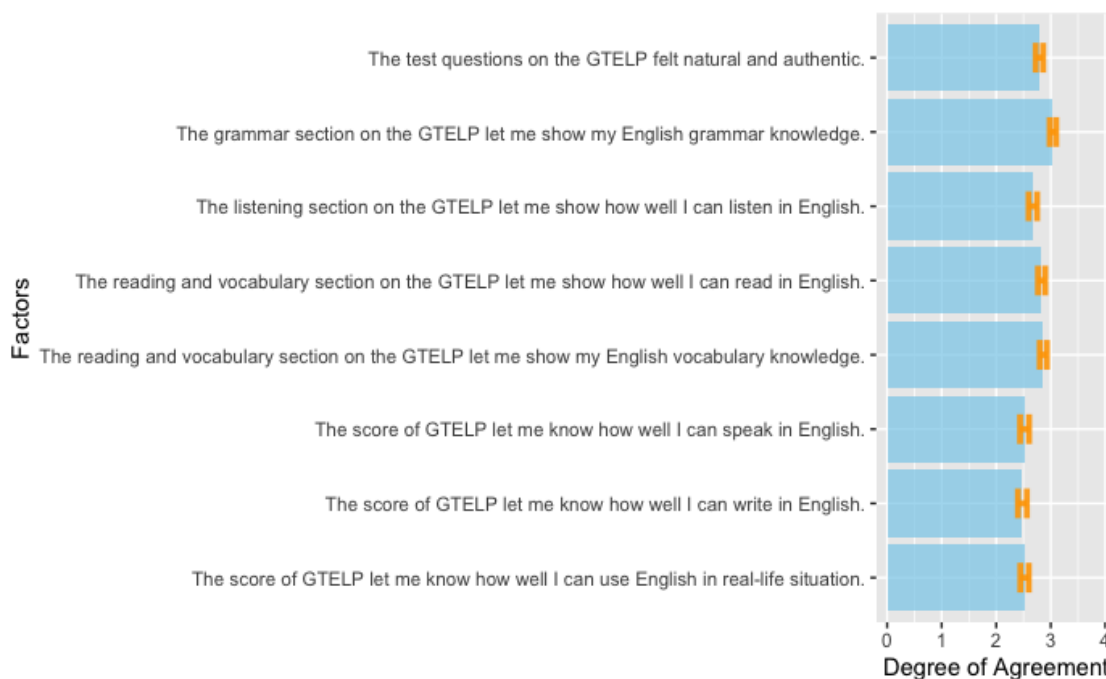**Stakeholders' Perceptions of the G-TELP Sections**

The test takers and instructors in the current study were asked about their beliefs about the G-TELP sections. For test takers, they were asked to indicate their general beliefs and what skills they used for each section of the test. As for test preparation course instructors, the survey asked them about what skills they have their students practice in the test preparation courses.

Test Taker Beliefs About the G-TELP

For this part of the survey, test takers rated the extent of their agreement with the given statements (see Figure 11 and Table 15) about the G-TELP, according to the 4-point scale. As mentioned before, 2.75 was set for the threshold of agreement.

Figure 11

Test Takers' Perceived Efficacy of the G-TELP

As Figure 11 shows, it is noticeable that test takers agree that the grammar section showed their grammar knowledge. The mean rating is 3.04, which indicates a large extent of agreement (see Table 15 for details). While test takers do not seem to believe that the listening

section showed how well they could listen in English (2.58), they believe that the reading and

vocabulary section showed how well they could perform in English reading comprehension and

reflected their vocabulary knowledge as the mean ratings were above 2.75. In terms of score

interpretation, test takers do not agree that the G-TELP score can show their speaking and

writing skills or their ability to use English in real-life communication.

Table 15

Test Takers' Perceived Efficacy of the G-TELP (N = 125)

Please indicate your agreement with the following statements.	Mean	SD	Min	Max
The test questions on the G-TELP felt natural (and authentic).	2.70	0.76	1	4
The grammar section on the G-TELP let me show my English grammar knowledge.	3.04	0.65	1	4
The listening section on the G-TELP let me show how well I can listen in English.	2.58	0.86	1	4
The reading and vocabulary section on the G-TELP let me show how well I can read in English.	2.79	0.73	1	4
The reading and vocabulary section on the G-TELP let me show my English vocabulary knowledge.	2.80	0.70	1	4
The score of the G-TELP let me know how well I can speak in English.	2.40	0.90	1	4
The score of the G-TELP let me know how well I can write in English.	2.36	0.92	1	4
The score of the G-TELP let me know how well I can use English in real-life situations.	2.45	0.89	1	4

Note. 1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree

Instructor Beliefs About the G-TELP

As Table 16 indicates, the G-TELP preparation course instructors somewhat disagree

with the given statements about the G-TELP. Two statements have a mean rating above 2.75 and

three have somewhere slightly below 2.75, while the others are relatively low. Notably,

instructors believe that they understand what the G-TELP scores mean. In addition, they believe

that test preparation helps students become ready for life at a university in an English-speaking

country.

Table 16

Instructors' General Beliefs About the G-TELP (N = 7)

Please indicate your agreement with the following statements.	Mean	SD	Min	Max
I understand what the G-TELP scores mean.	3.14	0.69	2	4
ITSC and G-TELP Korea adequately disseminate information about changes to the G-TELP.	2.29	0.76	1	3
ITSC and G-TELP Korea adequately disseminate information about the meaning of G-TELP scores.	2.57	0.53	2	3
Preparing to take the G-TELP prepares students for life at a university in an English-speaking country.	2.86	1.07	1	4
Preparing to take the G-TELP prepares students for professional life at a workplace or government agencies.	2.71	1.11	1	4
Users of G-TELP scores (administrators or employers—including public agencies, private sectors, and government institutions) understand how to use G-TELP scores.	2.71	0.49	2	3
Users of G-TELP scores look at subscores as well as total scores.	2.57	0.53	2	3
The G-TELP is an accurate predictor of how well a non-native English speaker will perform in an English-speaking context.	2.71	0.49	2	3

Note. 1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree

However, instructors did not have consensus on whether preparing for the G-TELP is

helpful for professional life. The mean rating was 2.71—five instructors agreed with them but

two disagreed. Likewise, instructors did not demonstrate strong consensus on whether the G-

TELP score users understand how to use the scores, again with the mean rating of 2.71. In

addition, they tended to disagree that score users (e.g., administrators or employers) look at the

subscores in addition to total scores. This seems to indicate that instructors have doubts about

test score users' assessment literacy for using the test scores. With rather lower mean ratings, instructors tended to disagree that ITSC and G-TELP Korea adequately disseminate information about changes to the G-TELP test and the score interpretation. Finally, they do not seem to strongly agree that the G-TELP is an accurate predictor of how a non-native English speaker will perform in an English-speaking context.

Based on the results discussed above, it seems that instructors generally believe that they have a decent level of understanding about the G-TELP and the meaning of the G-TELP scores while they believe that there is a lack of understanding about the test among test score users and that insufficient information is provided by ITSC and G-TELP Korea.

In addition to the general beliefs about the G-TELP test, instructors were also asked about what the G-TELP measures for each section. Table 17 displays the instructors' responses to this part of the survey. Despite the low mean rating for the G-TELP's ability to predict test takers' overall future English performance (the mean rating of 2.71, see Table 16), mean ratings for the G-TELP as a measure of skill-specific ability were above 3.00 for the listening section and reading/vocab section (see Table 17). This indicates that, on average, the instructors agreed that the listening and reading/vocab sections of the G-TELP allow test takers to show how well they can use English for the target skills or domains. For example, instructors strongly believe that, with the highest mean rating of 3.43, the listening section allows students to show how well they can listen in English. However, this finding is interesting because test takers do not strongly agree that the listening section allows them to demonstrate their listening skills. This contradiction may be caused by the excessive difficulty level of the listening section. That is, while instructors perceive the listening section to have a higher level of authenticity and item discrimination, test takers perceive it as too challenging and not accessible with their current

listening comprehension skills. As for the reading/vocab section, there was no contradiction between instructors and test takers. Both groups agree that the section allows test takers to show their reading ability and vocabulary knowledge. As for the grammar section, instructors also agree that the grammar section allows test takers to show their grammar knowledge. This belief is similar to that of test takers. Finally, instructors do not agree that the G-TELP score overall can show test takers' speaking and writing skills or their ability to use English in real-life communication.

Table 17

Instructors' Beliefs About What the G-TELP Measures (N = 7)

Please indicate your agreement with the following statements.	Mean	SD	Min	Max
The grammar section on the G-TELP allows students to show their English grammar knowledge.	2.86	1.07	1	4
The listening section on the G-TELP allows students to show how well they can listen in English.	3.43	0.53	3	4
The reading & vocabulary section on the G-TELP allows students to show how well they can read in English.	3.14	0.90	2	4
The reading & vocabulary section on the G-TELP allows students to show their English vocabulary knowledge.	3.14	0.69	2	4
The G-TELP overall allows students to show how well they can speak in English.	2.00	0.82	1	3
The G-TELP overall allows students to show how well they can write in English.	2.29	1.11	1	4
The G-TELP overall allows students to show how well they can use English in real-life situations.	2.43	0.79	1	3

Note. 1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree

Stakeholder Beliefs about Specific Skills Measured by Each Section

Test takers and instructors were also asked about specific skills in the surveys. Test takers were asked to rate the frequency of using the given skills in each section of the G-TELP
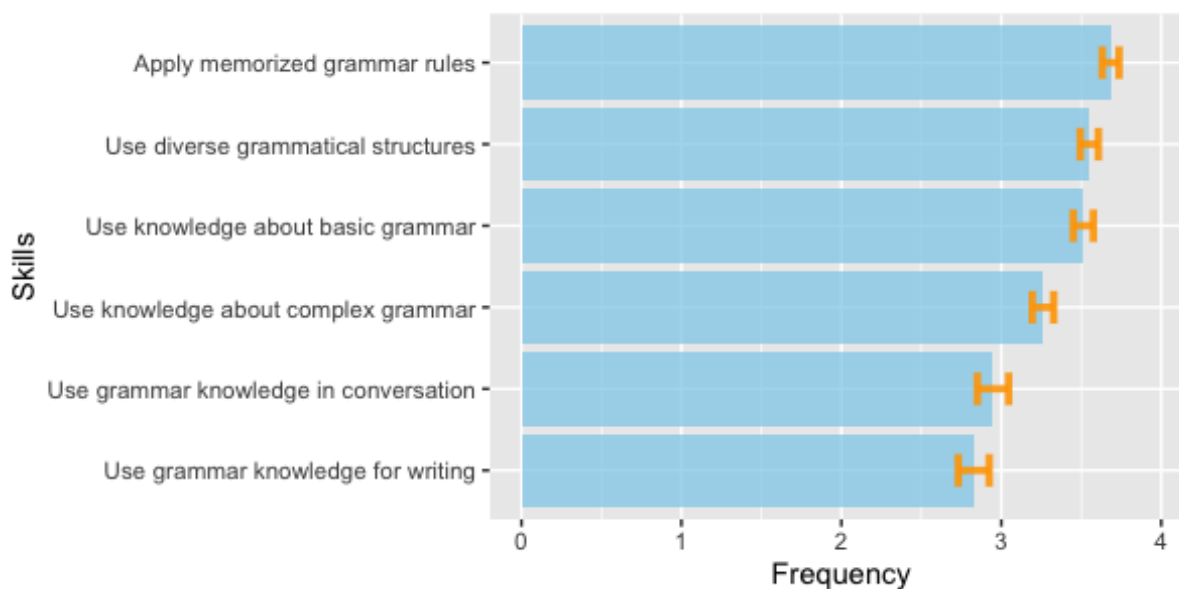
according to a 4-point scale (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often). Instructors were

asked about the same or similar set of skills with some adaptations to fit into teaching situations.

In the survey, the instructors were asked to indicate how often the skills were included in their

preparation courses. The results are presented below.

Grammar Section

As shown in Figure 12 and Table 18, test takers reported that the most frequently used

skill for the grammar section was *Apply grammatical rules I have memorized*. Once again, this

corroborates their perception about the effectiveness of the preparation strategy they have used

for the grammar section.

Figure 12

Frequency of Skills Utilized for the Grammar Section – Test Takers

Test takers and instructors all perceived studying and memorizing grammar rules as the

most effective strategy for this section. Consequently, test takers reported having frequently

applied the memorized rules to solve the questions. This was followed by other skills, including

using a range of diverse grammatical structures and utilizing knowledge about basic/somewhat

complex grammar rules.

