

TOEFL[®] Research Report

TOEFL-RR-84

ETS RR-18-44

Examining the Applications and Opinions of the *TOEFL ITP[®]* Assessment Series Test Scores in Three Countries

Juliya Golubovich

Florencia Tolentino

Spiros Papageorgiou

December 2018

The *TOEFL*[®] test is the world's most widely respected English language assessment, used for admissions purposes in more than 130 countries including Australia, Canada, New Zealand, the United Kingdom, and the United States. Since its initial launch in 1964, the TOEFL test has undergone several major revisions motivated by advances in theories of language ability and changes in English teaching practices. The most recent revision, the *TOEFL iBT*[®] test, contains a number of innovative design features, including integrated tasks that engage multiple skills to simulate language use in academic settings and test materials that reflect the reading, listening, speaking, and writing demands of real-world academic environments. In addition to the TOEFL iBT, the TOEFL Family of Assessments has expanded to provide high-quality English proficiency assessments for a variety of academic uses and contexts. The TOEFL Young Students Series (YSS) features the *TOEFL Primary*[™] and *TOEFL Junior*[®] tests, designed to help teachers and learners of English in school settings. The *TOEFL ITP*[®] Assessment Series offers colleges, universities, and others an affordable test for placement and progress monitoring within English programs.

Since the 1970s, the TOEFL tests have had a rigorous, productive, and far-ranging research program. ETS has made the establishment of a strong research base a consistent feature of the development and evolution of the TOEFL tests, because only through a rigorous program of research can a testing company demonstrate its forward-looking vision and substantiate claims about what test takers know or can do based on their test scores. In addition to the 20-30 TOEFL-related research projects conducted by ETS Research & Development staff each year, the TOEFL Committee of Examiners (COE), composed of distinguished language-learning and testing experts from the academic community, funds an annual program of research supporting the TOEFL family of assessments, including projects carried out by external researchers from all over the world.

To date, hundreds of studies on the TOEFL tests have been published in refereed academic journals and books. In addition, more than 300 peer-reviewed reports about TOEFL research have been published by ETS. These publications have appeared in several different series historically: TOEFL Monographs, TOEFL Technical Reports, TOEFL iBT Research Reports, and TOEFL Junior Research Reports. It is the purpose of the current TOEFL Research Report Series to serve as the primary venue for all ETS publications on research conducted in relation to all members of the TOEFL Family of Assessments.

Current (2018–2019) members of the TOEFL COE are:

Lia Plakans – Chair

Ayşegül Daloğlu
April Ginther
Luke Harding
Claudia Harsch
Lianzhen He
Volker Hegelheimer
Lorena Llosa
Carmen Muñoz
Yasuyo Sawaki
Randy Thrasher
Dina Tsagari

The University of Iowa

Middle East Technical University (METU)
Purdue University
Lancaster University
University of Bremen
Zhejiang University
Iowa State University
New York University
The University of Barcelona
Waseda University
International Christian University
Oslo Metropolitan University

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org Web site: www.ets.org/toefl



ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

RESEARCH REPORT

Examining the Applications and Opinions of the *TOEFL ITP*[®] Assessment Series Test Scores in Three Countries

Juliya Golubovich, Florencia Tolentino, & Spiros Papageorgiou

Educational Testing Service, Princeton, NJ

In this study, 249 users of the *TOEFL ITP*[®] assessment series (e.g., admissions officers, English-language teachers, academic staff) in Japan, Mexico, and Indonesia were surveyed about their uses and opinions of TOEFL ITP scores, followed by in-depth interviews with 21 of these users. Overall, the most common use of the test was as an exit requirement from English-language programs to demonstrate proficiency in English listening and reading. The majority of participants saw TOEFL ITP scores as very useful indicators of students' English-language proficiency. Interviews helped clarify the user needs met by this assessment and how test scores were actually applied. Study participants indicated that they need a relatively inexpensive and practical English-language assessment that also provides them with enough information to make decisions about test takers' proficiency in all relevant skill areas. Some of the ways interviewees talked about using TOEFL ITP scores were consistent with the recommendations of Educational Testing Service (ETS), while other uses (e.g., for workplace applications) were more questionable, as they might imply potentially higher stakes than were intended for this test. The results of the study highlight areas where TOEFL ITP users might need additional informational support with regard to score interpretation and use.

Keywords the *TOEFL ITP*[®] assessment series; English-language assessment; institutional tests; user perceptions; assessment literacy

doi:10.1002/ets2.12231

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), developers of educational and psychological assessments should provide evidence of an assessment's validity for its intended uses. Inherent in this standard is the notion that validity is a property of the ways in which assessment scores are interpreted and used, not a property of the assessment itself. This means that even when test developers provide end users with adequate evidence to support intended score-based inferences, misinterpretation or misuse of assessment scores can subsequently undermine those scores. Examining the extent to which users utilize scores for the recommended applications and how they actually make decisions based on scores can further developers' understanding of the consequences of assessment use and help identify ways to better inform assessment users. Notably, users may vary in what they know and understand about an assessment; how to apply it; and the impact of its use on individuals, institutions, and society more broadly (referred to as *assessment literacy*; Baker, 2016). Users with low assessment literacy can especially benefit from further support.

In addition to investigating users' understanding and applications of assessment scores, it is also valuable to examine their perceptions of an assessment's utility, including the extent to which it appears to nonexperts in the content area to assess what it is said to assess. An assessment has to be perceived as valid for its intended users to be willing to apply it for decision-making. User opinions are relevant to the social consequences of testing (Fulcher, 1997; Ginther & Elder, 2014). If an assessment is not perceived as useful, its developers might not be successful in bringing about the intended positive consequences as a result of the use of the scores.

The *TOEFL ITP*[®] Assessment Series

In the current study, we investigate users' applications of the *TOEFL ITP*[®] assessment series and their related perceptions of these tests. The *TOEFL ITP* assessment series¹ are tests administered by institutions such as colleges, universities, and English-language programs as well as by an Educational Testing Service (ETS) Preferred Network office for internal

Corresponding author: S. Papageorgiou, E-mail: spapageorgiou@ets.org

purposes (e.g., placement of students into appropriate levels of English-language classes). As of July 2017, TOEFL ITP tests were reportedly used by more than 2,500 institutions in more than 50 countries and were taken by more than 800,000 test takers.²

Institutions can administer the TOEFL ITP on their own, using their own facilities, as frequently as they desire. There are two levels of the test. Level 2 is intended for students having beginning to intermediate English-language skills, and Level 1 is for those having intermediate to advanced skills. Both levels of the TOEFL ITP contain three sections: (a) listening comprehension, (b) structure and written expression, and (c) reading comprehension. The listening comprehension section (30 items for Level 2 and 50 items for Level 1) measures the ability to understand English used in short and long conversations and in short talks or lectures. The structure and written expression section (25 items for Level 2 and 40 items for Level 1) measures the ability to recognize the usage of standard written English. The reading comprehension section (40 items for Level 2 and 50 items for Level 1) tests the ability to read and understand short academic passages written in English.³ Total scores range from 310 to 677 on the Level 1 test and from 200 to 500 on the Level 2 test. For simplicity, we refer to the test as the TOEFL ITP in our study, regardless of test level.

Results of institutional tests like the TOEFL ITP are typically not used outside of the institution administering the test. This is in contrast to tests like the *TOEFL iBT*® test (Internet based) and the International English Language Testing System (IELTS), which assess not just receptive but also productive English-language skills (writing and speaking; implying broader construct coverage) and are intended to inform high-stakes decisions (e.g., student admission for degrees in higher education where English is the medium of instruction). Seven claims about the intended uses of the TOEFL ITP are currently presented on the ETS website⁴:

- placement in intensive English-language programs requiring academic English proficiency at a college or graduate level;
- progress monitoring in English-language programs stressing academic English proficiency;
- exiting English-language programs by demonstrating proficiency in English listening and reading;
- admissions to short-term, nondegree programs in English-speaking countries where the sending and receiving institutions agree to use TOEFL ITP scores;
- admissions to undergraduate and graduate degree programs in non-English-speaking countries where English is not the dominant form of instruction;
- admissions and placement in collaborative international degree programs where English-language training will be a feature of the program; and
- scholarship programs, as contributing documentation for academic English proficiency.

Investigating Applications and Perceptions of TOEFL ITP Scores

There has been limited research to date examining institutions' actual uses of TOEFL ITP scores and perceptions of this assessment (though case studies of particular uses have been reported, e.g., Choi & Papageorgiou, 2014; Minton & Nishikawa, 2007). On the other hand, there is an accumulating body of research on the use of TOEFL iBT and IELTS scores (e.g., Ginther & Elder, 2014; Hyatt & Brooks, 2009; Malone & Montee, 2014; O'Loughlin, 2013; Stricker & Attali, 2010; Stricker & Wilder, 2012). Relative to assessments like TOEFL ITP, these tests have more involved development, administrative, and scoring procedures and are typically available at a higher price point. Findings regarding user experiences with these assessments that aid in making higher stakes decisions may not generalize well to user experiences with the TOEFL ITP or other institutional, lower stakes assessments (e.g., the Cambridge Michigan Language Assessment [CaMLA] English Placement Test and the Michigan Test of English Language Proficiency Series by CaMLA), which have a different set of intended uses.

Applications of TOEFL ITP Scores

It is important to examine the extent to which institutions are using TOEFL ITP scores for the applications recommended on ETS's website, how users are making decisions based on test scores when applying the assessment for a particular use, and which pieces of information about TOEFL ITP users consider. The popularity of various TOEFL ITP uses has implications for the support needs that TOEFL ITP users have. The questions of how users apply TOEFL ITP scores and use

the test's supporting informational materials also speak to users' TOEFL ITP-related assessment literacy (Inbar-Lourie, 2013). Certain applications of TOEFL ITP scores require that institutions determine where to set one or more cut scores. For example, if using the test to place students into English-language programs, institutions need to determine where on the score scale to set minimums that students have to reach to qualify for progressively higher levels in the program. Setting cut scores is left up to institutions, but these are complex decisions that present a variety of methodological options and require the involvement of qualified individuals (e.g., with relevant subject knowledge and/or with knowledge of measurement; Zieky & Perie, 2006). These are also socially consequential decisions, even in relatively lower stakes settings. When cut scores are set too high, students who should pass based on their true level of English-language skill fail; when cut scores are set too low, students who should fail actually pass (Zieky & Perie, 2006). Among these misclassified students, those who did not meet the cut (false negatives) may enter lower level programs for which they are overqualified and may not enjoy the benefits associated with passing; those who passed (false positives) may enter programs or receive benefits for which they are underqualified.⁵ The social consequences of test use are particularly concerning if a test is being misused due to low levels of assessment literacy among institutional users (Shohamy, 2001). A common finding noted by O'Loughlin (2013) is that institutional staff are not very knowledgeable about the meaning of English-language test scores.

Having information about how institutions are interpreting and applying TOEFL ITP scores (including where they are setting cut scores) would help to identify opportunities for supplying users with better or more useful information about the assessment. It would also allow gauging test users' actual understanding of appropriate test use so as to correct misconceptions if needed. To build on the earlier example, those who use the TOEFL ITP for placement or as an exit requirement within their own institutions probably need more support around how to set cut scores at appropriate levels than do those who help students qualify for external scholarships or programs that require a TOEFL ITP score (e.g., by placing them in English-language classes). In the latter cases, other institutions determine where to set cut scores. For those who support students seeking to attend programs at other institutions or to receive a scholarship, their support needs possibly revolve around having a wide enough variety of training materials and ensuring that these materials are adequately accessible to students.

Perceptions of TOEFL ITP Scores

As mentioned earlier, users are unlikely to apply an assessment that they do not perceive to be valid for its stated purpose(s). Therefore it is important to investigate users' perceptions of the TOEFL ITP's validity for the uses that ETS recommends.

Current Study

In the current study, we posed a set of research questions (RQs) to investigate how TOEFL ITP scores are applied, how users make decisions based on these scores, and the TOEFL ITP's perceived level of validity (RQs 1–3). Additionally, we tried to identify opportunities to better support test users (RQs 4–5) following recommendations in the field that test developers should offer users additional support when unmet needs are identified (Ginther & Elder, 2014):

RQ 1: How do institutions actually use (or misuse) TOEFL ITP scores?

RQ 2: How do users make decisions based on TOEFL ITP scores? How do users set cut scores?

RQ 3: Do users perceive TOEFL ITP to be valid for the uses outlined by ETS? What needs are met (or not met) by this test?

RQ 4: What available information about TOEFL ITP do users take into consideration when using the test, and how useful do users find this information to be?

RQ 5: What additional information needs do users have beyond the information provided by the TOEFL ITP program?

We sampled institutional TOEFL ITP users from Japan, Mexico, and Indonesia, as these are among the top countries in recent years in terms of test-taker volume for this assessment. We began with a user survey and subsequently conducted in-depth interviews with willing survey respondents. Survey data were intended to provide an overall summary of TOEFL

Table 1 Survey Respondents' Type of Institution, Primary Role, and TOEFL ITP-Related Activities

Response category	Percentage			
	Japan <i>n</i> = 65	Mexico <i>n</i> = 153	Indonesia <i>n</i> = 31	Total <i>N</i> = 249
Institution type				
Private college or university	66.2	17.6	32.3	32.1
Public college or university	20.0	30.1	35.5	28.1
Secondary school	3.1	3.9	0.0	3.2
Trade or vocational school	0.0	5.2	0.0	3.2
Other	10.8	43.1	32.3	33.3
Primary role				
Faculty/department administration	6.2	28.8	3.2	19.7
English-language teacher	16.9	20.3	9.7	18.1
Student support services	41.5	3.9	32.3	17.3
Academic teacher or researcher	23.1	9.8	19.4	14.5
University central administration	3.1	7.2	3.2	5.6
Other	10.8	30.1	41.9	26.5
Activities performed				
Build students' English-language skills	49.2	78.4	87.1	71.9
Place students into English-language classes	30.8	78.4	48.4	62.2
Advise admitted students	40.0	50.3	45.2	47.0
Give input on admissions policies	10.8	29.4	51.6	27.3
Answer admissions questions	10.8	31.4	35.5	26.5
Make decisions about admissions policies	13.8	28.8	35.5	25.7
Review applications	1.5	30.7	25.8	22.5
Other	30.8	17.6	9.7	20.1

Note. The primary role percentages add up to more than 100% because several respondents indicated more than one primary role.

ITP users' practices, opinions, and needs; the interviews were used to better understand the survey findings and address our questions more fully.

In the sections that follow, we discuss our study methodology. Subsequently, we present results of the survey and interviews and discuss the implications of current findings for understanding ways to better support institutional TOEFL ITP users. We also discuss limitations of the current work and areas for future research.

Method

Participants

Survey Participants

Survey participants were recruited and surveyed by ETS representatives in Japan (Council on International Educational Exchange), Mexico (Institute of International Education), and Indonesia (Indonesian International Education Foundation). A total of 249 individuals responded to the survey and indicated using the TOEFL ITP at the time.⁶ Of this sample, 65 were from Japan, 153 were from Mexico, and 31 were from Indonesia. Participants' background information in terms of institution type, role, and job activities, broken out by respondents' country, is summarized in Table 1. A majority of respondents were from a college or university (60.2%) and were working to build students' English-language skills (71.9%). Participants' average job tenure was 8.31 years ($SD = 7.29$ years). Among Mexican and Indonesian participants, 87.9% were "very confident" about the information they provided on the survey; 10.9% were "somewhat confident," and only 1.1% were "not at all confident." Japanese respondents, who received the survey first, were not asked this question.

Interview Participants

We contacted all the individuals in Japan and Indonesia who indicated on the survey that they would be willing to participate in an interview. Among survey respondents in Mexico, 71 said they would be willing to be interviewed. As this was not a manageable number of interviews to conduct, we later contacted only 18 of these people to try to set up an interview. We chose Mexican respondents to contact in a way that would representatively sample different types of institutions,

Table 2 Interviewees' Type of Institution, Primary Role, and TOEFL ITP-Related Activities

Response category	Percentage
Institution type	
Public college or university	26.1
Private college or university	26.1
Secondary school	4.3
Trade or vocational school	0.0
Other	43.5
Primary role	
English-language teacher	21.7
Faculty/department administration	17.4
Academic teacher or researcher	13.0
Student support services	8.7
University central administration	4.3
Other	47.8
Activities performed	
Build students' English-language skills	82.6
Advise admitted students	78.3
Place students into English-language classes	73.9
Answer admissions questions	60.9
Give input on admissions policies	60.9
Review applications	39.1
Make decisions about admissions policies	39.1
Other	21.7

Note. $N = 23$. The primary role percentages add up to more than 100% because several respondents indicated more than one primary role.

roles, and test uses (based on the frequencies of different institution types, roles, and test uses reported by survey respondents in Mexico). We also tried to take into consideration survey responses regarding the usefulness of TOEFL ITP scores in order to interview individuals who felt differently about the test's utility as an assessment of English-language skill.

A total of 21 individuals (three in Japan, 11 in Mexico, and seven in Indonesia) ultimately participated in a follow-up interview. Additionally, two individuals in Indonesia answered interview questions via e-mail because of poor audio quality during their scheduled telephone interviews. The institutions, roles, and activities for these 23 individuals (based on survey responses) are summarized in Table 2. The largest percentage was from other institutions (43.5%; e.g., language school) and performed other primary roles (47.8%; e.g., principal, manager, or director). Their average job tenure was 7.90 years ($SD = 6.59$ years). We coded respondents' institutions as either general education (including middle school, high school, undergraduate, or graduate; 10 respondents) or existing to provide language training or test preparation (10 respondents). Three transcripts were difficult to code for institution because these respondents were not asked a specific question about their institutions. We refer to interview participants using a letter and participant number (e.g., J1 to denote the first interview participant in Japan).

Measures

Survey Instrument

The survey consisted of a combination of multiple-choice and open-ended questions based in part on earlier studies investigating the use of test scores (e.g., Hyatt & Brooks, 2009; Malone & Montee, 2014; O'Loughlin, 2013). There were several differences between the Japanese and other two survey versions because data collection plans for Mexico and Indonesia were finalized several months after those for Japan, providing for an opportunity to revisit and tweak survey content. Minor changes were made to the survey to capture several additional important pieces of information (described later in this section). The first part of the survey asked about participants' institution, role, job tenure, and activities that pertain to assessing students' English-language skills, and current use of the TOEFL ITP assessment. It was not practical to identify, in advance, individuals who employ the test for a particular use. Thus, to target survey questions toward gathering information about a particular use, we asked respondents to indicate the test use with which they were most familiar at their institution and to respond to subsequent questions about the test with reference to that use. This way, we

hoped to capture user opinions of TOEFL ITP scores for each of the seven recommended uses. Notably, the Indonesian survey included an additional use (not one specifically listed on ETS's website) as a response option. At the Indonesian representative's request, we added the use "as one of the requirements for graduation from undergraduate and/or graduate programs in universities in a non-English-speaking country." We decided that it was appropriate to add this use because it was similar to the exit use recommended by ETS and, at the same time, more likely to be understood by participants in that context.

Participants (except those in Japan) were asked to indicate whether they used any other tests/information for their selected use of the TOEFL ITP assessment. They were asked about minimum section and total TOEFL ITP scores for the use they selected, if applicable. They then indicated their opinions of the minimum scores (if applicable) and TOEFL ITP scores' general usefulness as an indicator of students' English-language proficiency. Additionally, they were asked about the usefulness of various pieces of information pertaining to the test (e.g., how test takers can prepare for the TOEFL ITP, how long TOEFL ITP test scores are valid) and whether there is any other information about the test that the participant would find useful. These information-related questions were not asked with reference to any specific use of the TOEFL ITP but with reference to the potentially multiple ways in which individuals' institutions may be using the TOEFL ITP. Participants (except those in Japan) were also asked about their levels of confidence in their survey responses. The last section of the survey asked respondents if they would be willing to participate in a follow-up interview and requested their contact information. The survey was not anonymous, but confidentiality was assured. The survey version used in Indonesia is included in Appendix A. We note places where content was slightly different for the version(s) of the survey used in Japan and/or Mexico.

Because we were targeting contexts where English was not the individuals' native language, an experienced translation company was hired to translate the survey into Japanese, Indonesian (Bahasa), and Spanish. We then asked ETS representatives in the three countries to review and revise the translations as needed (with the original, English survey in mind) to ensure the survey was clear and accessible for participants in their native languages but that questions still had the intended meaning. The second author, a native Spanish speaker, served as an additional reviewer of the Spanish translation of the survey.

Interview Questions

Interview questions were prepared for each participant based on his or her survey responses as well as our RQs. However, there was a lot of overlap in the questions across participants. Approximately 15 questions were planned per interviewee to balance the need for detailed and high-quality information with the desire not to overburden participants. Interviews were meant to be semistructured in that there was a set of questions planned but follow-up questions were determined based on information interviewees provided during the course of the interview, so as to better understand their particular experiences with the assessment. A sample set of interview questions is included in Appendix B.

Procedure

Survey Procedure

Survey data were collected in Japan between October 2015 and January 2016, in Mexico between April and May 2016, and in Indonesia between April and June 2016. The survey was administered using an online platform for Mexico; a paper-and-pencil survey was used for Japan and Indonesia. The medium of administration was chosen by ETS's representatives considering accessibility issues in their locations.

Some participants provided survey comments in their native languages. The second author translated the Spanish comments into English. Comments provided in Japanese and Bahasa were translated into English by the same translation company that handled survey translation for these languages. As a token of appreciation, all participants who provided their contact information were informed about the results of the survey via a summary report.

Interview Procedure

Interviews were conducted between May and August 2016. The first and second authors conducted one-on-one interviews through conference calls. Most of the interviews were conducted in English. A Japanese–English interpreter (provided

by the translation company used for this study) was employed for two of the interviews with Japanese participants. In the case of Mexico, three interviews were conducted in Spanish by the second author.

Transcription of Interviews

Of the 21 interviews completed, eight were transcribed by the first and second authors. The three interviews conducted in Spanish were transcribed in Spanish and then translated by the second author into English. The remaining 13 interviews were transcribed by two transcription companies. They followed the *Chicago Manual of Style* or APA format and did full analysis and research of names, spellings, and technical terms to avoid errors. The first author reviewed the interview transcripts against the audio files to check quality and fill in words that were unnecessarily marked as inaudible. In quoting interview participants, we do not correct their responses for grammar to represent their responses with maximum fidelity.

Coding of Interviews

Interview data were coded using NVivo 11. At the broad level, the codes were defined according to the interview questions that were posed to participants (e.g., how the institution uses TOEFL ITP scores, utility of TOEFL ITP scores, other tests the institution uses, usefulness of different types of information available about TOEFL ITP). The first author read through each interview transcript one by one and assigned to various segments the applicable broad codes. Whenever the initial list of broad codes did not include a code covering a particular interview segment, the author generated an additional broad code to represent that content. After initially coding the interviews using broad codes, the author revisited the interview segments that had been assigned a particular broad code and generated more detailed subcodes to more fully represent the content of those interview segments.