Table 18

Frequency of Skills Utilized for the Grammar Section – Test Takers

Skills	Mean	SD	Min.	Max.	Don't know
Apply grammatical rules I have memorized	3.68	0.57	2	4	8
Use a range of diverse grammatical structures	3.55	0.62	2	4	5
Use knowledge about basic grammar rules (e.g., present and past verb forms, conjunctions, comparatives, superlatives)	3.51	0.69	1	4	6
Use knowledge about somewhat complex grammar rules (e.g., conditionals, auxiliary verbs, to-infinitives, gerunds, relative pronouns, etc.)	3.26	0.73	1	4	6
Use grammatical knowledge in conversational English	2.95	1.05	1	4	9
Use grammatical knowledge for English academic writing	2.83	1.04	1	4	9

Table 19 shows how frequently the instructors covered certain skills for their G-TELP

preparation courses. Instructors reported that, most frequently, they asked their students to work

on the practice questions and to memorize grammatical rules that appear on the test. During the

interview, the instructor Daeho shared how he uses practice questions in his course:

I can make the students to increase their points by 20 in the grammar

section. I have them take the practice tests with lots of grammar questions.

I review the questions that they got incorrect. Then, I repeat this process

for them and they learn, then the score definitely goes up.

The other instructor, Hyunwoo, explained more details about what grammatical rules are covered

and memorized by his students:

> I only cover the grammar rules that frequently appear on the test. For example,
>
> conjunctions or prepositions are relatively easy, but the G-TELP has more questions that
>
> assess knowledge about tense, including a distinction between present perfect and present
>
> perfect progressive or a distinction between future and future progressive. I focus a lot on
>
> tense so that students can understand the concepts. This will help them solve the grammar
>
> questions more easily.

These two instructors' teaching practices also support the rather frequent activity of explaining

grammar points from the survey.

Table 19

Frequency of Activities Covered in Class for the Grammar Section – Instructors

Activities	Mean	SD	Min.	Max.
Explain grammar points	3.29	0.76	2	4
Practice grammar questions	3.57	0.79	2	4
Practice formulas for structuring sentences	2.29	1.11	1	4
Practice a range of diverse grammatical structures	3.00	0.82	2	4
Memorize grammatical rules that frequently appear on the G-TELP	3.43	0.79	2	4

As discussed so far, instructors primarily focused on practicing and studying a certain set

of grammatical rules in their courses, and test takers also utilized the memorized grammar rules

during the actual test-taking process. However, test takers reported less frequent use of the

grammatical knowledge for conversational English or academic writing. Instructors also reported

the least frequent coverage of sentence structure using grammatical formulas. This indicates that

in preparation of the grammar section, grammar forms are heavily focused on while function and usage of grammar do not receive much attention. Regarding this tendency, the instructor Daeho expressed a concern:

> Even test takers without a proper level of English proficiency study the grammar rules and patterns, and they achieve a pretty decent score. I am not sure about how the test developers write items, but some important grammar rules are not even covered in the test. I sometimes wonder if we can really say that the test takers are capable of participating in meaningful communication only with that shallow grammatical knowledge they studied for this test.

Daeho's concern resonates with Hyunwoo's suggestions for possible change as well:

> The grammar section is not that challenging to the test takers. However, it might need some changes. While the TOEIC has shorter stimuli, like one sentence for one item, the G-TELP has more than two sentences. If test takers were to read the stimuli and completely understand to answer the questions, the G-TELP might be more difficult than the TOEIC. However, you can just read a few words right before and after the blank to find the correct answer. Currently, the grammar questions can be solved even without understanding the entire stimuli, which needs to be changed.

Daeho also agrees with Hyunwoo that the grammar section needs some improvement by including more essential grammar rules in the questions. Nevertheless, he was worried about the possible influences on potential test takers' choice of which English proficiency test they would take to satisfy requirements for job seeking. He said:

> I personally think the G-TELP is a pretty good proficiency test. However, as of now, it seems to only attract a certain group of people, such as those who are beginner-level
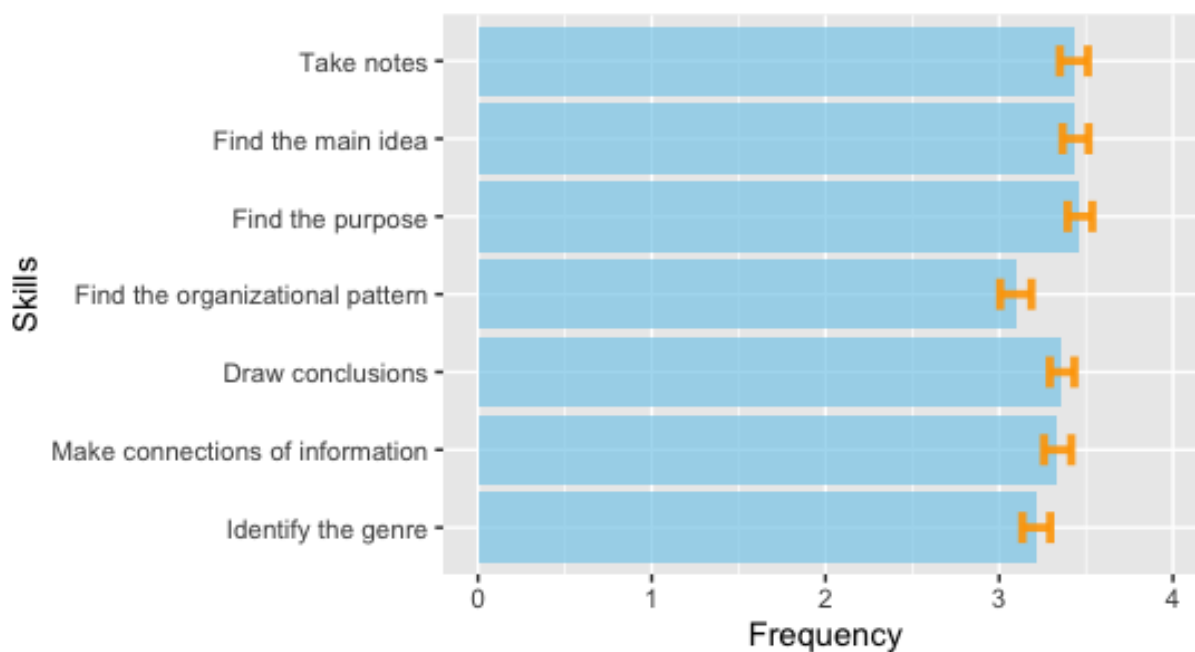
English learners and need to receive the average score of 32 points to satisfy certain requirements. They believe they can study grammar rules and patterns to achieve their goal within a short time. If changes were made to the G-TELP grammar section, it might lead to unexpected negative reactions. Test takers may think the G-TELP is too difficult and that they cannot achieve the target score. Then the G-TELP would lose potential test takers.

Based on the results and discussion above, it seems that stakeholders believe that the grammar section measures test takers' knowledge about a certain set of grammar rules (e.g., tenses or conjunctions), focusing on forms and patterns without much consideration of function or usage of grammar rules in conversational contexts or writing. This might cause the issue of construct underrepresentation, so further investigation is needed to determine whether the items represent the target construct of grammar knowledge that is relevant for the G-TELP Level 2 and whether the items and contents are ideal for the test taker population.

Listening Section

Figure 13 and Table 20 show the self-reported frequency of skills test takers used for the listening section. Interestingly, it seems that most of the skills have been utilized at a similar frequency. As Table 21 shows, instructors also used all the activities at a quite similar frequency. The skill of finding the speaker's purpose has the highest mean rating for both test takers (3.46) and instructors (3.71). This is probably because there are listening test items that explicitly ask test takers to identify the purpose of the speaker in listening passage. The skill of finding the main idea is also frequently used by the test takers. For each of the four passages in the listening section, test takers are asked to identify the main idea to respond to test items.

Figure 13

Test Takers' Frequency of Skills Utilized for the Listening Section

Table 20

Test Takers' Frequency of Skills Utilized for the Listening Section

Skills	Mean	SD	Min.	Max.	Don't Know
Take notes	3.43	0.86	1	4	11
Find the main idea	3.44	0.81	1	4	9
Find the speaker's purpose	3.46	0.75	1	4	11
Find the organizational pattern of a text (e.g., chronological order, process, etc.)	3.10	0.96	1	4	10
Draw conclusions based on what is implied	3.36	0.76	1	4	9
Make connections between pieces of information in a conversation or monologue	3.34	0.84	1	4	9
Identify the genre of the text (e.g., personal narrative, business negotiation, etc.)	3.21	0.87	1	4	8

Table 21

Instructors' Frequency of Activities in Class for the Listening Section

Activities	Mean	SD	Min.	Max.
Take notes	3.43	0.79	2	4
Find the main idea	3.43	0.79	2	4
Find the speaker's purpose	3.71	0.49	3	4
Find the organizational pattern of a text (e.g., chronological order, process, etc.)	3.14	0.90	2	4
Draw conclusions based on what is implied	3.14	0.90	2	4
Make connections between pieces of information in a conversation or monologue	3.00	0.82	2	4
Identify the genre of the text (e.g., personal narrative, business negotiation, etc.)	3.00	1.29	1	4

The least frequently used skill is to find the organizational pattern of a text with the mean rating of 3.10 (see Table 20). This might be because none of the test items directly asked test takers to identify the organizational patterns. During the interviews, test takers were found to be more concerned about understanding the speedy and rather authentic audio in terms of contents and meaning, rather than recognizing discourse-level characteristics of the listening passages.

The results should be interpreted with caution, though. That is because, as discussed earlier, some test takers tend to give up on the listening section and instructors usually focus on test-taking strategies or only grammar points that appear on the G-TELP. During the interview, it was discussed as a common test-taking strategy by six test takers and two instructors: with a beginner level of English ability, the listening section is not worth preparing for, so test takers with beginner-level listening ability give up and focus on other sections to achieve the target scores. As a certain portion of test takers skip this listening section and spend the time to solve questions in other sections, test takers' responses to this part of the survey would not be highly reliable.

Nevertheless, the test takers with a certain level of listening ability demonstrated a positive attitude toward the listening section in terms of authenticity, as in the following quotes:

Minji: While the TOEIC listening audio is enunciated, the G-TELP audio is speed as if I am listening to foreigners speaking.

Nahee: The G-TELP listening has different passage types including casual conversation, presentation of information, and business conversation. Topics are usually related to common knowledge, so I like it.

Hanna: I felt that the G-TELP listening was very similar to real-life English listening, so I think it is really helpful.
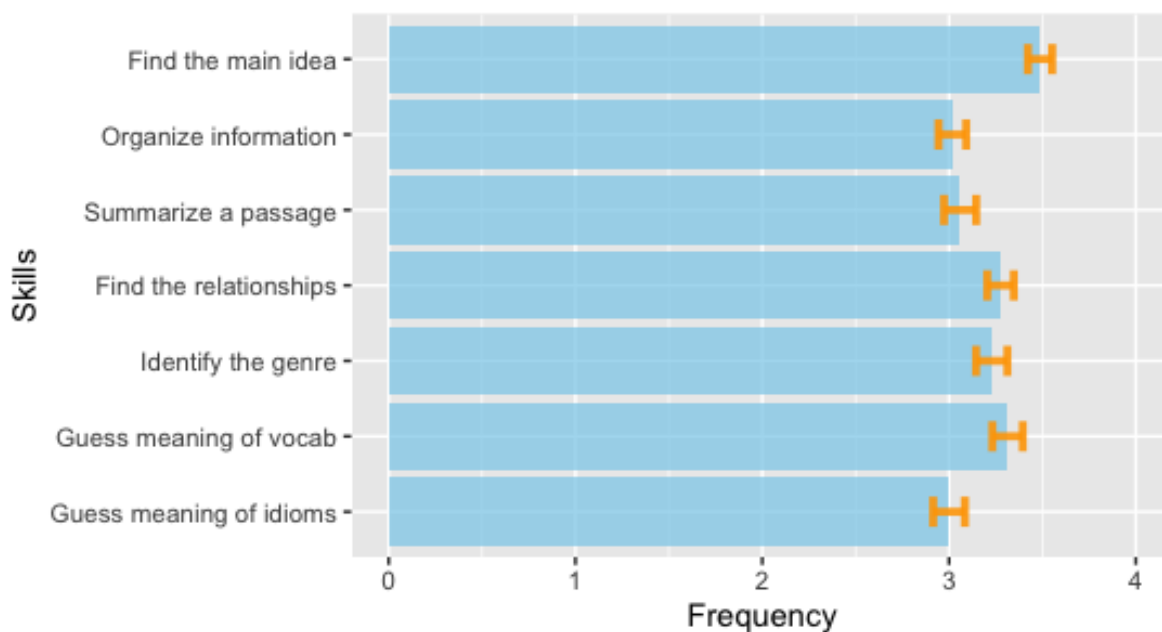
To summarize, due to most of the test takers' prevalent use of the "giving up" strategy, the current data cannot fully explain what specific skills the listening section can measure. Yet, the results provide some insights into validity issues of the test. Many test takers do not solve questions in the listening section when their listening ability is considered low or novice. Instead, since the questions are all multiple choices items, test takers end up making wild guesses rather than really solving the questions. This might be a threat to test validity since the questions are not measuring what they purport to test. Thus, it seems that the listening section can measure the listening ability only when the test takers are already quite proficient in English listening.