After the first author finished coding the interview transcripts, the three authors met to discuss the coding of one randomly selected transcript (I1). We reached consensus on the list of codes and their interpretation relative to the content of the transcript. Afterward, the second author reviewed all the coding completed by the first author with the goal of reaching 100% consensus on how the 23 transcripts should be coded. Any disagreements were discussed and resolved. Although independent coding of transcripts by the authors followed by assessment of intercoder agreement and reconciliation of disagreement may be viewed as a more rigorous approach, there is disagreement among qualitative researchers about the value of double-coding data and checking for interrater agreement (de Wet & Erasmus, 2005, provided a discussion). Furthermore, it is advisable for researchers to choose a qualitative analysis strategy that is feasible given their study's goals and resource constraints (Forman & Damschroder, 2008). We opted for the approach described rather than independent coding of the transcripts because of the complexity of the interview data and the associated coding scheme. There was a long list of codes that included multiple hierarchical levels, and many of the codes were interrelated. Therefore it would have been possible for a coder to overlook a particular code that could be applied to a certain interview segment. Our approach of having one person serve as a main coder and a second person double-check the coding allowed the second coder to better focus on what the first coder may have missed or misinterpreted, ensuring more thorough coding of the data (see Bandara, Indulska, Chong, & Sadiq, 2007; Kim, Addom, & Stanton, 2011; Silumbe et al., 2015, for similar approaches). We believe that the analysis approach chosen is appropriate for meeting our goal of highlighting users' various TOEFL ITP applications, opinions, and needs.

Results

In the sections that follow, we present study findings organized by RQ. Unless indicated otherwise, survey results (i.e., percentages) are reported with reference to the total number of individuals who responded to a particular question, which may be fewer than the total number of survey participants; that is, individuals who skipped a particular question are not included in the analyses. We report results at the overall and country levels (where possible). Institutions' practices and needs potentially vary across countries, as Japan, Mexico, and Indonesia differ in their educational, cultural, and economic backdrops for language testing. It is common to closely consider context when examining testing practices or test-related attitudes (e.g., Ryan et al., 2017; Steiner & Gilliland, 1996; Stricker & Attali, 2010). However, as explained later, we did not think it appropriate in this case to conduct statistical comparisons of results across countries.

Table 3 Specific Use of TOEFL ITP With Which Survey Respondents Said They Were Most Familiar

Use	Percentage			
	Japan <i>n</i> = 65	Mexico <i>n</i> = 151	Indonesia <i>n</i> = 30	Total <i>N</i> = 246
To place students in intensive English-language programs requiring academic English proficiency at a college or graduate level	18.5	9.3	3.3	11.0
To monitor students' progress in English-language programs stressing academic English proficiency	20.0	17.2	26.7	19.1
As an exit requirement from English-language programs to demonstrate proficiency in English listening and reading	10.8	34.4	16.7	26.0
To admit students to short-term, nondegree programs in English-speaking countries where the sending and receiving institutions agree to use TOEFL ITP scores	50.8	4.0	6.7	16.7
To admit students to undergraduate degree programs in non-English-speaking countries where English is not the dominant form of instruction	0.0	4.0	6.7	3.3
To admit students to graduate degree programs in non-English-speaking countries where English is not the dominant form of instruction	0.0	9.3	10.0	6.9
To admit and place students in collaborative international degree programs where English-language training will be a feature of the program	0.0	1.3	3.3	1.2
For scholarship programs, as contributing documentation for academic English proficiency	7.7	7.3	53.3	13.0
As one of the requirements for graduation from undergraduate and/or graduate programs in universities in a non-English-speaking country	– ^a	– ^a	20.0	– ^a
Other purposes	12.3	13.2	10.0	12.6

Note. Percentages for Japan and Indonesia add up to more than 100% because, given a paper-and-pencil survey, respondents in these countries were able to check off more than one use of the test.

^aThe corresponding use was only listed in the survey for Indonesia.

Interview findings are used to further elaborate on the results of the survey. Although we present information about the number of users who made a particular type of comment, these frequencies should not be interpreted as indicating relative importance of concerns. Rather, frequencies represent the amount of evidence we were able to gather to help us understand respondents' experiences or views. The types of comments participants made were partly a function of the set of initial and follow-up questions they were asked. As discussed earlier, questions varied somewhat across participants.

Research Question 1

Survey Results

Participants' responses about their main TOEFL ITP uses are summarized in Table 3. Participants in Mexico were most familiar with using TOEFL ITP as an exit requirement from English-language programs to demonstrate proficiency in English listening and reading (34.4%). Participants in Japan most commonly indicated using TOEFL ITP to admit students to short-term, nondegree programs in English-speaking countries where the sending and receiving institutions agreed to use TOEFL ITP scores (50.8%). Participants in Indonesia most often indicated using TOEFL ITP for scholarship programs, as contributing documentation for academic English proficiency (53.3%).

Interview Results

The interview data pointed to a number of questionable TOEFL ITP uses that were not captured during the survey. More than half of respondents (*n* = 14) mentioned applying the test for some type of workplace use, such as for job applications, selection of teachers (of which subject was not clear), and placement within the workplace. This suggests that TOEFL ITP

Table 4 Survey Respondents' Reported Use of Minimum TOEFL ITP Scores

Response	Percentage			
	Japan <i>n</i> = 64	Mexico <i>n</i> = 149	Indonesia <i>n</i> = 30	Total <i>N</i> = 243
Use a listening comprehension minimum	6.7	21.5	20.0	16.5
Use a structure and written expression minimum	6.7	22.1	20.0	16.9
Use a reading comprehension minimum	6.7	23.5	20.0	17.7
Use a total score minimum	103.3	75.2	70.0	67.5
Use no absolute minimum scores	110.0	26.8	36.7	34.6
Do not know	– ^a	0.7	0.0	0.4

^aThe corresponding option was not provided in the survey for Japan.

Table 5 TOEFL ITP Total Score Minimums in Use as Specified by Survey Respondents and Subsequently Aligned to Common European Framework of Reference for Languages Levels

Range	Percentage			
	Japan <i>n</i> = 28	Mexico <i>n</i> = 86	Indonesia <i>n</i> = 19	Total <i>N</i> = 133
337–459 (A2)	46.4	26.7	47.4	33.8
Ranging from A2 to B1	0.0	2.3	5.3	2.3
Ranging from A2 to B2	0.0	2.3	0.0	1.5
Ranging from A2 to C1	0.0	1.2	0.0	0.8
460–542 (B1)	46.4	31.4	36.8	35.3
Ranging from B1 to B2	0.0	5.8	0.0	3.8
543–626 (B2)	7.1	30.2	10.5	22.6

Note. The six main levels range from A1, indicating breakthrough or beginner, to C2, indicating mastery or proficiency.

users may be applying the test for purposes that are not purely institutional. Additionally, participants mentioned using the TOEFL ITP for program evaluation (three respondents) and to admit students into a teacher training course (one respondent).

Research Question 2

Survey Results

Fewer than half of survey respondents (34.6%) indicated not using absolute minimum TOEFL ITP scores (see Table 4). Further analyses showed that it was most common for respondents to indicate not using minimum scores when applying TOEFL ITP specifically for monitoring students' progress in English-language programs: 59.5% of respondents who indicated being most familiar with that use of TOEFL ITP said they did not use absolute minimum scores.

Respondents who indicated using a total score minimum for their application of TOEFL ITP (67.2%) indicated that these minimums ranged from 350 to 677, with the majority of respondents (87.5%) providing a single total score minimum and the rest indicating a range of minimum total scores (possibly ranges were provided because the minimum score depended on the test use or situation). To aid interpretability of the minimum scores survey respondents listed, and because we expected that institutions may have considered the proficiency levels defined by the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) when setting minimum scores, we aligned their minimum scores to the CEFR levels (see Table 5).⁷ The CEFR describes communicative language ability in terms of six main levels. For each level, it describes learners' expected performance in terms of language activities (e.g., reception [reading and listening], production and interaction [speaking and writing]) and language communication competences (e.g., linguistic, sociolinguistic, and pragmatic competence). These six main levels range from A1, indicating breakthrough or beginner, to C2, indicating mastery or proficiency. Minimum total scores listed by the respondents corresponded mostly to the A2 (33.8%) and B1 (35.3%) proficiency bands.

The majority of respondents (82.3%–83.5%, depending on TOEFL ITP test section; see Table 4) indicated not using minimum section scores for the test. For those who did use section minimums, minimums were typically listed as a single

Table 6 TOEFL ITP Section Score Minimums in Use as Specified by Survey Respondents and Subsequently Aligned to Common European Framework of Reference for Languages Levels

Range	Percentage			
	Japan	Mexico	Indonesia	Total
Listening comprehension section				
Lower than 38 ^a	0.0	0.0	20.0	3.4
38–46 (A2)	50.0	27.3	20.0	27.6
Ranging from A2 to B1	0.0	4.5	0.0	3.4
47–53 (B1)	50.0	18.2	40.0	24.1
Ranging from B1 to B2	0.0	4.5	0.0	3.4
54–63 (B2)	0.0	45.5	20.0	37.9
Structure and written expression section				
Lower than 32 ^a	0.0	0.0	20.0	3.3
32–42 (A2)	50.0	4.3	0.0	6.7
Ranging from A2 to B1	0.0	4.3	0.0	3.3
43–52 (B1)	50.0	34.8	60.0	40.0
Ranging from B1 to B2	0.0	4.3	0.0	3.3
53–63 (B2)	0.0	52.2	20.0	43.3
Reading comprehension section				
Lower than 31 ^a	0.0	0.0	20.0	3.3
31–47 (A2)	50.0	21.7	0.0	20.0
Ranging from A2 to B1	0.0	4.3	0.0	3.3
48–55 (B1)	50.0	60.9	80.0	63.3
Ranging from B1 to B2	0.0	0.0	0.0	0.0
56–62 (B2)	0.0	13.0	0.0	10.0

Note. $N = 2$ for Japan, 22 (listening comprehension section) to 23 (structure and reading sections) for Mexico, 5 for Indonesia, and 29 (listening comprehension section) to 30 (structure and reading sections) in total.

^aOne respondent in Indonesia indicated minimum section scores that did not reach even the level of the lowest proficiency band (A2) in the Common European Framework of Reference for Languages. It seems inconsistent, however, with the 450 total minimum score (corresponding to an A2 level) this person listed.

score (86%–87% of respondents) versus as a range of scores and were the same across test sections. Table 6 summarizes section minimum scores. Among those who actually had any minimum score requirements (section or total) and had an opinion about their appropriateness, 70.0% believed them to be “about right.” Of the remaining respondents, 22% thought the minimum scores were “too low,” and 8% thought they were “too high.”

To find out whether decisions were based completely on TOEFL ITP scores or only partially, we asked Mexican and Indonesian survey participants about additional tests or information they considered when applying TOEFL ITP scores. Close to half (45.9%) indicated using other tests or additional information for the TOEFL ITP use(s) they selected. Additional tests or methods used mainly included other commercially available language tests (e.g., the *TOEIC*® tests, the *TOEFL*® suite, IELTS, Business Language Testing Service) and internal methods (e.g., language tests, interviews). However, it was challenging to classify a particular open-ended answer cleanly into one of the categories that we applied post hoc.

Interview Results

Interview data provided additional information regarding how users make decisions based on TOEFL ITP scores and how they set cut scores. We examined these questions within the specific uses reported by interviewees. Findings are organized into categories of approved and questionable uses.

Application of TOEFL ITP Scores for Approved Uses

When TOEFL ITP scores were used as a graduation/exit requirement, institutions tended to set minimum scores based on considerations like what major or career different students were in, what kinds of scores students would need to pursue

further studies after graduating from the institution, and what kinds of scores seemed reasonable to require based on students' abilities and institutional officials' opinions. Three individuals talked about students being required to take the test as many times as necessary to achieve an institutional minimum score requirement. However, institutions did make exceptions. One respondent from Mexico described making an exception to minimum score requirements for graduation from a doctoral program for a hardworking student who, despite what were apparently her best efforts, just could not reach the institution's minimum TOEFL ITP score of 550.

Individuals who described using TOEFL ITP in the context of student graduation/exit did not necessarily use these test scores to make graduation/exit-related decisions. This was partly a function of institution type. An individual employed at a language institute who mentioned a graduation/exit use of the test explained that the institute provided students with training to help them meet other institutions' minimum graduation/exit score requirements. A second individual employed at a high school described offering graduating students the test as a certification that they could present to meet universities' requirements; in this case, TOEFL ITP scores were treated as part of a portfolio of diplomas and certificates students compiled. Another consideration in not using TOEFL ITP scores for graduation/exit-related decisions was students' language ability and whether they could afford retesting. One individual mentioned that her institution got rid of minimum score requirements for graduation (though students still took the test upon graduation) because it was an obstacle for students to have to pay to take the test several times after failing to reach the minimum score the first time. After eliminating minimum requirements, the institution hoped that 65% or more of students could reach 450 points on the test at the time of their graduation.

When TOEFL ITP scores were used for scholarship applications for students looking to study in an English-speaking country, these scores were typically treated as an initial application requirement (unless students already had a TOEFL iBT or IELTS score), and students subsequently had to take another English-language proficiency test used for admission purposes, such as the ones just mentioned. For example, one respondent explained,

In Australian scholarships . . . they can have the TOEFL ITP score for the first selection process of the application. . . . And when they have their international English test already like iBT TOEFL test or IELTS . . . they can go directly for the next test when they have been announced to be selected for the next process. (I3, English-language teacher at an educational institution)

But when students applied for a scholarship to study in a non-English-speaking country, TOEFL ITP scores could be sufficient for showing working knowledge of English, as one individual explained:

If they're going to study in Spain, for example, in order to get a scholarship—even though they're not going to study in English, they still want the students to have a working knowledge, and so they will take the ITP. (M6, faculty/department administrator at a public college or university)

Scholarship providers differ in the minimum TOEFL ITP scores they require. One person explained it this way:

It depends very much on the country where student or lecturer want to pursue their studies. . . . If they want to study in U.S. . . . the student should have at least 550 for the Fulbright Scholarship. And some other countries let's say like Australia, the minimum requirement for the TOEFL is only 530. But again, you know, this is for . . . preliminary process. . . . The students send all application or requirements to the Australian consul in Jakarta. And then they do kind of short listing. (I4, university central administrator at a public college or university)

Interviewees did not have a lot of insight into why minimum scores were set at particular levels for scholarships. This was not surprising, because respondents were not involved in the decision-making processes of scholarship providers.

Institutional users who placed students into English preparation courses generally set multiple cut scores on the test that dictated into which level a student should be placed. The following is an illustrative quote:

We do an assessment test first to see which level they are in and then after that, we will tailor lesson for them so that it will improve their skills in ITP. We check in which area they are weak, whether they are weak in the listening, grammar, or reading. . . . We put minimum scores if they want to prepare for the TOEFL ITP because we have to

make sure that the level of student who want to take the ITP test is not a beginner. . . . So basically, if the students already know English, at least we can help them to learn about TOEFL ITP. But, if they are still a beginner, of course we cannot help them to prepare for their TOEFL ITP test. (I5, academic administrator at a language institute)

Regarding the use of TOEFL ITP scores for admission to nondegree programs in other countries or for study abroad, interviewees mentioned that scores informed external institutions about students' ability to handle either ESL classes or subject classes that would be taught in English. For example, one participant said,

Most of our students also take international programs during their high school studies and many of the universities where they go, they require for a TOEFL score no matter what country they go to—for example, I could mention international programs in France—some of the universities require—even if they are going to be in France, most of their classes are given in English. They need to have a certain level of English. That's when they require a TOEFL score—an updated TOEFL score. . . . Not all of our students are enrolled in international programs, but the ones that do take these programs, they do have to take TOEFL ITP test. (M7, faculty/department administrator at a high school)

When TOEFL ITP scores were used for admission to graduate programs in non-English-speaking countries, students had to meet a certain minimum score to be accepted or to achieve full acceptance after being conditionally accepted to the institution (i.e., to become a matriculated student). The following comment illustrates this:

Some of the [graduate] students who enter the school as a nonmatriculated student, they have to be matriculated so that's the purpose we use TOEFL ITP. . . . As a nonmatriculated student, people can participate the three course, like 9 credit, before they submit the matriculated document. And then those times the students can come to the school without reaching the 575 [the minimum score], but still can take the class. But most of them who cannot to reach the 575 by the timing of the matriculated students, they will be failed. . . . Even if we tried to admit them, to be a matriculated student, most of the students . . . will not be succeeded in the class anyway, so they will be like naturally dropped down from the class, most of the case. (J1, faculty/department administrator at a university)

Using TOEFL ITP for progress monitoring involved considering how much improvement in test scores students were able to demonstrate over time and generally did not seem to involve the use of minimum scores (except as a related end goal). For example, one participant explained,

We didn't have any entrance test yet to measure the initial TOEFL score. But maybe next year we would like to begin, to start looking at the students' baselines information about TOEFL proficiency test. So that they know from the beginning, the gap between their initial score with the target score that they have to fulfill before they graduate. (I7, employee within the language center of a public college or university)

Application of TOEFL ITP Scores for Potentially Questionable Uses

As stated earlier, the majority of interviewees mentioned applying the test for some type of workplace use. Some respondents described employers asking for TOEFL ITP scores for hiring or promotion, while in other cases, it was not clear whether employers requested scores or the respondent thought having a TOEFL ITP score listed on a student's curriculum vitae would help the student when applying for jobs. It appears that in Indonesia and Mexico, TOEFL is a more well-known test than TOEIC, even though TOEIC is the more appropriate assessment for a workplace context. One respondent from Mexico commented, "I think the majority of institutions and companies here in Mexico go for ITP, for TOEFL. I've seen it" (M4, English-language teacher within a public institution). The interviews did not shed much light on how TOEFL ITP scores were actually used for decisions when applied for workplace purposes, making it difficult to judge the extent to which these uses may be (in)appropriate.

The participants who mentioned using the TOEFL ITP for program evaluation talked about judging the utility of their English-language programs in improving students' scores, but it was not really clear whether these evaluations were done in a formal manner, what actions were taken based on the results of such program evaluations, and whether other factors were taken into account (e.g., it is unlikely that TOEFL ITP scores would be appropriate to use as the sole criterion in

decisions to change the curriculum). The following quote illustrates what individuals said about applying TOEFL ITP scores for program evaluation:

It is important for us as the language training provider for the students, for the candidates, whether the training is useful, or effective to improve the students' TOEFL score. In other words, to improve the students' language proficiency. And if there is no significant improvement for example, then we have to evaluate our course and training. (I7, employee within the language center of a public college or university)

An interviewee who mentioned using TOEFL ITP scores to admit students into a teacher training course described this use as follows:

We require 600 points minimum on the TOEFL to enter our teacher training course. . . . We use a language certification as a first filter. You cannot even put your documents up on the platform unless you have 600 points on the TOEFL or a B2 level in one of the European exams, for example, or for French, or different languages. Then we still do our own entrance exam where we're specifically having people write. We have an oral interview. We also do a grammar exam where we're looking at very specific points. . . . We want those students to enter a teacher training course and be able to be student teachers in our groups, and not have language issues. (M6, faculty/department administrator at a public college or university)

The quote illustrates that the individual had an understanding of the TOEFL ITP's limitations for her context and was therefore using it in conjunction with other selection tools, and only as an initial screening tool. This may represent an acceptable use of the assessment, even though it is not technically aligned with one of the recommended uses listed on ETS's website.

Related to, but separate from, the question of how institutions themselves actually apply TOEFL ITP scores for decision-making, a number of respondents mentioned that students could take the TOEFL ITP at their institutions to meet a requirement at another institution. For example, one individual explained,

It is not that we use [TOEFL ITP]. Students come to us when they need to take the exam as a requirement to be admitted in a master program or to validate their knowledge in English when they are required to know a language at college level. They take the certification exam, they get the certification and then submit it to their universities. . . . For example, [names institution] know that we administer the TOEFL exam so they send us their students. They have their own language center, so they send us their students just when they need a proof (this case a certification) of their knowledge in English. They take the exam and if they get the required score then good, if not they have to stay with them and take classes in their language center. (M9, director at a language school)

Such an application of the test might represent a more high-stakes use than originally planned and may be problematic, as it is not clear whether all the necessary security measures are in place at these institutions, which are essentially being treated as secure test centers.

Research Question 3

Survey Results Regarding TOEFL ITP's Perceived Validity

We asked survey respondents how well TOEFL ITP scores seem to indicate students' English-language proficiency. The vast majority of participants responded that TOEFL ITP scores were either "very useful" (64.1%) or "somewhat useful" (28.3%) as indicators of students' English-language proficiency. A minority of respondents said that the scores were just "slightly useful" (6.3%) or "of little or no use" (1.3%). To more directly address perceptions of test validity for specific uses of the TOEFL ITP, we also examined average perceptions of scores' utility by test use with which individuals were most familiar. For these analyses, we only included respondents who picked one most familiar use. Regardless of test use, on average, respondents reported that TOEFL ITP scores are somewhere between very useful and somewhat useful.

Interview Results Regarding TOEFL ITP's Perceived Validity

Most of the interviewees ($n = 20$) were asked about their perceptions of the utility of the TOEFL ITP. Many of them talked about it being at least a slightly useful tool. Subthemes were not coded for this interview question, so we present our analyses of participants' responses without frequencies.