Reading and Vocabulary Section

As shown in Figure 14 and Table 22, the most frequently used skill was finding the main idea. Understanding the main idea is not only helpful for solving the questions directly asking test takers to identify the main idea of the reading passages, but also useful in solving other types of questions because understanding the main idea would facilitate comprehension of detailed information.

Figure 14

Frequency of Skills Practiced – Reading & Vocab

Table 22

Test Takers' Frequency of Skills Utilized for the Reading & Vocab Section

Skills	Mean	SD	Min.	Max.	Don't know
Find the main idea	3.49	0.70	1	4	8
Organize information	3.02	0.78	1	4	13
Summarize a passage	3.06	0.94	1	4	7
Find the relationships between ideas (e.g., compare/contrast, cause/effect)	3.28	0.76	1	4	9
Identify the genre of the text (e.g., newspaper article, encyclopedia entry, business letter)	3.23	0.91	1	4	7
Guess accurate meaning of vocabulary	3.31	0.88	1	4	7
Guess appropriate meaning of idiomatic expressions	3.00	0.92	1	4	8

Test takers perceived that guessing the accurate meaning of vocabulary was the second most frequently utilized skill for this section. As for this skill, the highly proficient test taker,

Sujin, shared a positive attitude. She achieved 93 points out of 100 in the reading and vocab

section, with only two items incorrect. Sujin said:

> One unique characteristic of the G-TELP reading section was the vocabulary item
>
> that asks you to identify a synonym. In other English proficiency tests that I have
>
> taken so far, this kind of questions usually had distractors that have completely
>
> different meanings from the correct answer. However, the G-TELP has distractors
>
> that can also be a meaning of the target words, which means you have to find the
>
> particular meaning in the context of the passage. I think this is really helpful for
>
> learning. Usually, people just memorize synonyms and do not understand about
>
> the actual usage of a word. I think the G-TELP helps me to learn the accurate
>
> meaning of a certain word depending on the context.

Table 23 displays instructors' ratings for how frequently they use the activities. The mean

rating for *Find the main idea* is similar to that of test takers. *Identify the genre of the text* was the

least frequently used activity in class as reported by instructors just as test takers utilized that

skill least frequently.

However, instructors reported that more frequent activities include *Guess accurate*

meaning of vocabulary, *Guess appropriate meaning of idiomatic expressions*, and *Make*

connections among pieces of information in reading passages. These practices seem to be

common teaching activities for the instructor Daeho. During the interview, he commented:

> For the reading section, you must read and understand. Otherwise, you cannot
>
> solve the questions. Test-taking strategies or memorizing patterns do not really
>
> help. In the end, when I see the G-TELP test takers, I think the casting vote is the
>
> reading section. To achieve the target score, they have to focus on the reading

section to make up for the missed points in the listening section. However, some
students really can't comprehend the reading passages. Especially when solving
the question, they cannot paraphrase the information from the reading passages.
To be better prepared for the reading section, test takers need to increase their
background knowledge because it facilitates reading comprehension. Also, they
should practice more vocabulary, expressions, and sentence structures. This is
essential for improving reading comprehension skills.

Table 23

Instructors' Frequency of Activities for the Reading & Vocab Section

Activities	Mean	SD	Min.	Max.
Find the main idea	3.43	0.53	3	4
Organize information	3.14	0.90	2	4
Summarize a passage	3.00	1.00	2	4
Find the relationships between ideas (e.g., compare/contrast, cause/effect, etc.)	3.29	0.76	2	4
Identify the genre of the text (e.g., newspaper article, encyclopedia entry, business letter)	2.86	0.69	2	4
Guess accurate meaning of vocabulary	3.57	0.79	2	4
Guess appropriate meaning of idiomatic expressions	3.57	0.53	3	4
Make connections among pieces of information in reading passages	3.57	0.53	3	4

As discussed so far, stakeholders believe that the reading and vocabulary section assesses
the general ability to read and comprehend text in English (e.g., find the main idea or find the
purpose). Stakeholders also seem to believe that the section measures the test takers' knowledge

about the meaning of vocabulary and their ability to guess the accurate meanings of the words in

a certain context.

Summary: Research Question 2

To summarize the findings for the RQ2, test takers tend to believe that the G-TELP

grammar section allows them to show their grammar knowledge and the reading/vocab section

allows them to show their reading skills and vocabulary knowledge. However, the listening

section does not seem to measure well the test takers' ability to listen in English. Nevertheless,

instructors believe that the listening section allows test takers to demonstrate their English

listening ability well. This discrepancy in their responses might be due to test takers' perceptions

of difficulty and the test-taking strategy of "giving up" on the listening section. As for specific

skills that are measured in the G-TELP, test takers believe that the grammar section assesses

their ability to apply memorized grammar knowledge, with the highest mean rating of 3.68. For

the listening and reading/vocab sections, no particular skills were found salient with relatively

lower ratings. However, one thing that should be mentioned is that test takers can have enough

time for either the grammar or reading/vocab sections by giving up on the listening section. This

prevents the test from measuring their target skills within a uniformly controlled period of time.

RQ3. How do different stakeholders perceive English language ability required for

professional purposes?

RQ3 addressed how three different groups (test takers, instructors, and score users)

perceive English language ability required for professional purposes. The perceptions of test

takers and instructors were explored during the individual interviews, and the score users' beliefs

were examined through a part of the survey. Due to the limited volume of data from instructors

and score user groups, the RQ3 could not be fully answered with this current study. However, to

partially answer the RQ3, the researcher here discusses some relevant opinions shared by the

participants in relation to how different stakeholders perceive English language requirements for

professional purposes in the Korean context and whether they perceive the G-TELP as a relevant

test to satisfy the requirement.

Test Taker Perceptions

Test takers were asked to rate how relevant the G-TELP is for different population groups.

Figure 15 and Table 24 show that the majority of respondents think the G-TELP is a relevant test

for different groups.

Figure 15

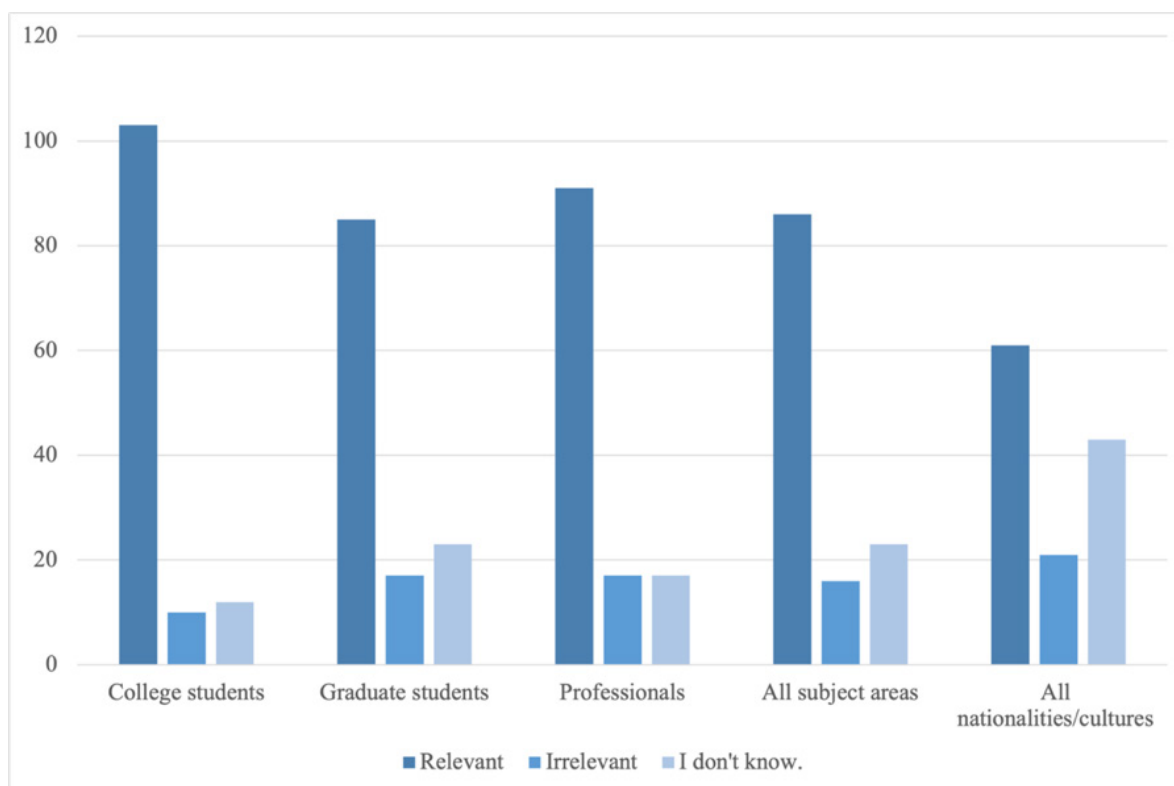*Test Takers' Perceived Relevancy of the G-TELP for Different Groups*

Table 24

Test Takers' Perceived Relevancy by Different Population Groups

Groups	Relevant (%)	Not Relevant (%)	Don't know (%)
College Students	103 (82.40)	10 (8.00)	12 (9.60)
Grad Students	85 (68.00)	17 (13.60)	23 (18.40)
Professionals	91 (72.80)	17 (13.60)	17 (13.60)
Students in all subject areas	86 (68.80)	16 (12.80)	23 (18.40)
All nationalities/cultures	61 (48.80)	21 (16.80)	43 (34.40)

Table 25

Test Takers' Perceptions About Washback (N = 125)

Please indicate your agreement with the following statements.	Mean	SD	Min	Max
Preparing for the G-TELP helped prepare me for English ability required in workplaces.	2.76	0.83	1	4
The questions on the G-TELP asked me to use English in ways that I have had to use in workplaces or I will have to use in future career.	2.63	0.83	1	4

Note. 1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree

The greatest number of respondents (82.40%) think the G-TELP is relevant for college students. However, it is interesting that the lowest number of respondents (48.89%) think the G-TELP is relevant for all nationalities/cultures and 34.40% reported that they do not know whether it is relevant or not for that particular group. The respondents were not asked to provide reasons for their answers, so it is not clear why they believe it is more relevant for college students, while they are unsure about the relevancy of the G-TELP for all nationalities and cultures in addition to Korean. This might be because most of them are currently college students (20.80%) or job seekers who just graduated (42.40%). In addition, the test takers considered taking the G-TELP to be a "rite of passage" to satisfy the requirements for national public

official exams or certification processes approved by the government. This might have led them to think that the G-TELP seems to be relevant for the Korean context while they are not sure about if it would be acceptable for other countries or cultures.

A part of the survey is also related to test takers' perception that taking the G-TELP was a "rite of passage." According to Table 25, the respondents tended to agree that preparing for the G-TELP helped them prepare for the English ability required in workplaces. However, they did not show a tendency to agree that the questions reflected the English they would use in a future career. The respondents seem to believe that preparing for the G-TELP was only done to satisfy the requirement for job applications or certification, not to prepare for the English ability that is necessary for their responsibilities at workplaces. This also can be interpreted that the G-TELP, in addition to other proficiency tests such as the TOEIC, is used for gatekeeping purposes to decide if someone is qualified to take public official exams or certification exams.

All of the interview participants mentioned this gatekeeping function of the G-TELP and other English proficiency tests. One test taker, Joon, during the interview, commented that using English proficiency scores would be a fair way to evaluate applicants. He said:

> To be honest, at my previous workplace, I didn't use English at all. I just take the G-TELP because I have to satisfy the requirement. I don't think preparing for the G-TELP would help increase my English performance required for my future job. I don't think about this issue that much because it is just a rite of passage to land a job.