Perceptions of test utility seemed closely tied to users' struggles to find the right balance between considerations of test quality (based on the amount of information a test provides and inferences institutions can make about test takers' language proficiency) and considerations of test cost and practicality of its use. Even though users realized that the TOEFL ITP does not assess productive skills, they appreciated the amount of information it provides for a relatively low price. One interviewee described the dilemma this way:

I think that the paper-based version of the TOEFL test—that's my opinion—has the limitation of not measuring directly the productive skills, although they are somewhat measured throughout the Structure and Written Expression. I think it's mostly a receptive test. It's a multiple-choice test. The students are not producing anything. Especially for scientists who are going to give presentations, present papers, attend conferences, etc., production is really necessary. I have had students who do well in reading, but they cannot write because they don't have that experience. From that point of view, I have made here the recommendation that we should go to the iBT test. It's a form that really measures the level of students' ability in a more wholesome fashion. Here, the problem would be that it's a much more expensive test, and we do not have the infrastructure to apply for the certification as an iBT testing center, if that possibility exists. ITP is practical; it's easy to apply; it's easy to administer; there are many centers where people can get it; it's cost accessible. Still, I feel like it's not really measuring the skills that researchers or scientists should have. (M4, English-language teacher within a public institution)

Importantly, users' satisfaction with the skills assessed by the TOEFL ITP seemed to be partly a function of how they applied test scores, that is, what inferences about students they needed to make. In contrast to the earlier quote from the interviewee (M4) dissatisfied with TOEFL ITP scores as a basis for making inferences about students' ability to function effectively as researchers/scientists, consider the following quote from an interviewee who used the test for placing students into courses (based on the English skill area where they needed to improve):

I guess that the TOEFL ITP is better [than other exams] because we have listening, we have writing and structure, we have reading, and we can check how are they, the students, doing with the listening, with the writing expression, and the comprehension and reading comprehension, and everything. This is very important for the students and for us to get to the point . . . what they really need. If they need structure, we teach the structure and grammar and everything. They are not having correct reading comprehension—we will work on that. (M5, department administrator at a language school)

Still within the context of the question of the TOEFL ITP's utility, some talked about the test, as a measure of English proficiency, serving as a gatekeeper to opportunities students want. If students could receive a high enough score, it opened up doors for them. The following quote illustrates this idea:

I think every part of Indonesia, English is the most famous language and because they know this . . . most of countries in the world offer so many scholarships to Indonesian students. That you know, there is a minimum requirement in the language, English. So, when students are able to communicate and having a good TOEFL score, there are passport for them and they can study overseas. . . . And if they have 550 TOEFL score then they—I can say that there is guarantee that they will obtain the scholarship to study everywhere in the world. . . . But it is I think big opportunity for them to study in U.S. So that's why I say I think it is very useful. (I4, university central administrator at a public college or university)

Interview Results Regarding Needs Met (or Not Met) by TOEFL ITP

When comparing TOEFL ITP to other, externally developed tests of English-language proficiency, it was most common for interview participants to mention the issue of cost and the fact that the TOEFL ITP is less expensive relative to other

tests (e.g., TOEFL iBT, IELTS; 10 respondents). Thus the TOEFL ITP may meet the need for an affordable test better than some other commercially available English-language assessments. The following is an illustrative quote:

It's very useful for us to have an exam that's not as expensive as the iBT would be that still permits us to have an international exam for all the different uses. . . . For some people to pay for an iBT, that's a lot of money . . . especially if you have to retake it, which is often the case for the ITP with the doctorate of law students. They don't always get it the first time. They have to retake it and retake it. Some people end up really investing a lot of money into their language certification, because they didn't get it the first time, because they didn't know what to expect or they just didn't have the level, obviously. To have an exam that's not terribly expensive is really helpful. (M6, faculty/department administrator at a public college or university)

One respondent from Indonesia pointed out that when there was no external pressure to take an exam other than the TOEFL ITP, students preferred the TOEFL ITP (to evaluate their English-language proficiency) because of its lower price. But if students needed to take a four-skill test to be able to meet an institutional requirement (e.g., for study abroad, for scholarships), they would choose another test.

In addition to price, five respondents commented that they appreciated TOEFL ITP's more flexible test administration relative to some other commercially available English-language assessments, as illustrated by the following quote:

We don't have the registration to have the TOEFL iBT . . . TOEFL ITP is better because we can place it at our school or at a university, whatever. Where they need it, we can go. For us, iBT is very difficult because they need to have a lot of computers and another approval from TOEFL for us to get this test. (M5, department administrator at a language school)

Additionally, two respondents expressed agreement with the 2-year expiration policy for confirming TOEFL ITP certificates because language ability does not stay constant over time; the expiration policy forces people to maintain and periodically recertify their English skills. Some language tests administered by other test providers, on the other hand, provide certificates that do not expire, a policy that these respondents did not find appropriate. One individual from Indonesia pointed out that the TOEFL ITP is an appropriate assessment for students with relatively low English-language skills, implying that the TOEFL ITP targets the appropriate level of difficulty for the population in that context.

While the TOEFL ITP may be less expensive and may be easier to administer relative to other tests, and some respondents may appreciate its 2-year policy for the language certificate, there appeared to be some unmet needs. Six respondents mentioned the narrower construct coverage of TOEFL ITP relative to some other tests, because the test does not evaluate productive skills. The following is an illustrative quote:

We're in the process of being authorized as a center for the iBT. . . . Obviously, [the iBT is] an exam with different characteristics. It's integrated skills, very much focused in on being able to analyze information and write about it or speak about it, which is something that students have to do in academic life. It's very different from answering multiple-choice questions [with the ITP]. (M6, faculty/department administrator at a public college or university)

Two respondents made additional comments about perceived shortcomings of the TOEFL ITP relative to some other tests, including that it is less flexible because the audio input in the listening section of the test is played only once (some tests administered by other providers play the audio twice) and that security measures are not as sophisticated as those for the TOEFL iBT. A third respondent mentioned the logistical challenge of obtaining the optional certificate (provided in addition to the score report), whereas other exam providers offer a certificate to all test takers.

Six respondents aptly commented on the TOEFL ITP and other assessments of English-language proficiency having different applications or meeting different user needs; they spoke about each test having its own place. This idea is illustrated in the following quote:

[TOEFL iBT is] a very good exam. It is more complicated to apply. Much, much more complicated to apply, more expensive, etc., that kind of thing, and it requires a different kind of preparation for students. I think each one has its own place. I like the ITP, because of the flexibility that we have in application. We can set our own dates. It's not

complicated to apply. It's very, very useful in that sense. (M6, faculty/department administrator at a public college or university)

Research Question 4

Survey Results

The majority of respondents perceived the various pieces of information ETS provides about the TOEFL ITP to be at least “slightly useful”:

1. The content of the different sections of TOEFL ITP (92.5%; 6.7% do not use)
2. How test takers can prepare for TOEFL ITP (83.8%; 11.8% do not use)
3. How TOEFL ITP is administered (84.4%; 14.8% do not use)
4. How TOEFL ITP section and total scores are calculated (89.9%; 7.6% do not use)
5. How TOEFL ITP section and total scores are reported (89.0%; 9.7% do not use)
6. The security of the TOEFL ITP test administration and score reports (84.3%; 14.0% do not use)
7. How long TOEFL ITP scores are valid (87.8%; 8.4% do not use)
8. The minimum TOEFL ITP test scores for entry into specific courses at [the respondent's] institution (67.1%; 27% do not use)
9. The relationship between TOEFL ITP test scores and scores on other English-language proficiency tests (89.0%; 7.6% do not use)
10. The validity and reliability of TOEFL ITP test scores (90.3%; 8.8% do not use)
11. The recognition of TOEFL ITP locally and internationally (88.7%; 10.1% do not use)

Interview Results

To further address the question of the utility of TOEFL ITP-related information, we asked interview participants to explain why they do or do not find a particular type of available information useful. When asking interviewees this type of question, we referenced how useful they had rated a piece of information when they completed the survey. Interview findings are organized by information type. Information types that were more frequently discussed by interviewees are presented first.

How TOEFL ITP Scores Relate to Scores on Other English Proficiency Tests

Nineteen participants were asked to explain why they did or did not find information on how TOEFL ITP scores relate to scores on other English proficiency tests to be useful. Eighteen of the 19 had indicated on the survey that they found this information at least slightly useful; 1 indicated that she did not use this information.

Several respondents who indicated using this type of information talked about comparing TOEFL ITP scores to other tests (e.g., IELTS) to answer test takers' questions or make decisions about whether students meet institutional requirements.⁸ One respondent indicated using this type of information to make the case for why her institution should use TOEFL ITP scores (to help decision makers understand what a TOEFL ITP score means). The following is an illustrative quote:

It's useful for me in the sense that as a program designer, as a curriculum designer of this institution, knowing the different equivalencies. Some people come with different types of tests and you can't take the TOEFL just as the one authority. It's useful for me to know how the different tests are scored and what the equivalents are and things of that sort in order for me to make informed judgments based on the program that we have, on hiring teachers, and of the sort. (M1, academic teacher or researcher at a public college or university)

Interview responses of several participants from Indonesia and Japan did not correspond well to their survey responses regarding the utility of information about how TOEFL ITP scores relate to scores on other English proficiency tests.⁹ Specifically, two Indonesian individuals responded on the survey that they found this type of information to be useful but then did not have an explanation of how they actually used that information; one of these individuals indicated that other

institutions need to use this sort of information but she does not. A third respondent indicated on the survey that she did not use information about TOEFL ITP's relationship with other English proficiency tests but talked about using it when asked during the interview: “[Nonnative speakers] should have to take TOEFL IBT or TOEFL ITP in our university. . . . We have like a score sheet to transfer, to compare the IBT scores to ITP scores in our school” (J1, faculty/department administrator at a university).

Security of the TOEFL ITP Administration and Score Report

Fifteen participants were asked to explain why they did or did not find information about the security of the TOEFL ITP administration and score report to be useful. Fourteen of the 15 had indicated on the survey that they found this information at least useful; one indicated not using this information. Participants talked about following the security protocols recommended by ETS when administering the test. For example, one participant said,

Every time any TOEFL is administered at our institution, we do follow the procedures with timing, with checking the IDs, going around monitoring, making sure that they're filling out the answer sheets correctly with their names. At the end, we check IDs one more time to make sure that the name on the answer sheet is the same one on the test book. We follow all sorts of procedures every single time we administer the test. (M1, academic teacher or researcher at a public college or university)

Test administrators may even go beyond the security measures recommended in the ETS procedures to further strengthen security and discourage fraud. The following quote demonstrates this:

We ask for two IDs instead of one, and things like that, to make sure that we don't have like one student taking the test for somebody else. . . . So we make sure that both IDs are official, that there is a signature and the signature matches. We have a good security procedure to make sure that students don't cheat. (M8, English language teacher at a private university)

Recognition of TOEFL ITP Nationally and Internationally

Ten participants were asked to explain how they used information about the recognition of the TOEFL ITP nationally and internationally. These participants had all indicated on the survey that they found this information at least useful. Two more participants spoke about the importance of the TOEFL ITP's reputation in the context of answering another question. Individuals talked about TOEFL ITP scores being recognized and accepted in their country and overseas, getting a boost in institutional reputation for administering the test, and using the test's reputation to encourage individuals to obtain proof of their English-language proficiency. The following is an illustrative quote:

When students or candidates say “Why do we have to do the TOEFL?” . . . I say, “Because it's accepted all over the place. . . . It's an international standard. If you're going into a doctorate program, you need to be thinking globally.” It's useful to be able to show them that's what the situation is. . . . For [name redacted], it's very important, this whole globalization concept of academia being a global and not necessarily a local or national situation. (M6, faculty/department administrator at a public college or university)¹⁰

A couple of participants, however, seemed to misunderstand the question posed, as their responses were unrelated to the recognition of the test.