Another test taker, Nahee, shared her opinion about why English test scores are necessary for job applications. She is preparing for the certification exam to become a certified public labor attorney, and the minimum requirement is the average score of 65 points on the G-TELP. She said:

> To be honest, for a public labor attorney's job, English language proficiency is not essential. However, I believe that you should have at least a basic English ability to read and comprehend. The certification is only available to those who prove that they have the basic level of English language ability. I think the English language requirement is the preliminary round for the certification process.

Sujin, who is preparing for exams for certified public accountant (CPA), also shared a similar perspective:

> I am not sure if English would be frequently used on the job as a CPA. If you have a client overseas or need to collaborate with a foreign company, then you need it. I heard that it would give you an advantage. In fact, the minimum score is 65 on the G-TELP or 700 on the TOEIC, which is not a higher score. That's just average. I think the requirement of 65 does not really evaluate English ability, but rather, they want to know if you have the basic knowledge about English.

Another test taker, Yoon, is preparing for exams for tax accountant certification. Her perception also resonates with the previous two test takers:

> I think English test scores are required because lawyers or accountants are professionals. Especially for a tax accountant, I don't think a higher score is necessary. That is why they required the minimum of the G-TELP 65. I feel it's just average. I did not even prepare for the test that much, but I still received the minimum score. If you really prepare for the test, you can achieve higher scores. Maybe the G-TELP is not a good test for employers, but to test takers, it's a great test.

> The interview data suggest that test takers interpret the English language ability required for professional purposes as not being necessary for linguistic performance in English

communication. Instead, test takers perceive the required English test scores as "the evidence of

self-discipline and efforts to prepare for the test." This perspective is similar to score users'

interpretation of the TOEIC scores (Im & McNamara, 2017) that the test scores are indicators of

"a reflection of an applicant's effort" rather than of English proficiency. Also, test takers believe

that the required English proficiency for professional purposes is "the basic knowledge about

English with the ability to learn English on the job when it becomes necessary." These

perceptions are values attached to the test scores, but not necessarily recommended by ITSC.

The perception of taking the G-TELP as a "rite of passage," is closely related to the test-

taking strategies discussed earlier. Figure 16 shows the disproportionate distribution of self-

reported scores across the three sections, which indicates that the respondents in the survey have

utilized the test-taking strategies of "giving up on the listening section" and "focusing on the

grammar section" to achieve the target scores.

As summarized in Table 26, the grammar section has more respondents who achieved

scores between 90-100 (27.20%) than the listening and reading/vocab sections. Listening section

in particular has a large portion of respondents who received lower scores, less than 50 (23.20%)

or around 50-60 points (26.40%). However, there are a smaller number of respondents who

received scores around 60-70 points (13.60%), with a slightly higher number of people with

scores of 70-80 points (28.00%). This might be because beginner-level test takers just gave up on

the section while higher level test takers made a sincere effort to solve the questions in the

listening section. One instructor, Daeho, commented on this phenomenon:

> The G-TELP has a gap in measuring test takers' English ability. It measures abilities
>
> across three different sections pretty well for the test takers at a quite proficient level.

However, for those who do not have that level of proficiency, their performances are not

measured very well.

Nevertheless, to satisfy the requirements for professional purposes such as job applications,

certification processes, public official exams, or on-the-job training decisions, only the average

scores are used. As the grammar section is perceived to be easier for improving scores than other

sections, by memorizing grammar rules, test takers focus on that section and sacrifice the

difficult listening section. In the end, these test preparation and test-taking practices are chosen

because test takers consider the process to be a "rite of passage" rather than a way to achieve a

balanced English ability across skills.

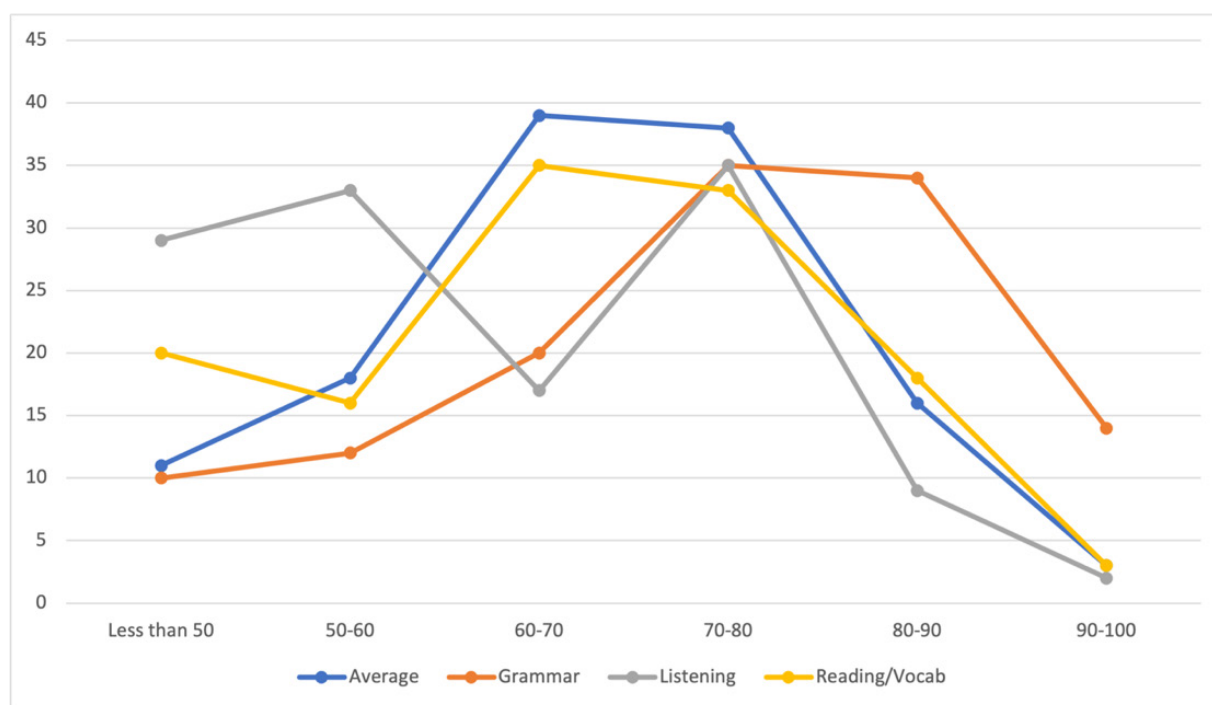**Figure 16**

Self-Reported G-TELP Scores

Table 26

Self-Reported G-TELP Scores (N = 125)

Score range	Average (%)	Grammar (%)	Listening (%)	Reading/Vocab (%)
Less than 50	11 (8.80)	10 (2.40)	29 (23.20)	20 (16.00)
50-60	18 (14.40)	12 (8.00)	33 (26.40)	16 (12.80)
60-70	39 (31.20)	20 (9.60)	17 (13.60)	35 (28.00)
70-80	38 (30.40)	35 (16.00)	35 (28.00)	33 (26.40)
80-90	16 (12.80)	34 (28.00)	9 (7.20)	18 (14.40)
90-100	3 (2.49)	14 (27.20)	2 (1.60)	3 (2.40)

Instructor Perceptions

Instructors were also asked to rate how relevant the G-TELP is for different population groups or purposes. Table 27 shows that instructors agree that the G-TELP is relevant for college students or undergraduate-level learners.

Table 27

Instructors' Perceived Relevancy by Different Population Groups or Purposes

Groups	Relevant	Not Relevant	Don't know
At the pre-university level (high school level below)	3	3	1
At the undergraduate level	6	1	0
At the graduate level	4	1	2
For vocational studies	1	2	4
For job application	4	0	3
For promotion	3	1	3

About using the G-TELP for different purposes, instructors were not in agreement. For example, some instructors believe that the G-TELP is relevant for job applications while others are not sure. With this limited data, it would be hard to describe how instructors perceive the G-TELP as a proficiency test required for professional purposes.

As for the use of the G-TELP test scores for professional purposes, the two instructors that were interviewed demonstrated their understanding about how test takers perceive the use of the test. The instructor Hyunwoo shares his story:

One time, I organized a review session to cover some of the latest G-TELP test items. I explained a lot of grammar points and made sure that I provided a lot of information for learning English. However, lots of students were not happy. They just wanted to know the answers to calculate their scores. After that, I made sure I met their needs.

Hyunwoo realized that the G-TELP test takers were more interested in getting the items correct to achieve the target scores rather than increasing their English knowledge and ability. After several years of teaching the G-TELP preparation courses, Hyunwoo has developed this perspective:

I believe that being able to employ some test-taking strategies and studying to the test have both advantages and disadvantages. For example, the grammar section is easier to prepare for and accessible to test takers. Practically speaking, the grammar section is the key to achieving the target score. Especially, those people who prepare for army public official exams only need to receive the average score of 32 or above. To save their time to study other subjects, they focus a lot on the grammar section to prepare for the G-TELP. The preparation practice is a little bit deformed. Personally, I think 32 is too low for a minimum requirement, and it's better not to require that test score.

Hyunwoo lamented this reality that test takers are employing the strategies of "give up on the listening and focus on the grammar" and even suggested raising minimum scores. The other instructor, Daeho, also suggested improving the grammar section so that it can cover more

grammar rules, but he was worried that changes to the grammar section would cause the loss of potential test takers:

Once test takers think the test overall is difficult and it is not easy to improve test scores, they would not choose the test. It's a serious issue. Anyway, from test takers' perspective, the G-TELP is a good test, since you receive your score within a short time, and you spend less time for preparation. In the end, what they need to do is to satisfy the requirements with the minimum score.

In addition, Daeho expressed concerns about the way the G-TELP scores are perceived and used by the score users:

Currently, G-TELP 65 could be achieved by Grammar 90, Listening 40, and Reading/vocab 65. If an English proficiency test is an ideal test, the score should be reported differently. If a test taker receives the score of G-TELP 65, that person should be able to receive 65 or similar points across sections.

When he was asked about the gatekeeping function of the G-TELP, he said:

Well, not a lot of people know the G-TELP yet, so it is hard to discuss. For example, the G-TELP score of 87 is a very high score. Let's suppose this person submitted a job application. The employers or people in charge probably won't understand the meaning of the score because they are mostly familiar with the TOEIC. The G-TELP should be promoted to college students and companies to expand the market base.

As Daeho did, all the participants in the interviews compared the G-TELP to the TOEIC as a test used for professional purposes. The instructor Hyunwoo said:

For the TOEIC, test takers need to solve questions with a lot of time pressure,

which makes it difficult to measure their true ability. Also, because the TOEIC

has more questions and the test is longer, there might be an unintended fatigue

effect. Thus, to take the TOEIC, test takers need to train themselves to solve

questions under time pressure. I believe the G-TELP does not require this training,

so I think it is more relevant to measure general English ability.

Hyunwoo's comparison between the G-TELP and the TOEIC is concerned with the validity

issue. Being unable to answer the questions only because of time pressure would probably result

in less valid results of measuring reading skills (De Luca et al., 2013). Hyunwoo believes that the

TOEIC might have validity issues because test takers need to complete 100 questions within a

time limit and also because sitting down for 2.5 hours for the test would cause fatigue. By

contrast, he believes the G-TELP would be able to maintain the integrity of assessing the reading

skill with a lower number of questions and with more time available.

Score User Perceptions

Test score users were also asked to rate the relevancy of the G-TELP, but due to their

lack of familiarity with the test itself, the majority of responses were "I don't know." Only one

respondent who works for the HR department at a conglomerate considered it to be relevant for

vocational studies and promotion. He further explained that "since it was officially announced

that the G-TELP is acceptable for several public and private institutions, I think it can be

relevant."

One thing that should be mentioned is that English language test score users in this study

were not familiar with the G-TELP. As Table 28 shows, most score users reported their lack of

familiarity with the test, by choosing "Don't know" as an answer to the question about how

relevant the G-TELP is for each population group. In response to open-ended questions, two of

the participants actually reported that they had never heard of this test, and another two reported

lack of familiarity and knowledge about the test. All of the respondents in private sectors

reported that their institution was not officially utilizing the G-TELP scores for the hiring or

promotion process. As for why the G-TELP is not officially used, the respondents were asked to

provide the reasons. One respondent said, "We already have a lot of other proficiency tests

available for use, such as the TOEIC," and another respondent commented, "The G-TELP is not

very widely known, so the majority of applicants are submitting their TOEIC or TEPS scores."