How Long TOEFL ITP Scores Are Valid

Nine participants were asked to talk about how they used information regarding how long TOEFL ITP scores are valid. On the survey, they had all responded that this information is at least slightly useful. Two additional participants commented on the issue while responding to a different question. Interviewees talked about the expiration policy forcing people to periodically retake the test and took the opportunity to explain whether they agreed with the policy. For example, one participant explained,

Sometimes referrals say, “Why do we have to take this exam again?” It’s useful to have an international standard to say, “TOEFL says its 2 years,” . . . because in 2 years your language level can change a lot, either up or down. It’s very useful to be able to say, “There’s been research behind this and that’s the standard for TOEFL.” Some of our exams that have also a limitation or the graduate programs says, “Even if your certificate says it’s for life, we want it to have a 2-year or a 3-year maybe date.” Then it’s useful to go back to you as support. (M6, faculty/department administrator at a public college or university)

Minimum TOEFL ITP Scores for Entry Into Classes

Seven participants were asked to explain how they used information about the minimum TOEFL ITP scores for entry into classes. These individuals had indicated on the survey that this information was at least “somewhat useful.” Individuals mostly spoke about using TOEFL ITP scores to make judgments about placement; nothing they said indicated that they actually received or used guidance from the TOEFL ITP program on the minimum scores for entry into classes, suggesting that this question was not a very meaningful one for them. One respondent specifically mentioned that various stakeholders were the ones who chose scores that would be used for placement into classes at different levels:

We use the chart that — you know, in national meetings, teachers or coordinators representing the different campuses agreed on the scores needed to be placed in the different levels. That’s what we use to show — that table in order to standardize the levels and the scores needed to be placed in specific levels. (M7, faculty/department administrator at a high school)

Validity and Reliability of TOEFL ITP Scores

Six participants were asked to talk about how they use information about the validity and reliability of TOEFL ITP scores. With the exception of one individual from Japan who said she did not use this information, the other five had indicated on the survey that this information was at least slightly useful. On the basis of their interview responses, participants apparently did not use information about the TOEFL ITP’s validity and reliability per se. Participants indicated taking it as a given that the test is of high quality when it comes to measuring the skills it focuses on but did point out that it does not measure all four language skills. One participant said,

I believe that there are no questions about the validity and reliability of the test, actually, because this is sort of international standardized test. . . . TOEFL test is very good test actually. It must measure English proficiency. But, the question is whether the test content relevant to what the students trying to achieve. . . . Some students with high TOEFL test, we can’t guarantee that is also successful in their further studies . . . Someone with high TOEFL score, we can’t guarantee that he can communicate in English effectively as well. (I7, employee within the language center of a public college or university)

Decision makers were not always focused on issues of reliability or validity but focused on more practical matters instead. A participant who rated this type of information as only slightly useful said during the interview that even though she considered validity very important, decision makers were not particularly concerned about it; they were more interested in receiving information about other schools’ use of TOEFL ITP scores (for benchmarking).

TOEFL ITP Score Reports

Two participants were asked to explain how they use information about TOEFL ITP score reports, as they had indicated finding this information useful or very useful. One indicated that the score reports were informative for understanding in which specific skill (e.g., listening, reading, structure) a student needed further practice. The second did not provide a particularly informative response to the question asked.

How Students Can Prepare for TOEFL ITP and Content of Different Sections of the Test

One participant was asked to talk about how she used information regarding what students can do to prepare for the TOEFL ITP. She indicated that the available online resources and books were adequately informative about the types

of questions on the test and how to answer them. One participant was invited to comment on how she used information about the content of different sections of the test. She talked very generally about reading to students the instructions provided at the beginning of each section of the test to explain what they should or should not do. On the survey, both participants had indicated finding the respective pieces of information very useful.

Research Question 5

Survey Results

Survey participants indicated wanting various additional information or resources related to the TOEFL ITP. Some more common themes were additional test preparation materials for students, normative TOEFL ITP scores for various groups (e.g., by educational level, national, international), level of demand for TOEFL ITP scores from universities or employers, and how TOEFL ITP scores relate to scores on other tests (e.g., TOEFL iBT, TOEIC, Cambridge English tests) or to the CEFR levels.

Interview Results

Interviewees were also asked to talk about their outstanding information or support needs. Nine interviewees suggested that they would welcome support to help their students understand why English-language proficiency is a critical skill to acquire rather than viewing TOEFL ITP as an obstacle. This is illustrated in the following quote:

They see English only as an obstacle to graduate. I think we need to work on making them realize that in scientific research, English is a much-needed tool. You can succeed, but you will be much more successful if you master English. (M4, English-language teacher within a public institution)

Relatedly, five participants wanted support promoting or advertising the TOEFL ITP (e.g., promotional materials) to attract more test takers who would like to validate their English proficiency.

Interviewees wanted support around better preparing their students to take the TOEFL ITP. Three respondents asked that ETS provide training for teachers or a forum for institutional users to meet so that users may increase their knowledge of the TOEFL ITP (e.g., the subskills needed for each section) and improve the quality of their instruction. One interviewee wanted to learn how much of a score gain is reasonable to expect upon retest.

Participants also talked about wanting more test preparation materials for their students, including up-to-date practice books (four respondents), information about useful test-taking strategies (one respondent), and detailed feedback on their test results (three respondents). For example, one participant commented,

Whenever we posted results for students and if they saw no big difference—if they notice that the score they had gotten was the same or lower than a prior one they used to have, but what was the problem? . . . Sometimes they are—“Is it possible for me to take a look at my exam?” Of course, I told them—I will always tell them we can’t do that, but if we could be given some feedback on, for example, reading—“Well, the problem is inference questions or questions about details,” that would help a lot. That would let us help our students improve on those skills. (M7, faculty/department administrator at a high school)

To facilitate decision-making, participants asked for easy access to information relating TOEFL ITP scores to scores on other exams or the CEFR (four respondents). For example, one participant explained the value of having information about the relationship of test scores with the CEFR levels¹¹:

For example, the Supervisor’s Manual, it would be good to have table of comparisons for the different—not only for the different exams that there are out there, but also, to the Common European Framework. A lot of people in this region work with the Common European Framework. They are familiar with the levels and the different can do statements that the framework has . . . it would be good to try and align the TOEFL . . . whether it’s the ITP, the iBT, whatever it is . . . with the Common European Framework, since it has become quite a force of curriculum

development and goal setting within language programs in this area. (M1, academic teacher or researcher at a public college or university)

A respondent in Japan wanted to know the level at which other universities set their TOEFL ITP score requirements so that she could help decision makers at her institution benchmark their own expectations for students.

Interviewees also wanted to better understand ETS's decisions about the design of the TOEFL ITP. Four respondents wondered about the choice to use test topics that are not very familiar for their culture, as some users have concerns that the content of the test may disadvantage their students. One participant explained, “[TOEFL ITP] is very contextual for the Western countries for reading. It’s history. It’s all about America. . . . For instance, the listening and the reading, it’s all very Westernized. . . . We are trying to give understanding to our students about the context” (I1, English-language teacher at a private college or university). Three interviewees wondered why the audio stimuli in the listening section are only played once and why students do not get extra time to circle answers while they listen to the audio; to one individual, this seemed to suggest that the listening section tests students’ memory. One respondent wondered about the possibility of a shorter form of the TOEFL ITP assessment.¹²

Discussion

Limited research to date has examined how institutions are actually applying TOEFL ITP scores, whether and how they use cut scores to facilitate decision-making, and the extent to which they perceive the test to be a valid assessment of English-language skills. Additionally, researchers have not examined the related questions of how users are utilizing the TOEFL ITP-related information provided by ETS and whether they have any unmet informational needs. The current study used survey and interview data from institutional TOEFL ITP users in Japan, Mexico, and Indonesia to start to address these questions.

Research Question 1

Survey findings regarding institutions’ uses of TOEFL ITP scores suggested that there may be differences in the popularity of different test applications across countries. Additionally, more than half of the interviewees described not using the assessment for purely institutional purposes. They described applying the test for job applications, work placement, initial screening of teacher training candidates, and English-language program evaluation. Unfortunately, our small sample size of interviewees did not permit estimation of how widespread these unintended uses may be, so further research is needed to address this issue.

Research Question 2

Related to the question of how users make decisions based on test scores, survey results indicated that some users take multiple pieces of information into consideration, such as whether individuals show an adequate level of English proficiency during an interview. Taking into consideration other sources of evidence of individuals’ proficiency, not just test scores, is consistent with recommendations both by ETS and non-ETS researchers regarding the use of scores from other tests in the TOEFL family of assessments (Kokhan, 2012; Papageorgiou & Cho, 2014). Yet, more than half of the survey respondents did not indicate using any tests or information other than TOEFL ITP scores for the purpose for which they were applying the test. One related consideration is that many respondents were helping students to meet a score requirement of an external institution, so they may have been unable to report other informational inputs institutions may actually consider. We did not break out survey responses to the question about use of additional information by TOEFL ITP application because of small sample sizes for these analyses.

As mentioned earlier, we found out from interviewees that users may be applying TOEFL ITP scores to make certain decisions for which this assessment was not intended. Before determining how to address this concern (e.g., advising users against certain applications, conducting research to support the validity of additional interpretations and uses of TOEFL ITP scores), we need to better understand why and how institutions apply the TOEFL ITP in nonrecommended ways. Regarding the “why,” demand from employers and students may be one of the driving forces behind these applications and seems related to users’ need for an accessible (especially in terms of cost) and well-recognized assessment, even if,

content-wise, it is not the ideal assessment for the targeted application. Knowing the “how” is important, because these nonrecommended uses of the assessment (some of which sound like higher stakes applications than intended for the TOEFL ITP given the security measures in place) may be more or less problematic depending on situational factors. For example, using TOEFL ITP scores as part of the selection process for employees who only need to read English texts or listen to input in English may be less problematic than using these scores to select English-language teachers. The concern in the latter case would be that the TOEFL ITP does not assess the ability to speak or write in English, which are important skills for English teachers to have. The TOEFL iBT might be a more appropriate test for assessing the language proficiency level of future English-language teachers, as one of its intended uses, screening of international teaching assistants in university contexts, is somewhat relevant (see Xi, 2007).

Users often understand the inability of tests like the TOEFL ITP to provide rigorous test security and comprehensive content coverage in the limited time for testing. However, to the extent that misconceptions about the test also contribute to its “misuse” (i.e., applications of the test that are inconsistent with ETS’s recommendations), it would be valuable to provide users with more information. Specifically, ETS should help educate score users about the importance of using a test in ways that are evidenced to be valid and reliable, and this should be done using language that is accessible to a nontechnical audience. The TOEFL ITP program may consider instituting workshops similar to the Propel workshops that are being developed for TOEFL iBT users. In fact, interviewees voiced an interest in learning more about the TOEFL ITP via some sort of workshop or seminar.

Also related to the issue of potentially problematic TOEFL ITP-related practices, some interview participants talked about testing students on behalf of other institutions (though this was only the case for Mexico and Indonesia; Japanese institutions conduct their own test administrations¹³), despite the fact that the TOEFL ITP is intended for internal use by institutions. The fourth recommended test use on ETS’s website—admissions to short-term, nondegree programs in English-speaking countries where the sending and receiving institutions agree to use TOEFL ITP scores—suggests that scores are shared directly between institutions. On the other hand, to test students on behalf of other institutions may represent a more high stakes use than intended for the TOEFL ITP. It could be problematic if there are inadequate security measures in place at these institutions to prevent test-taker misconduct. It is possible that students tested elsewhere are allowed to present their own copies of score reports to their home institutions (vs. score reports issued directly to administering institutions).¹⁴ If the situations in Mexico and Indonesia are such that institutions are not always able or willing to administer the TOEFL ITP themselves, it would be important to provide users with recommendations to enhance test security in situations where institutions administer the test on behalf of other institutions.

Related to the question of where, if at all, test users set TOEFL ITP minimum (or “cut”) scores, we found that the majority of survey respondents were using minimum scores. But, not surprisingly, use of cut scores was partly dependent on test application (e.g., whether it requires some type of pass–fail decision). Unfortunately, we did not find out very much from interview participants regarding how cut score decisions are made. Interview data did seem to suggest that minimums cited by survey respondents may have often been ones set by external institutions (e.g., institutions in the United States or other English-speaking countries) and that the role of the respondent was to prepare students via English-language training to achieve that target.¹⁵ When users work with cut scores set by external institutions rather than their own, they are unlikely to know how cut scores were set.