One public official who works for a government agency (the interviewee Changmin) reported, "I

am aware of the test, the G-TELP, but there has been no case where G-TELP scores were

submitted to us so far." Changmin strongly insisted that the G-TELP should be more widely

known if this score is to be used.

Table 28

Score Users' Perceived Relevancy by Different Population Groups or Purposes

Groups	Relevant	Not Relevant	Don't know
At the pre-university level (below high school level)	0	0	6
At the undergraduate level	0	0	6
At the graduate level	0	0	6
For vocational studies	1	0	5
For job performance evaluation	0	0	6
For promotion	1	0	5

As the researcher was already aware that score users might not be familiar with the G-

TELP test, a set of questions was asked with sample items and passage types.

For the grammar section, score users were asked to evaluate the relevancy of the difficulty level. Five out of six considered the level to be relevant while one respondent believed it is easy. As for the listening and reading/vocabulary sections, sample passages of different text types were presented to the participants. Then, they were asked to rate the usefulness and necessity of the skills to comprehend each of the text types or genres (see Table 29). All of the score users believed that the listening skills required for understanding conversations were very necessary while comprehending presentations and explanations of process were necessary to a lesser extent. As for the reading skills, all the respondents said the reading comprehension of business letters is a very necessary skill for workplaces. While comprehension skills for magazine articles were still perceived as necessary, understanding biography and encyclopedia articles was considered less necessary. Based on this result, it can be said that inclusion of business letters in the reading section is a case of positive construct representation.

Table 29

Perceived Necessity of English Comprehension Ability Assessed in the G-TELP

Sections	Text Types	Not Necessary	Rarely Necessary	Somewhat Necessary	Very Necessary
Listening	Conversation	0	0	0	6
	Presentation	0	0	4	2
	Explanation of Process	0	0	2	4
Reading & Vocab	Biography	0	3	3	0
	Magazine article	0	0	2	4
	Encyclopedia article	0	2	3	1
	Business letter	0	0	0	6

Score users were also asked to share their opinions on whether speaking and writing skills should be included in a proficiency test and why. Five of them said that speaking and writing skills should be included for the following reasons:

- Speaking ability is essential for virtual meetings with foreign companies, and writing ability is important for writing business letters.

- Korean learners usually lack speaking ability, so it should be assessed in the test.

- Speaking and writing skills are critical for communicating with foreign business partners.

The three reasons were shared by the respondents in private sectors. One respondent from government agency said that overseas training decision requires evidence of English-speaking ability, so speaking ability should be included in the test. Assessing speaking and writing skills would improve construct validity because professionals at workplaces believe those skills are necessary and useful. The same issue was discussed in Lee (2018) as public officials on the job think speaking is the most necessary skill, which is not assessed in popular tests such as the TOEIC.

Summary: Research Question #3

Most test taker participants indicated that the G-TELP would be relevant for most of the population to be used for professional purposes. Another important finding is that test takers consider the test a "rite of passage," which not only assesses the "language" aspect or "communicative" ability, but also assesses the learning ability and affective elements (e.g., motivation or determination to prepare for an English proficiency test). By achieving the required cut-off score, test takers can prove themselves as someone who possesses a certain level
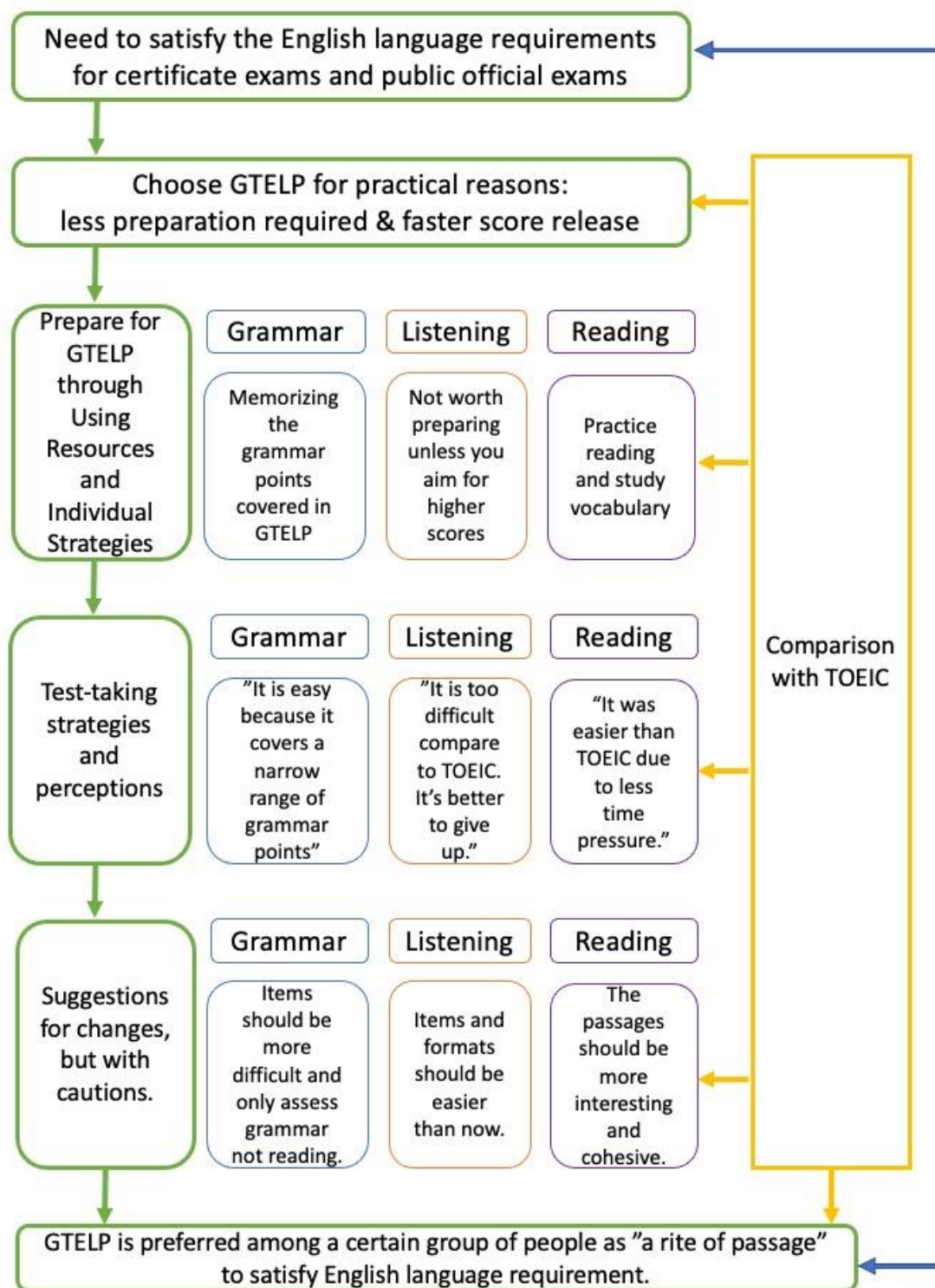
of basic English knowledge and the ability to learn in order to further develop their English skills when it becomes necessary on the job. Being aware of test takers' perception, instructors can also support them to successfully pass the stage with useful strategies. Score users provided valuable suggestions that the G-TELP should be more widely known while construct validity can be improved with major and minor changes in the test.

Summary of Findings

This section summarizes the overall findings with the visual representation of stakeholder beliefs around the G-TELP test (see Figure 17) and discusses some issues that need attention from test developers and decision makers for test use.

Test takers of the G-TELP choose to take the test to satisfy the preliminary requirement and to become qualified for taking the certification exams or civil service exams. Their decision to take the G-TELP over other tests is attributable to several factors, such as shorter test length, faster score release, and less time and effort needed to achieve the minimum target score. The reference for those factors was the TOEIC. As 83 test takers (66.40%) in this study have also taken the TOEIC, during the interviews, participants from all the stakeholder groups made comparisons between the two tests based on their test preparation and test-taking experiences.

Test takers and instructors all believed that the grammar section is easy and that it only covers a narrow range of grammar points. This characteristic was also pointed out in Lee (2018), where experts evaluated the G-TELP grammar section as restricted in coverage of grammar points. This might cause validity issues in terms of test use because when test takers need to communicate in English at their future workplaces, they might need grammar knowledge that is not currently covered in the G-TELP.

Figure 17

Summary of Stakeholder Beliefs About the G-TELP

As for the listening section, test takers in general perceived it to be extremely difficult. This might be due to the extended length of the listening section, with six or seven questions per each passage. For this same reason, English language education experts in Lee's (2018) study evaluated the G-TELP listening section as lacking authenticity. The long listening passages might pose a threat to validity because test takers with better working memory capacity would perform better or because test takers are likely to experience too much cognitive load, which might cause unexpected variances in assessing their listening ability.

The reading section did not receive much attention during the interview compared to other two sections for which stakeholders had strong opinions and beliefs. However, test takers were mostly positive about the section partly due to the fact that they could spend more than the designated time of 40 minutes working on it. Some test takers reported enjoying the various topics of the G-TELP more than the TOEIC passages, which were written in the context of business communication. Nevertheless, the instructor Hyunwoo suggested possible changes because the G-TELP reading sections have a restricted coverage of topics such as history, science, or autobiography, as reported in Lee (2018). Hyunwoo suggested including more reading passages that deal with topics such as business or economics. That way, the test can be more valid for the majority of test takers who would need knowledge about those topics as professionals in public/private sectors or as public accountants.

Participants in the study consider this test "a rite of passage" or a qualification they have to fulfill for the next steps in their employment or certification. This is similar to the findings from Kim et al. (2019), where Korean job seekers perceived the TOEIC test as a "gate." This also resonates with the findings from the Sinclair et al. (2019) study, which analyzed the participants' narratives around the TOEIC or IELTS test preparation and test-taking, using the

Bordieuan framework. The international student participants in that study referred to the high-stakes English tests as "filters" or a method of "screening" people, which clearly shows that the language tests and related policies are structures "capable of guiding and constraining [subjects'] practices or representations" (Bourdieu, 1989, p.14). As the high-stakes tests were obstacles to overcome and potential supports for learning opportunities at the same time for the international students in the Sinclair study, the participants in this study also had to overcome the obstacles of English tests to land a job or to become a professional. This nature of "high-stakes" encouraged them to use whatever strategies necessary to succeed in the "rite of passage" or to clear the first hurdle.

Recommendations and Conclusion

This study examined the beliefs of three groups of stakeholders on the G-TELP: test takers, test preparation course instructors, and score users. The purpose of the study was to investigate what these groups believe about the G-TELP in terms of effective preparation and test-taking strategies, each section of the test, and the English ability required for professional purposes.

Limitations of the Current Study

This study aimed to investigate stakeholder beliefs about the G-TELP through surveys and individual interviews. Throughout the data collection process, challenges in recruitment and participation rates led to some limitations of the results. For example, it was not possible to interview any score users who were familiar with the G-TELP and had experiences of utilizing the G-TELP scores. The difficulty of recruitment was also true for test preparation course instructors as well, so only two instructors agreed to participate in the interviews. In the end, the data for this study were limited to a small sample size. However, the sample size for the test

taker group exceeded the target numbers in both surveys and interviews. Future research should use more systematic and effective recruitment strategies to have more participants from different stakeholder groups, as a larger number and greater variety of stakeholders would provide a better understanding of stakeholders' beliefs (Malone & Montee, 2014). As for the data collected, this study only used respondents' self-reported English proficiency levels as a proxy. The comparisons among different proficiency levels made in this study might not be valid. In future studies, it would be desirable for the researcher to obtain G-TELP test results from ITSC, so that more rigorous research can be conducted. In addition, this study only relied on surveys and interviews as research methods. Future research can include other research methods such as focus groups or stimulated recalls for data triangulation to gain a clearer picture of stakeholder beliefs. Finally, to make sure the current use of the G-TELP for professional purposes is legitimate and fair, more research is necessary to explore target domains using observations or logging techniques to identify what kinds of functions professionals need to perform in English on their jobs. The results from the research should be communicated with ITSC to improve the test for more legitimate test use.

Recommendations

As the G-TELP has been utilized by a relatively small group of populations, these beliefs were shaped by a limited number of people. Nonetheless, a few recommendations emerged from both quantitative and qualitative analysis of the current data.