An outstanding question is whether users typically set TOEFL ITP cut scores based on the alignment of these scores to the CEFR levels (several interview participants mentioned doing this). Using the CEFR to set cut scores on the TOEFL ITP may not be ideal given (a) that the performance levels and descriptors of the CEFR are intentionally broad and generic so that they can be applied in a variety of educational contexts and (b) the interpretation of the framework and its levels is up to test users (Papageorgiou, Tannenbaum, Bridgeman, & Cho, 2015; Powers, Schedl, & Papageorgiou, 2017).

The important takeaway from the minimum score results and respondents’ related perceptions is that institutional test users can probably benefit from additional support around setting cut scores. Only a minority of test users reported the use of minimum section scores for the TOEFL ITP. Institutions may benefit from setting *both* total and section score requirements; ETS recommends taking both into account. Total scores are more reliable, as they are based on all test items, and are a better estimate of overall language proficiency (Papageorgiou & Cho, 2014). However, as Bridgeman, Cho, and DiPietro (2016) have shown, section scores contain important information for each language skill that may be informative for understanding students’ specific profiles. Supporting test users’ cut score decisions does not have to mean

working with them on a case-by-case basis but can be a matter of providing them with accessible information on the factors to take into consideration when making these decisions on their own.

Research Question 3

The majority of survey respondents had favorable opinions about the usefulness of TOEFL ITP scores for determining students' English-language proficiency, regardless of the specific use for which they were applying the assessment. Interviewees' comments about the TOEFL ITP's usefulness illustrated their need for a relatively inexpensive and practical English-language assessment that also provides them with enough information to make decisions about test takers' proficiency in all relevant skill areas. It can be challenging for users to find the right balance between a test with wide construct coverage (in terms of the skills tested) and accessibility; while the TOEFL ITP is an affordable test and is relatively easy to administer owing to its paper-and-pencil format, it does not assess productive skills. For some users, depending on how they use the test, this limitation is not a big concern. Others talked about this limitation being problematic and voiced a desire to switch to using the TOEFL iBT. However, users face the challenges of not necessarily having the facilities or equipment for TOEFL iBT administration and test takers pushing back on paying more for an English-language assessment, especially because many of them have to take the test multiple times to achieve targeted scores. As discussed earlier, the TOEFL ITP is apparently applied even more widely (for more uses) than intended by ETS; we believe this is at least partly due to it being so accessible and students preferring it to other tests (if given an option).

Research Questions 4 and 5

Related to the question of information usefulness, the majority of survey participants reported positive perceptions about the utility of various pieces of information provided about the TOEFL ITP assessment. Respondents did indicate wanting additional information about the TOEFL ITP, such as how test scores relate to scores on other English-language assessments or to the CEFR. Information about how TOEFL ITP Level 1 scores relate to the CEFR is available on the ETS website¹⁶ or from publications (Powers *et al.*, 2017; Tannenbaum & Baron, 2011); corresponding information for TOEFL ITP Level 2 is under development at the time of writing. Some users may be unaware of the existence of this information for Level 1 or, perhaps, have difficulty processing this information in English.

Interviewees talked about the value of information about the TOEFL ITP's relationship with other English-language assessments (e.g., TOEFL iBT, TOEIC, IELTS, language tests administered by CaMLA) or the CEFR. This type of information helps institutions provide students with different options of tests they can take to demonstrate a certain level of language proficiency and meet some type of requirement (e.g., scholarship, graduation). Participants tended to assume that a TOEFL ITP score can simply be converted into a score on another English-language test. However, it is not appropriate to directly compare scores across tests that differ in terms of their content coverage, intended applications, and stakes (Dorans, 2008). Relatedly, users may have some misunderstandings about the CEFR. First, they may not realize that there is no official body accrediting claims exam providers make about the alignment of their scores to the CEFR levels and that the framework's interpretation is up to test users (Papageorgiou *et al.*, 2015). The CEFR is a product not of the European Union (as one respondent in Mexico incorrectly referenced) but of the Council of Europe, an international organization promoting democracy and protecting human rights and the rule of law in Europe,¹⁷ which cannot make binding laws. Second, although the CEFR levels serve as a convenient way to obtain some understanding of the comparative difficulty of different tests (based on the way test score ranges are linked to the framework), tests claiming alignment to the same CEFR levels should not be considered equivalent in terms of content and difficulty (Papageorgiou & Tannenbaum, 2016). Therefore it might be useful for test users to be provided with nontechnical promotional materials that explain what kinds of inferences about TOEFL ITP scores relative to other tests are more or less appropriate to make.

Interviewees felt informed about the TOEFL ITP's security measures and, if needed, were able to implement additional security measures beyond those recommended by the TOEFL ITP program to further discourage test-taker misconduct. It is important to point out that test-taker attempts to game the test are likely to be a particularly high concern when the test is applied in a relatively high stakes fashion (not for purely institutional purposes) because of higher incentives to gain an unfair advantage. Users would benefit from further support around preventing or detecting test-taker misconduct. A respondent raised the specific question of how much of a score gain is reasonable to expect upon retest; such information could help detect suspicious score gains.

On the basis of their responses, interviewees associated the TOEFL name with quality; they did not question that the TOEFL ITP must be a valid and reliable test. It is not completely clear how their opinions were formed, though it was likely through some combination of ETS's reputation as a test publisher, the popularity of the TOEFL ITP in their countries and internationally, and observations of students' abilities relative to their TOEFL ITP scores. Interviewees' biggest concern was that the TOEFL ITP is not designed to measure all the skills they consider critical. Respondents also raised the question of whether the Western focus of the test content may be disadvantaging their test takers. There is no evidence from routine and ongoing differential item functioning analyses as part of the operational administration of the TOEFL ITP to show that cultural background unfairly affects test scores. From a test design perspective, routine test development procedures also aim to ensure that items can only be answered based on information provided in the passage/audio rather than from background knowledge. Furthermore, published research with the TOEFL iBT, an assessment that targets similar language use domains and test-taker ages as the TOEFL ITP, has suggested that background knowledge does not meaningfully impact performance (Liu, Schedl, Malloy, & Kong, 2009). Such investigations should continue, and perhaps findings should be communicated to stakeholders in an accessible way to address concerns about whether TOEFL ITP content disadvantages a particular group of test takers.

Beyond the informational needs or concerns already discussed, interviewees saw a need for easier access to useful information. It is not safe to assume that simply putting information, such as policy updates, on ETS's website is adequate for reaching users. Teachers and students may not visit the website on a regular basis and may not know to check it for updates. One piece of information that users wanted to be more accessible was how TOEFL ITP scores relate to the levels of the CEFR. Users would likely find it helpful also to have this information in printed form (e.g., in the supervisor manual or packet of information provided to test centers).

Methodological Strengths of Current Study

There is value in combined survey and interview data collections such as ours. Surveys are a practical means of quickly gathering data from a large number of test users (many more than would be feasible to interview), summarizing their views and experiences, and highlighting areas in need of further probing via interviews. But using interviews, we were able to learn more than the results of our survey could tell us about why users perceive the TOEFL ITP or information ETS provides about the test to be more or less useful, how decisions are made based on test scores, and the extent to which users feel adequately supported. Furthermore, analyses of participants' interview responses relative to their survey responses highlighted the fact that survey responses did not always represent the complexity of institutions' actual uses of the TOEFL ITP assessment. This underscores the value of interviews for gaining deeper understanding of phenomena and the drawbacks of relying solely on questionnaires with selected responses.

Limitations and Future Directions

A few limitations of our study should be noted. One limitation pertains to sampling of participants. Ours were convenience samples, rather than representative samples, of test users in Japan, Mexico, and Indonesia, as we do not have information about how the population of users breaks out in terms of background variables (e.g., type of institution, role, type of test use). The responses provided by survey and interview participants accurately reflect the experiences and opinions of the broader population of institutional TOEFL ITP users to the extent that these samples actually reflect the nature of the broader population. Given that only three individuals from Japan were ultimately available for interviews, our interview findings are probably least representative of the experiences and needs of TOEFL ITP users in Japan.

Furthermore, the sample of participants in this study was just a subset of the institutional test users invited to participate. Given the anticipation of having to provide information about test use and related opinions, the individuals who agreed to participate (especially those who agreed to an interview) were likely more knowledgeable about the assessment and had more experience with it than the test users who declined the invitation. The fact that our respondents were likely the more knowledgeable ones among the user population implies that any unmet information needs that they highlighted are likely even more pressing concerns for the broader population of test users. Another issue related to individuals self-selecting to participate in our study is that the users who agreed to participate may have had more positive views about the TOEFL ITP or about ETS in general. This is a potential source of bias in our results; it is possible that if other test users had participated in the study, less positive views on the test would have been found.

Another limitation stems from the challenges associated with trying to ask respondents about multiple test uses within a single survey. This was the most efficient approach to simultaneously addressing RQs about a variety of test uses, but it put the burden on respondents to follow instructions—specify just one use with which they were most familiar, and answer a number of subsequent questions specifically with regard to that one use. Web survey programming allowed respondents in Mexico to select no more than one use of the test. However, respondents in Japan and Indonesia, who took a paper-and-pencil survey as per ETS representatives' request, had the opportunity to overlook the instructions and check off more than one use of the test. When analyzing perceptions of score utility by test use, we removed individuals who checked off more than one use of the TOEFL ITP in their institutions. Yet, it is still possible that some of the analyzed individuals responded with multiple test uses in mind (despite following instructions and only checking off one when asked). It would be beneficial for future research to replicate our findings, but to do so using surveys targeting one TOEFL ITP use at a time, potentially starting with the most popular one.

Likewise, because of the choice to ask participants about multiple test uses within the same survey, we are not able to effectively address the support needs associated with a specific use of the TOEFL ITP. Participants' responses about the usefulness of various pieces of information ETS provides or other types of information needs they have were not tied to one particular use of the test. Future research that addresses specific TOEFL ITP uses can further investigate informational support needs associated with a particular test application.

We did not conduct cross-country statistical comparisons of survey results and caution against attempts at cross-country comparisons. We used a convenience sample, and the composition of that sample (e.g., in terms of institution type, respondent role, and respondent activities) differed across countries. Given that test uses and opinions are likely related to respondents' backgrounds, we did not think it was appropriate to compare how these different samples responded. Future studies that examine TOEFL ITP uses one at a time and that have more control over participant sampling from the larger population would be better positioned to make direct comparisons across countries. One way to encourage more TOEFL ITP users to participate in future studies would be to use current interview findings to enhance the relevance and clarity of survey questions.

As highlighted earlier, even though we included information on the number of interviewees who mentioned a particular theme, we caution against using these frequencies to judge certain themes to be more or less important than others. One reason is that our sample is, again, not necessarily representative of the full population of institutional test users. A second reason is that users were not all asked the same questions. We tried to probe for more information about individuals' own survey responses and to ask questions about issues where we expected a greater variety of responses. Thus certain questions were asked more frequently than others. More frequent themes are simply ones where we were able to gather more evidence of a particular TOEFL ITP use, perception, or outstanding need.