Recommendation 1: Employ More Effective Strategies for Marketing and User Education

The current study revealed that English proficiency test score users had almost no knowledge about the G-TELP. Preparation course instructors also pointed out the current status of the G-TELP and said they hope it becomes more widely known and chosen by more English

proficiency test takers in Korea. Currently, even though they have heard of the test, they might need extra information. For example, interpretation of the G-TELP scores is different from that of the TOEIC scores, which most Koreans are familiar with. This necessitates a wider dissemination and education of the score interpretation of the G-TELP. The more test user education the company provides to potential score users, the more they can raise the recognition of the test as well. As raising awareness of particular tests among users takes a considerable time (Ginther & Elder, 2014), ITSC should develop effective strategies for marketing and test user education. Especially for user education, it is recommended that the company plan for both in-person and virtual workshops in collaboration with applied linguists to increase assessment literacy of all stakeholder populations through different channels.

Recommendation 2: Investigate Construct Validity of the Test

Despite the use of the G-TELP for professional purposes in Korea, test takers considered the test a "rite of passage" rather than an assessment of their English ability. To ensure the legitimacy and appropriateness of the test use, it seems important to improve construct validity to argue for more widespread use of the test. The target domain of the G-TELP is general communication contexts within a wide range of tasks. On the official webpage of G-TELP Korea, it is stated that the G-TELP, particularly Level 2, is used for English language requirements of civil service exams and the target contexts are "the contexts of task completion at workplaces or overseas training." Although there will be differences among different contexts, the investigation into the workplaces would provide valuable information for target domain of the test. For example, Korean public officials were found to believe that speaking skill is the most important skill in their work responsibilities (Han, 2015; Kim, 2007). This perspective was also reported from the public official interviewee in the current study. Nevertheless, the G-TELP Level 2

currently is separate from the G-TELP Speaking and Writing tests. It is important to note that including a speaking section would affect practicality of the test (e.g., more expensive registration fee or more wait time for score release). It would not be a practical suggestion that can be reflected soon. If the G-TELP continues to aim for faster assessment of general English language ability of the test takers, an alternative would be to investigate if the G-TELP scores can effectively predict speaking performance ability.

Recommendation 3: Investigate Each Section for Improvement

The grammar section should be investigated to determine if it is as level-appropriate as ITSC claims it to be. Currently, participants in this study perceive it to be too easy and believe it only assesses a limited range of grammar knowledge. However, the instructor Hyunwoo claimed that the perception might be formed because test takers do not have to completely understand the meaning and the context to find the correct answer. To improve the grammar section, it seems necessary to ensure the integrity of measuring the "grammar knowledge" without issues such as construct underrepresentation and construct irrelevant variances (e.g., excessive practice of item patterns or memorization of item patterns).

As for the listening section, participants in this study had the least positive evaluation in terms of test preparation and difficulty level. It was found that beginner-level test takers were most likely to make wild guesses for the section rather than sincerely answering the questions. Although the length and the speed of passages were positively evaluated by two high-performing test takers during the interview, they suggested that compared to the TOEIC test the G-TELP lacks items that assess comprehension skills for shorter audios and the G-TELP asks too many questions (6-7 questions) per passage while the TOEIC has 3 questions per each passage. It was also pointed out that not providing questions on the test sheet seems to require more cognitive

load for the section, which might harm the integrity of measuring the listening comprehension ability.

The reading section did not cause many strong reactions or perceptions from the stakeholders. Some test takers and an instructor suggested including passages of more diverse topics, which would increase motivation and attention of test takers during the test. The test score users rated the business letter test types as necessary and useful, which supports the construct of the reading section. Future research can add more evidence of construct validity by examining what text types and topics are more relevant for professional purposes. However, there might be some reliability issues of the reading section because test takers are allowed to use more than 40 minutes by using the time designated for the listening section. This might need more investigation.

Recommendation 4: Accommodate Stakeholders' Needs

In addition to validity issues, practicality of the test should be considered. All the instructors and score users who participated in the current study suggested that the G-TELP should use more effective marketing and promotion strategies. Several test takers reported that the registration fee is too expensive without any discount coupons. A few suggested more frequent testing administrations and more testing sites.

Another important aspect is the impact of the test. Currently, test takers have reported that they gain information about the test through searching the reviews of test takers or freely available YouTube lectures. The majority of information from those sources encourages them to use test-taking strategies, including giving up on the listening section and memorizing only the grammar points that are covered on the test. More research is necessary to examine the impact of

these strategies on English learning or future performances, but positive washback should be

promoted to improve the quality of the test.

References

Alderson, J. C. (2010). A survey of aviation English tests. *Language Testing, 27*(1), 51-72.

    https://doi.org/10.1177/0265532209347196

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback.

    Language Testing, 13(3), 280-297. https://doi.org/10.1177/026553229601300304

American Educational Research Association, American Psychological Association, & National

    Council on Measurement in Education. (2014). *Standards for educational and*

    psychological testing. American Educational Research Association.

Ata, A. W. (2015). Knowledge, Education, and Attitudes of International Students to IELTS: A

    Case of Australia, *Journal of International Students, 5*(4), 488-500.

    https://doi.org/10.32674/jis.v5i4.410

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what

    we count counts. *Language Testing, 17*(1), 1-42.

    https://doi.org/10.1177/026553220001700101

Bachman, L. F. (2007). What is the construct? The dialectic abilities and contexts in defining

    constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner

    & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–72). University of Ottawa.

Baker, B. A., Tsushima, R., & Wang, S. (2014). Investigating language assessment literacy:

    Collaboration between assessment specialists and Canadian university admissions

    officers. *Language Learning in Higher Education, 4*(1), 137.

    https://doi.org/10.1515/cercles-2014-0009

Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL-iBT writing tasks. *Language Testing, 31*(2), 241-259. https://doi.org/10.1177/0265532213509810

Barkaoui, K. (2015). *Test-takers' writing activities during the TOEFL iBT writing tasks: A stimulated recall study* (TOEFL iBT Research Report No. 25). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12050

Barnes, M. (2016). The Washback of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education, 41*(7), 158-174.

Bourdieu, P. (1989). Social space and symbolic power. Sociological theory, 7(1), 14-25. https://doi.org/10.2307/202060

Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.

Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, *16*, 104–122. https://doi.org/10.1080/10627197.2011.584042

Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39-62. https://doi.org/10.1177/0265532207083744

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: CUP/Council of Europe

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 107-145. https://doi.org/10.1191/0265532204lt278oa

DeLuca, C., Cheng, L., Fox, J., Doe, C., & Li, M. (2013). Putting testing researchers to the test: An exploratory study on the TOEFL iBT. *System, 41*, 663–676. https://doi.org/10.1016/j.system.2013.07.010

Eom, M., Lang, Y., & Xie, C. (2017). Chinese and Korean college students' perceptions of standardized English tests. *The Asian Journal of Applied Linguistics. 4*(1). 117-128. Retrieved from https://caes.hku.hk/ajal/index.php/ajal/article/view/410

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. *ETS Research Report Series, 2014*(2), 1-39. https://doi.org/10.1002/ets2.12037

Green, A. (2013). Washback in language assessment. *International Journal of English Studies, 13*(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17-27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly, 32*(2), 329-337. https://doi.org/10.2307/3587587

Hyatt, D. (2013). Stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK. *Journal of Further and Higher Education, 37*(6), 844-863. https://doi.org/10.1080/0309877X.2012.684043

Im, G. H. (2019). *Stakeholder voices: Validity argument for score meaning and use of the test of English for international communication (TOEIC) in international business workplaces*

in South Korea. [Doctoral dissertation, Queen's University].

https://qspace.library.queensu.ca/handle/1974/26679

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.

https://doi.org/10.15858/engtea.72.2.201706.71

International Testing Services Center (ITSC) (2015). http://www.itsc-group.com/

Jamieson, J., Taylor, C., Kirsch, I., & Eignor, D. (1999). *Design and evaluation of a computer-based TOEFL tutorial (TOEFL research report No. RR-66)*. Princeton, NJ: Educational

Testing Service. https://doi.org/10.1002/j.2333-8504.1999.tb01799.x

Jeon, M. (2012). Globalization of English teaching and overseas Koreans as temporary migrant

workers in rural Korea. *Journal of Sociolinguistics, 16*, 238–254.

https://doi.org/10.1111/j.1467-9841.2011.00527.x

JTBC News. (March 9, 2016). JTBC Newsroom. Retrieved from

http://news.jtbc.joins.com/article/article.aspx?news_id=NB11189537&pDate=20160309

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*, 527–

535. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues

and Practice*, *21*, 31–35. https://doi.org/10.1111/j.1745-3992.2002.tb00083.x

Kang, I. H., Baek, H., & Yoo, D. (2011). A study on the practicality of open competitive exam

for local public servants: Based on a survey of general administrative local public

servants' awareness. *지방정부연구, 15*(3), 145-160. Retrieved from

https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE01800301

Kim, E.-Y. J. (2017). The TOEFL iBT writing: Korean students' perceptions of the TOEFL iBT

writing test. *Assessing Writing*, *33*, 1–11. https://doi.org/10.1016/j.asw.2017.02.001

Kim, M., Choi, D. I., & Kim, T. Y. (2019). South Korean Jobseekers' Perceptions and (De)

motivation to Study for Standardized English Tests in Neoliberal Corporate Labor

Markets. *The Asian EFL Journal, 21*(1), 84-109.

https://www.elejournals.com/category/asian-efl-journal/

Lee, J. (2018). An Analysis Study of the Test Usefulness of Five Standardized EFL Tests for

Korean Civil Service Employment Exam. *Studies in English Education, 23*(3), 739-768.

http://dx.doi.org/10.22275/SEE.23.3.04

Llosa, L., & Malone, M. E. (2017). Student and instructor perceptions of writing tasks and

performance on TOEFL iBT versus university writing courses. *Assessing Writing, 34*, 88-

99. https://doi.org/10.1016/j.asw.2017.09.004

Ma, J., & Cheng, L. (2015). Chinese Students' Perceptions of the Value of Test Preparation

Courses for the TOEFL iBT: Merit, Worth, and Significance. *TESL Canada Journal*,

33(1), 23. https://doi.org/10.18806/tesl.v33i1.1227

Macqueen, S., Pill, J., & Knoch, U. (2016). Language test as boundary object: Perspectives from

test users in the healthcare domain. *Language Testing*, *33*, 271–288.

https://doi.org/10.1177/0265532215607401

Malone, M. E., & Montee, M. (2014). Stakeholders' beliefs about the TOEFL iBT® test as a

measure of academic language ability. *ETS Research Report Series, 2014*(2), 1-51.

http://dx.doi.org/10.1002/ets2.12039

McIver, J., & Carmines, E. G. (1981). Unidimensional scaling (Vol. 24). Sage.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting

teachers' judgments: A validity study of a curriculum-embedded performance assessment in

kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73-95.

https://doi.org/10.3102/00028312038001073

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–

103). American Council on Education & Macmillan.

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13, 243–256.

Mihara, K. (2015). An analysis of the differences among L2 listening comprehension test

formats. *Language Testing in Asia, 5*(1), 12. https://doi.org/10.1186/s40468-015-0021-5

Ministry of Personal Management (2018). https://www.gosi..kr

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, *23*, 5–12.

https://doi.org/10.3102/0013189X023002005

Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from

interpretive research traditions. *Educational Researcher*, *25*, 20–28, 43.

https://doi.org/10.3102/0013189X025001020

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review

of Research in Education, 30*, 109–162. https://doi.org/10.3102/0091732X030001109

Murray, J. C., Cross, J. L., & Cruickshank, K. (2014). Stakeholder perceptions of IELTS as a

gateway to the professional workplace: The case of employers of overseas trained teachers.

IELTS Research Reports Online Series, 78. https://www.ielts.org/-/media/research-

reports/ielts_online_rr_2014-1.ashx

O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33-56. https://doi.org/10.1191/0265532202lt219oa

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Silalahi, R. M. (2014). IELTS teachers and test takers' attitudes towards IELTS writing tasks. *Proceedings of ISELT FBS Universitas Negeri Padang, 2*, 350-357. http://ejournal.unp.ac.id/index.php/selt/article/view/6726

Sinclair, J., Larson, E. J., & Rajendram, S. (2019). "Be a Machine": International Graduate Students' Narratives around High-Stakes English Tests. *Language Assessment Quarterly, 16*(2), 236-252. https://doi.org/10.1080/15434303.2019.1628238

Stoynoff, S. (2009). Recent Developments in Language Assessment and the Case of Four Large-Scale Tests of ESOL Ability. *Language Teaching*, *42*(1), 1–40. https://doi.org/10.1017/S0261444808005399

Stricker, L. J., & Attali, Y. (2010). Test Takers' Attitudes About the TOEFL iBT™. *ETS Research Report Series*, *2010*(1), i–16. https://doi.org/10.1002/j.2333-8504.2010.tb02209.x

Stricker, L. J., Wilder, G. Z., & Rock, D. A. (2004). Attitudes about the computer-based Test of English as a Foreign Language. *Computers in Human Behavior, 20*(1), 37-54. https://doi.org/10.1016/S0747-5632(03)00046-3

Taki, Y. (2017). Effects of Explicit and Implicit Strategy Instruction on Reading Strategies. *松山大学論集, 29*(5), 25-48. https://core.ac.uk/download/pdf/230512753.pdf

Taylor, L. (2005). Washback and impact. *ELT Journal, 59*(2), 154-155.

https://doi.org/10.1093/eltj/cci030

Tran, T. T., & Nguyen, H. T. (2018). Investigating the Vietnamese EFL Learners' Affective And

Cognitive Attitudes Towards The IELTS Listening Test. *Advances in Social Sciences

Research Journal, 5*(7), 390-401. https://doi.org/10.14738/assrj.57.4912

Wall, D., & Horák, T. (2008). The impact of changes in the TOEFL examination on teaching and

learning in Central and Eastern Europe: Phase 2, coping with change. *ETS Research

Report Series, 2008*(2), i-105. https://doi.org/10.1002/j.2333-8504.2008.tb02123.x

Wang, Y. (2019). The Impact of TOEFL on Instructors' Course Content and Teaching Methods.

TESL-EJ, 23(3), 3. https://www.tesl-ej.org/wordpress/issues/volume23/ej91/ej91a2/

Yu, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., . . ., Fang, L. (2017). *Preparing for

the speaking tasks of the TOEFL iBT test: An investigation of the journeys of Chinese test

takers (Research Report No. RR-17-19).* Princeton, NJ: Educational Testing Service.

https://doi.org/10.1002/ets2.12145

Zahari, D. A., & Dhayaalan, J. (2016). The perception of Malaysian ESL tertiary level students

on the IELTS test. *Asian Journal of Education and Training*, *2*(1), 1-6. Retrieved from

https://ideas.repec.org/a/aoj/asjoet/2016p1-6.html

Zahedkazemi, E. (2015). Construct validation of TOEFL iBT (as a conventional test) and IELTS

(as a task-based test) among Iranian EFL test-takers' performance on speaking modules.

Theory and Practice in Language Studies, 5, 1513–1519.

https://doi.org/10.17507/tpls.0507.27

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 from https://www.youtube.com/watch?v=0L6Lr2BzuSQ

Appendix A.

Test Taker Survey

1. What is the highest level of education you have completed?
• High school/secondary school diploma
• College/undergraduate degree (AB, BA, BS)
• Graduate degree (Master's)
• Graduate degree (Doctoral, etc.)
• Other (please specify)

2. What was/is your major? (open-ended)_____

3. At which level are you currently studying or where are you working? (Select all that apply.)
• College/undergraduate level
• Graduate level (Master's)
• Graduate level (Doctoral, etc.)
• I am looking for a job.
• I work at government agencies.
• I work in private sector.
• I am self-employed
• Other (please specify)

4. Apart from any English classes you took up to high school, how would you describe your English learning experiences? Please be specific. (e.g. I majored in English, I am a member of English study group, I participated in study abroad program, …. etc.)

_____

_____

_____

5. Please tell us about your English proficiency level.
5-1. For each of the subskills, please indicated your ability: advanced, intermediate, or beginner

Subskills	advanced	intermediate	beginner
Listening			
Reading			
Speaking			
Writing			

5-2. For each of the subskills, please choose the most relevant description of your English ability.

	1	2	3	4	5	6
L i s t e n i n g	I can recognize familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programs on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programs. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signaled explicitly. I can understand television programs and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
R e a d i n g	I can understand familiar names, words and very simple sentences, for example, on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialized articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally, or linguistically complex texts such as manuals, specialized articles, and literary works.

			letters.			
Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skillfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.

			book or film and describe my reactions			
Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Note. Adapted from The self-assessment grid illustrating the levels of proficiency described in the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2019).

6. Have you taken any of the following tests? (Select all that apply.)
• TOEIC (Test of English for International Communication)
• IELTS (International English Language Testing System)
• TOEFL iBT (Test of English as a Foreign Language)
• TEPS - The Test of English Proficiency developed by Seoul National University (only for Korean)
• None of the above
• Others (please specify)

7. If you have, what was your best overall score?
Which test? _____
Score? _____

8. Why did you or are you planning to take an English proficiency test? (Select all that apply.)
- My company requires English test scores for promotion.
- My company requires English test scores for overseas training or overseas positions
- My university/college requires English test scores for graduation.
- English proficiency test scores are required for job application.
- English proficiency test scores are not required, but it would be helpful for getting a job.

9. Have you ever taken G-TELP before? Or Are you planning to take G-TELP?
YES---→ all the following questions.
If NO -→ go to the very end of the survey

10. Why did you choose the G-TELP over other tests?
- I was prepared in school or class to take G-TELP.
- I can receive my scores faster by taking G-TELP than other tests.
- The G-TELP is shorter than other tests.
- The G-TELP does not require writing and speaking.
- The G-TELP registration fee is less than the fees for other tests.
- The G-TELP shows my language abilities better than other test scores.
- Other (please specify)

11-1. Did you prepare for the G-TELP?
- Yes
- No

11-2. For how many months did you prepare for the G-TELP?
- 1 month or less
- 1-2 months
- 3-4 months
- 5-6 months
- More than 6 months
- I did not prepare for the test.

12. Indicate the importance of the following activities when preparing to take the G-TELP. (1=Not important at all 2=Not important, 3=Important, 4=Very important)

Activities	1	2	3	4
• Read academic articles and books in English				
• Read practical articles in English				

• Read business letters in English				
• Practice academic vocabulary				
• Practice business English				
• Speak with native English speakers				
• Practice essay writing				
• Study many different subjects in English				
• Take a practice G-TELP				
• Take a class specifically for the G-TELP in person				
• Take an online class specifically for the G-TELP				
• Practice the G-TELP question format				
• Read online tutorial on how to take the G-TELP				
• Other (please specify)				

13. Have you ever taken any of the following courses to prepare for G-TLEP? (Select all that apply.)
 • an offline course with "G-TELP" in the title
 • an online course with "G-TELP" in the title
 • an English course that prepares you for English tests in general?

If YES to the first two options → 13.1. Otherwise, go to 14.

13.1. How useful were the different skills you practiced in your English class for the G-TELP?
(1=Not useful, 2=Somewhat useful, 3= Useful, 4= Very useful, Didn't do this)

Skills	1 Not useful	2 Some what useful	3 Useful	4 Very Useful	Didn't do this
• Reading					
• Listening					
• Grammar					
• Vocabulary					
• Practice G-TELP tests					
• Test-taking skills (taking notes, answering multiple-choice questions, etc.)					
• Other (please specify)					

14. Please indicate how difficult the sections on the G-TELP were for you.
(1=Very easy, 2=Easy, 3=Moderate, 4=Difficult, 5=Very difficult)

Skills	1 Very easy	2 Easy	3 Moder ate	4 Difficu lt	5 Very difficu lt

	1	2	3	4	
• Reading & Vocabulary					
• Listening					
• Grammar					

15. On the reading and vocabulary section of the G-TELP, how often did you use the following skills:
(1= Never, 2=Rarely, 3=Some, 4=Often)

Skills	1 Never	2 Rarely	3 Some	4 Often	Don't know
• Find the main idea					
• Organize information					
• Summarize a passage					
• Find the relationships between ideas (e.g. compare/contrast, cause/effect, etc.)					
• Identify the genre of the text (e.g. newspaper article, encyclopedia entry, business letter..etc.)					
• Guess accurate meaning of vocabulary					
• Guess appropriate meaning of idiomatic expressions					

16. On the listening section of the G-TELP, how often did you use the following skills:
(1= Never, 2=Rarely, 3=Some, 4=Often)

Skills	1 Never	2 Rarely	3 Some	4 Often	Don't know
• Take notes					
• Find the main idea					
• Find the speaker's purpose					
• Find the organizational pattern of a text (e.g. chronological order, process… etc.)					
• Draw conclusions based on what is implied					
• Make connections between pieces of information in a conversation or monologue					
• Identify the genre of the text (e.g. personal narratives, business negotiation, etc.)					

17. On the grammar section of the G-TELP, how often did you use the following skills:
(Often, Some, Rarely, Never, Don't know)

Skills	1 Never	2 Rarely	3 Some	4 Often	Don't know
• Apply grammatical rules I have memorized					
• Use a range of diverse grammatical structures					

	1	2	3	4	5
• Use knowledge about basic grammar rules (e.g. present and past verb forms, conjunctions, comparatives, superlatives)					
• Use knowledge about somewhat complex grammar rules (e.g. conditionals, auxiliary verbs, to-infinitives, gerunds, relative pronouns, etc.)					
• Use grammatical knowledge in conversational English					
• Use grammatical knowledge for English academic writing.					

18. Please indicate your agreement with the following statements.
(1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree)

Statements	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
• The test questions on the G-TELP felt natural (and authentic).				
• The reading and vocabulary section on the G-TELP let me show how well I can read in English.				
• The reading and vocabulary section on the G-TELP let me show my English vocabulary knowledge.				
• The listening section on the G-TELP let me show how well I can listen in English.				
• The grammar section on the G-TELP let me show my English grammar knowledge.				
• I had enough time to answer the questions on the grammar section.				
• I had enough time to answer the questions on the listening section.				
• I had enough time to answer the questions on the reading & vocabulary section.				
• The score of G-TELP let me know how well I can speak in English.				
• The score of G-TELP let me know how well I can write in English.				
• The score of G-TELP let me know how well I can use English in real-life situation				
• It is important to include a speaking section on a test of English as a foreign language.				
• t is important to include a writing section on a test of English as a foreign language.				
• Preparing for the G-TELP helped prepare me for English ability required in workplaces.				
• The questions on the G-TELP asked me to use				

English in ways that I have had to use in workplaces or I will have to use in future career.				

19. What was your best score on the G-TELP (out of 300 or 100%)?

Section	150 or less (below 50%)	150–180 (50-60%)	180–210 (60-70%)	210–240 (70-80%)	240–270 (80-90%)	270 or greater	I don't know or remember.
Overall							
Section	**below 50**	**50-60**	**60-70**	**70-80**	**80-90**	**90-100**	**I don't know or remember.**
Grammar							
Listening							
Reading/Vocabulary							

20. How much did you worry about taking the G-TELP
- Very much
- Some
- Very little
- Not at all

21. What are the factors that affect your performance on the G-TELP? (Select all that apply)
- Time pressure
- Length of the test
- Fear of tests
- Unfamiliarity of topics
- Distraction caused by other test takers
- Difficulty of language on the test
- Other (please specify)

22. In your opinion, is the G-TELP appropriate for the following groups? (Yes/No/No opinion)
- Professionals
- Post-graduates
- College/undergraduates
- Students in all subject areas
- All nationalities/cultures

23. What would you like to change in G-TELP? (e.g. test registration process, test content, score report formats, or any other aspects)
_____

24. Please share any of your opinions about English proficiency test.
_____

Thank you.

Appendix B.

Instructor Survey

First, we'd like to know more about your educational background and your teaching background.

1. What is the highest level of education you have completed?

• College/undergraduate degree (AB, BA, BS)

• Graduate degree (Master's)

• Graduate degree (Doctoral, etc.)

• Other (please specify)

2. Where did you earn the highest degree? _____ (in Korea or any other countries?)

3. What was your major?_____

4. How many years have you taught English (either in Korea or any other countries)?

• less than 1 year

• 1-2 years

• 3-5 years

• 5-7 years

• 7-9years

• 10 years or more

Next, we'd like to know about your current position and program.

5. Which best describes your current position? (Select all that apply.)

• Instructor at an educational institutions

• Instructor at an offline language academy

• Instructor at an online language academy

• Private tutor

• Other (please specify)

6. Where do you currently teach? (Select all that apply.)

• Test preparation center

• Private institution

• University-based EFL institution

• Tutoring company

• Self-employed

• Other (please specify)

7. How would you characterize your English teaching program? (Select all that apply.)
 • A university program for students who are trying to graduate
 • A program for students who are looking for jobs
 • A program for students who are preparing for a civil service exam
 • An online course for test-taker who are preparing for G-TELP
 • Other (please specify)

8. Do you teach or have you ever taught course(s) with "G-TELP" in the title or related to G-TELP preparation?
 • Yes
 • No

9. Which of the following best describes your G-TELP classes? (Select all that apply.)
 • A constant cycle of skills classes
 • A G-TELP preparation class held for a set number of weeks (e.g. 6-week course)
 • Students enroll until they receive their desired score on the G-TELP
 • Other (please specify)

10. Are your G-TELP classes designed exclusively for G-TELP preparation?
 • Yes
 • No, G-TELP preparation is only one component of the class

The purpose of this survey is to find out about your experience preparing students to take the G-TELP. In this section of the survey, we will ask you questions about the test and how you prepare students.
11. Have you ever taken G-TELP yourself?
 • Yes
 • No
If yes – 11.1. How many times did you take it? _____

12. How familiar are you with the G-TELP? Select the best option:
 • Very familiar
 • Somewhat familiar
 • Somewhat unfamiliar
 • Very unfamiliar

13. Do you have adequate access to materials/ information about the G-TELP?
 • Yes
 • No
 • Why or why not? (open-ended response)_____

14. Think about students preparing to take the G-TELP. How useful is it for these students to practice the following skills in relation to the G-TELP?
(1=Useful, 2=Somewhat useful, 3=Not very useful, 4=Not useful at all, Don't know)

Skills	1 Not useful	2 Some what useful	3 Useful	4 Very Useful	Don't know
• Reading					
• Listening					
• Grammar					
• Vocabulary					
• Practice G-TELP tests					
• Test-taking skills (taking notes, answering multiple-choice questions, etc.)					
• Other (please specify)					

15. What is the most important skill to teach students preparing to take the G-TELP? (open-ended response)

_____

16. Please rank the sections of the G-TELP based on the importance of preparing for them prior to the test. (Most important, Second most important, Third most important)
 • Listening section
 • Reading & Vocabulary section
 • Grammar section

16.1. Why do you think your first choice is the most important? (open-ended response)

_____

17. Do your students report that any of the following factors affect their performance on the G-TELP?
(1=None, 2= Not many, 3= Some, 4=Many)

Factors	1 None	2 Not many	3 Some	4 Many
• Time pressure				
• Length of the test				
• Students' fear of tests				
• Unfamiliarity of topics				
• Distraction caused by other test takers				
• Difficulty of language on the test				
• Other (please specify)				

18. Think about how you prepare students for the reading and vocabulary section of the G-TELP. Approximately how often do you incorporate the following tasks into your instruction?
 (1= Never, 2=Rarely, 3=Some, 4=Often)

Skills	1 Never	2 Rarely	3 Some	4 Often
• Find the main idea				
• Organize information				
• Summarize a passage				
• Find the relationships between ideas (e.g. compare/contrast, cause/effect)				
• Identify the genre of the text (e.g. newspaper article, encyclopedia entry, business letter..etc.)				
• Guess accurate meaning of vocabulary				
• Guess appropriate meaning of idiomatic expressions				
• Make connections among pieces of information in reading passages				
• Other (please specify)				

19. Think about how you prepare students for the listening section of the G-TELP. Approximately how often do you incorporate the following tasks into your instruction?
 (1= Never, 2=Rarely, 3=Some, 4=Often)

Skills	1 Never	2 Rarely	3 Some	4 Often
• Take notes				
• Find the main idea				
• Find the speaker's purpose				
• Find the organizational pattern of a text (e.g. chronological order, process)				
• Draw conclusions based on what is implied				
• Make connections between pieces of information in a conversation or monologue				
• Identify the genre of the text (e.g. personal narratives, business negotiation, etc.)				
• Other (please specify)				

20. Think about how you prepare students for the grammar section of the G-TELP. Approximately how often do you incorporate the following tasks into your instruction?
(1= Never, 2=Rarely, 3=Some, 4=Often)

Skills	1 Never	2 Rarely	3 Some	4 Often
• Explain grammar points				
• Practice grammar questions				
• Practice formulas for structuring sentences				
• Practice a range of diverse grammatical structures				
• Memorize grammatical rules that frequently appear on the G-TELP				
• Other (please specify)				

21. Please indicate your level of agreement with the following statement:
(1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree)

Statements	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
• The listening section on the G-TELP allows students to show how well they can listen in English.				
• The reading & vocabulary section on the G-TELP allows students to show how well they can read in English.				
• The reading & vocabulary section on the G-TELP allows students to show their English vocabulary knowledge.				
• The grammar section on the G-TELP allows students to show their English grammar knowledge.				
• The grammar section on the G-TELP allows students to show how well they can speak in English.				
• The grammar section on the G-TELP allows students to show how well they can write in English.				
• The grammar section on the G-TELP allows students to show how well they can use English in real-life situation				
• It is important to include a speaking section on a test of English as a foreign language.				
• It is important to include a writing section on a test of English as a foreign language.				

22. Please indicate your level of agreement with the following statements.
(1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree)

Statements	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
• I understand what the G-TELP scores mean.				
• ITSC and G-TELP adequately disseminates information about changes to the G-TELP.				
• ITSC and G-TELP adequately disseminates information about the meaning of G-TELP scores.				
• Preparing to take the G-TELP prepares students for life at an English-speaking university.				
• Preparing to take the G-TELP prepares students for professional life at a workplace or government agencies.				
• Users of G-TELP scores (administrators or employers – including public agencies, private sectors, government institutions) understand how to use G-TELP scores.				
• Users of G-TELP scores look at subscores as well as total scores.				
• The G-TELP is an accurate predictor of how well a non-native English speaker will perform in an English-speaking context.				

23. Is the G-TELP appropriate to students' future English language needs: (Yes/No/Don't know)
 • At the pre-university level (below high school level)
 • At the undergraduate level
 • At the graduate level
 • For vocational studies
 • For job application
 • For promotion

24. What would you like to change in G-TELP? (e.g. test registration process, test content, score report formats, or any other aspects)
_____

25. Please share any of your opinions about English proficiency test.
_____

Thank you.

Appendix C.

English Proficiency Test Score User Survey

1. In which type of institution do you currently work?
• trade or vocational school
• community or junior college
• private college or university (4-year)
• public college or university (4-year)
• a private company
• a governmental institution
• Other (please specify)

2. In which city is your institution located? _____

3.1. School
Approximately how many people does your institution graduate each academic year?
• 100 students or fewer
• 101-500 students
• 501–1000 students
• 1001–1,500 students
• 1,501 students or more
• Don't know

3.2. Company or governmental institutions
Approximately how many people does your institution hire each year?
• 100 people or fewer
• 101-200 people
• 201–500 people
• 501–1,000 people
• 1,001 people or more
• Don't know

4. What is your current job title? (open-ended response) _____

5. How long have you worked in your current position?
• 1year or less
• 2 years
• 3 years
• 4 years
• 5 years or more

6. Are you involved in setting language requirements for graduation or job application?
Yes → 6.1
No → 6.2

6.1. How long have you worked with tasks related to setting language requirements?
• 1 year or less
• 2–3 years
• 4–6 years
• 7 years or more

6.2. Please describe the process of setting language requirement as much as you know

_____

7. As part of your position, which of the following activities do you perform? (Select all that apply.)
• Recruit students or employees
• Review applications
• Answer questions from applicants
• Place admitted/hired people
• Build people's English language skills
• Give input on policies for language requirements
• Other (please specify)

Please answer the following questions about admissions requirements for your institution.
8. Which of the following tests does your institution accept to meet entrance/graduation/hiring requirements for English language proficiency? (Select all that apply.)
• TOEIC (Test of English for International Communication)
• IELTS (International English Language Testing System)
• TOEFL iBT (Test of English as a Foreign Language)
• TEPS - The Test of English Proficiency developed by Seoul National University
• None of the above
• Others (please specify)

9. Why did your institution decided to accept the G-TELP scores? (Select all that apply.)
• It is a good measure of English proficiency.
• There are many applicants who take the test.
• Other institutions also decided to accept the score.
• Upper institutions or authorities recommended the test.
• Other (please specify)

9-1. If your institution does not accept G-TLEP scores, please describe why.

_____

10. How does your institution use G-TELP scores? (Select all that apply.)
• Entrance requirement
• Student placement
• Graduation requirement
• Job application requirement
• Promotion
• Other (please specify)

11. What is the minimum total score on the G-TELP that is required to enter your institution or satisfy requirements set for your purpose?
• 150 or less (below 50%)
• 150–180 (50-60%)
• 180–210 (60-70%)
• 210–240 (70-80%)
• 240–270 (80-90%)
• 270 or greater
• No minimum score required
• Other (please specify)

If are not familiar with the G-TELP, respond to the question 12-1.
12. Please indicate your level of agreement with the following statements.
(1= Strongly disagree, 2= Disagree, 3= Agree, 4= Strongly agree)

Statements	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
• I am familiar with the content of the G-TELP.				
• The content of the G-TELP reflects what students/employees need to be able to do at workplaces.				
• I understand how to interpret G-TELP scores.				
• I look at both composite and subscores on the G-TELP.				
• The test publisher disseminates adequate information about the meaning of G-TELP scores.				
• The G-TELP is a good predictor of how well people will perform at their position or at my institution.				
• I am confident using G-TELP scores to make decisions.				

12-1. The following is an example item of the Grammar section of the G-TELP. Based on English language ability required for your workplace, indicate your opinions about difficulty level.

- Extremely Easy (1)
- Somewhat Easy (2)
- Relevant (3)
- Somewhat Difficult (4)
- Extremely Difficult (5)

My friend has asked me four times if I am going to her wedding, and I keep answering "Yes". For almost a year, she has been preparing for the big event _____ at the end of this year.

 (a) which it will be held
(b) what will be held
(c) that will be held
(d) who will be held

12-2. The following is the text types on the Reading/Vocab section of the G-TELP. Based on English language ability required for your workplace, indicate your opinions about usefulness.

Statements	1 Very Unnecessary	2 Somewhat Unnecessary	3 Somewhat Necessary	4 Very Necessary
Biography				
Magazine article				
Encyclopedia article				
Business letter				

12-3. The following is the text types on the Listening section of the G-TELP. Based on English language ability required for your workplace, indicate your opinions about usefulness.

Statements	1 Very Unnecessary	2 Somewhat Unnecessary	3 Somewhat Necessary	4 Very Necessary
Conversation				
Presentation				
Explanation of Process				

13. Indicate your level of agreement with the following statements. (Strongly agree, Agree, Disagree, Strongly disagree, Don't know)

Statements	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
• I am familiar with how people study English to prepare for graduation or job application.				
• I am familiar with the way people prepare for the G-TELP.				
• Students who have taken the G-TELP have better language skills than those who have not.				
• Candidates need to prepare for the G-TELP using materials designed specifically for the G-TELP.				
• Candidates have a better chance of getting a good score on the G-TELP if they attend a preparation course.				

14. The G-TELP is a good measure of English language proficiency: (Select all that apply.)
 • At the pre-university level (high school and below)
 • At the undergraduate level
 • At the graduate level
 • For vocational studies
 • For performance evaluation at work
 • For hire or promotion

15. Do you agree or disagree?
The G-TELP is a good measure of English language proficiency
• Yes
• No

Explain why you agree or disagree (open-ended response)

_____

16. Additional Comments (open-ended response)

_____

Thank you.

Appendix D.

Guiding Questions for Individual Interviews

* This is not an exhaustive list.

Test takers
  • Are you a student or are you employed?
  • Have you taken the G-TELP Level 2?
  • When did you take it? Why did you take it?
  • Do you have a copy of your score report? What was your score (range)?
  • If you don't have to talk about your score, but do you think the score you got shows how well you know English?
  • Have you taken a test preparation course? What preparation courses have you taken?
  • What do you do in a typical G-TELP preparation class?
  • How well do you think these classes prepared you to take the G-TELP?
  • How well do you think these classes have prepared you for your future plans?
  • How difficult or easy was the G-TELP for you?
  • Did you feel prepared to do well on the G-TELP?

Test Preparation Course Instructors
  • Have you taken the G-TELP Level 2?
  • How long have you been teaching G-TELP preparation courses? What other courses are you taking?
  • How do you prepare students for the G-TELP? Can you give me an example?
  • What do you think is the most important thing you can do to prepare students for the G-TELP? Can you describe this?

Score Users
  • Have you taken the G-TELP Level 2? How aware are you and your colleague of the process and content of the G-TELP?
  • Do you consider the G-TELP a useful indicator of English proficiency appropriate to work in your institution/company?
  • How does your institution/company set standards or requirement for English language proficiency?
  • Does job applicants' English language proficiency play a critical role in decision-making process? To what degree?
  • How does your institution/company respond to any additional identified English language needs for completing required tasks?