Owing to the language barrier between the interviewer and interviewee for the interviews conducted in English (because English was not the interviewees' native language), there were some instances of interviewees apparently misunderstanding questions and giving seemingly irrelevant responses. By the same token, it is possible that the interviewer did not always interpret interviewees' responses as intended. Furthermore, owing to poor audio quality and/or heavily accented speech, some responses could not be captured fully in transcripts, and two interviews could not be conducted via telephone as planned. Subsequently, some interviewee comments may not have been coded correctly, or at all. However, we can be fairly confident that the themes that occurred relatively frequently capture real user experiences or needs.

Given that it was not possible to probe every aspect of users' TOEFL ITP uses, experiences, or needs, there is room for more research into these issues. In particular, it would be useful to gain a better understanding of the more questionable uses of TOEFL ITP that emerged from our interviews. Test developers can then determine how to best approach the situation to improve assessment literacy and help ensure valid applications of test scores.

Acknowledgments

This study was conducted with funding by the TOEFL ITP program. The first author was employed at ETS when the study was conducted. We would like to thank the TOEFL ITP representatives in Japan (Council on International Educational Exchange [CIEE]), Mexico (Institute of International Education [IIE]), and Indonesia (Indonesian International Education Foundation [IIEF]) for their assistance with the data collection. We are also grateful to Don Powers, Margarita Olivera Aguilar, and Yeonsuk Cho for their helpful feedback on an earlier version of this report.

Notes

- 1 The acronym *ITP* stands for “institutional testing program.” However, only the acronym is currently used. *TOEFL ITP Assessment Series* is also used to refer to both TOEFL ITP tests targeting different proficiency levels.
- 2 http://www.ets.org/toefl_itp/faq/
- 3 For more information, see https://www.ets.org/toefl_itp/content/
- 4 http://www.ets.org/toefl_itp/use
- 5 While some number of misclassifications can be expected whenever score-based decisions are made because of the inherent measurement error contained in any test score, the goal is to determine which types of misclassification (false positives or false negatives) are more consequential in a given situation and set cut scores that minimize that type of error (Zieky & Perie, 2006).
- 6 We are unable to calculate participant response rates because ETS representatives declined to tell us how many institutions they invited to complete the survey.
- 7 For the alignment of TOEFL ITP Level 1 scores to the CEFR levels, see Tannenbaum and Baron (2011).
- 8 As we discuss later, test users need support to understand what types of inferences about TOEFL ITP scores relative to other tests are more or less appropriate to make.
- 9 This potentially reflects a misunderstanding of the questions — as translated into their native language on the survey or as asked during the interview in English — or inability to convey their experiences adequately in English during the interview.
- 10 Although the respondent did not specifically mention TOEFL ITP in her response, we assume she was referring to this test because the question asked was about TOEFL ITP and she spoke at other points about the use of TOEFL ITP scores for graduate admissions.
- 11 This information is currently available online for TOEFL ITP Level 1 and is under development for TOEFL ITP Level 2.
- 12 Information about test length and time limits for TOEFL ITP Level 1 and Level 2 is available here at https://www.ets.org/toefl_itp/content
- 13 ETS’s TOEFL ITP representative in Japan informed us that Japanese institutions always administer TOEFL ITP on their own campus, bringing in test administrators from outside the institution if additional support is needed.
- 14 TOEFL ITP provides two copies of the score reports, one for the student (pink color) and one for the administering institution (green color).
- 15 Close to half of our interview participants represented language training or test preparation institutions.
- 16 https://www.ets.org/toefl_itp/research
- 17 <http://www.coe.int/en/web/about-us/do-not-get-confused>
- 18 In the survey for Mexico, the response option read “Public University (4 years or more).”
- 19 In the survey for Mexico, the response option read “Private University (3 years or more).”
- 20 This option is unique to the survey for Indonesia. It was not included for Japan or Mexico.
- 21 This question was not included in the survey for Japan.
- 22 This response option was not included in the survey for Japan.
- 23 This response option was not included in the survey for Japan.
- 24 Definitions of *validity* and *reliability* were not included in the survey for Japan.
- 25 This question was not included in the survey for Japan.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baker, B. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *Canadian Journal of Applied Linguistics*, 19, 63–83.
- Bandara, W., Indulska, M., Chong, S., & Sadiq, S. (2007, June). *Major issues in business process management: An expert perspective*. Paper presented at the 15th European Conference on Information Systems, St. Gallen, Switzerland.
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, 33, 307–318. <https://doi.org/10.1177/0265532215583066>
- Choi, I., & Papageorgiou, S. (2014). *Monitoring students’ progress in English language skills using the TOEFL ITP assessment series* (Research Memorandum No. RM-14-11). Princeton, NJ: Educational Testing Service.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.
- de Wet, J., & Erasmus, Z. (2005). Towards rigour in qualitative analysis. *Qualitative Research Journal*, 5, 27–40.
- Dorans, N. J. (2008). The practice of comparing scores on different tests. *R&D Connections*, 6, 1–5.

- Forman, J., & Damschroder, L. (2008). Qualitative content analysis. In L. Jacoby & L. A. Siminoff (Eds.), *Advances in bioethics—Empirical methods for bioethics: A primer* (pp. 39–62). Oxford, England: Elsevier.
- Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, *14*, 113–139. <https://doi.org/10.1177/026553229701400201>
- 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* (Research Report No. RR-14-44). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12037>
- Hyatt, D., & Brooks, G. (2009). *Investigating stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK* (IELTS Research Report No. 10). Manchester, England: British Council.
- Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, *30*, 301–307. <https://doi.org/10.1177/0265532213480126>
- Kim, Y., Addom, B. K., & Stanton, J. M. (2011). Education for eScience professionals: Integrating data curation and cyberinfrastructure. *International Journal of Digital Curation*, *6*, 125–138. <https://doi.org/10.2218/ijdc.v6i1.177>
- Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, *29*, 291–308. <https://doi.org/10.1177/0265532211429403>
- Liu, O. L., Schedl, M., Malloy, J., & Kong, N. (2009). *Does content knowledge affect TOEFL iBT reading performance? A confirmatory approach to differential item functioning* (Research Report No. RR-09-29). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2009.tb02186.x>
- Malone, M. E., & Montee, M. (2014). *Stakeholders' beliefs about the TOEFL iBT test as a measure of academic language ability* (Research Report No. RR-14-42). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12039>
- Minton, T., & Nishikawa, S. (2007). The introduction of a standardized proficiency test (TOEFL ITP) into the English curriculum. *Bulletin of Liberal Arts & Sciences, Nippon Medical School*, *37*, 55–77.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, *30*, 363–380. <https://doi.org/10.1177/0265532213480336>
- Papageorgiou, S., & Cho, Y. (2014). An investigation of the use of TOEFL Junior® Standard scores for ESL placement decisions in secondary education. *Language Testing*, *31*, 223–239. <https://doi.org/10.1177/0265532213499750>
- Papageorgiou, S., & Tannenbaum, R. J. (2016). Situating standard setting within argument-based validity. *Language Assessment Quarterly*, *13*, 109–123. <https://doi.org/10.1080/15434303.2016.1149857>
- Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). *The association between TOEFL iBT test scores and the Common European Framework of Reference (CEFR) levels* (Research Memorandum No. RM-15-06). Princeton, NJ: Educational Testing Service.
- Powers, D., Schedl, M., & Papageorgiou, S. (2017). Facilitating the interpretation of English language proficiency scores: Combining scale anchoring and test score mapping methodologies. *Language Testing*, *34*, 175–195. <https://doi.org/10.1177/0265532215623582>
- Ryan, A. M., Reeder, M. C., Golubovich, J., Grand, J., Inceoglu, I., Bartram, D., . . . Yao, X. (2017). Culture and testing practices: Is the world flat? *Applied Psychology*, *66*, 434–467. <https://doi.org/10.1111/apps.12095>
- Shohamy, E. (2001). *The power of tests*. London, England: Longman.
- Silumbe, K., Chiyende, E., Finn, T. P., Desmond, M., Puta, C., Hamainza, B., . . . Bennett, A. (2015). A qualitative study of perceptions of a mass test and treat campaign in Southern Zambia and potential barriers to effectiveness. *Malaria Journal*, *14*, 1–11. <https://doi.org/10.1186/s12936-015-0686-3>
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*, 134–141. <https://doi.org/10.1037/0021-9010.81.2.134>
- Stricker, L. J., & Attali, Y. (2010). *Test takers' attitudes about the TOEFL iBT* (Research Report No. RR-10-02). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2010.tb02209.x>
- Stricker, L. J., & Wilder, G. Z. (2012). *Test takers' interpretation and use of TOEFL iBT score reports: A focus group study* (Research Memorandum No. RM-12-08). Princeton, NJ: Educational Testing Service.
- Tannenbaum, R. J., & Baron, P. A. (2011). *Mapping TOEFL ITP scores onto the Common European Framework of Reference* (Research Memorandum No. RM-11-33). Princeton, NJ: Educational Testing Service.
- Xi, X. (2007). Validating TOEFL iBT Speaking and setting score requirements for ITA screening. *Language Assessment Quarterly*, *4*, 318–351. <https://doi.org/10.1080/15434300701462796>
- Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service.

Appendix A

Survey Questions for Indonesia

Survey Part A: About Your Position

To better understand your responses, we would like some information about your role in your institution.

1. In which type of institution do you currently work? (Select one.)
 - a. Public college or university (4-year)¹⁸
 - b. Private college or university (4-year)¹⁹
 - c. Trade or vocational school
 - d. Secondary school
 - e. Other (please specify):
2. How would you classify your current primary role in your institution? (Select one.)
 - a. University central administration
 - b. Faculty/department administration
 - c. Academic teacher or researcher
 - d. English-language teacher
 - e. Student support services
 - f. Other (please specify):
3. How long have you been in your current role?
4. Which of the following activities do you perform that require considering students' English-language proficiency? (Select all that apply.)
 - a. Review applications
 - b. Answer admissions questions
 - c. Advise admitted students
 - d. Place students into English-language classes
 - e. Build students' English-language skills
 - f. Give input on admissions policies
 - g. Make decisions about admissions policies
 - h. Other (please specify):

Survey Part B: About the Use of TOEFL ITP

5. Do you currently use the TOEFL ITP in your job?
 - a. Yes (please go to Question 6)
 - b. No (please go to Question 16)
6. How do you use the TOEFL ITP test in your current job? (Please select ONE use with which you are MOST familiar.)
 - a. *To place* students in intensive English-language programs requiring academic English proficiency at a college or graduate level
 - b. *To monitor* students' progress in English-language programs stressing academic English proficiency
 - c. As an *exit* requirement from English-language programs to demonstrate proficiency in English listening and reading
 - d. *To admit* students to *short-term, non-degree programs in an English-speaking country* where the institutions sending and receiving students agree to use TOEFL ITP scores
 - e. *To admit* students to *undergraduate degree programs in a non-English-speaking country* where English is not the dominant form of instruction
 - f. *To admit* students to *graduate degree programs in a non-English-speaking country* where English is not the dominant form of instruction
 - g. *To admit and place* students in *collaborative international degree programs* where English-language training will be a feature of the program

- h. For *scholarship* programs, as contributing documentation for academic English proficiency
 - i. As one of the requirements for *graduation from undergraduate and/or graduate programs in universities in a non-English-speaking country*²⁰
 - j. Other purposes (Please specify):
7. Are other tests or additional information used for the purpose you selected in Question 6 above?²¹
- a. Yes (Please specify):
 - b. No

Please answer the questions that follow (8–10) while thinking ONLY about the test use you selected in question 6 above.

8. For the use of TOEFL ITP scores in your current job, what are the minimum required TOEFL ITP scores? (Please provide the ACTUAL minimum scores or indicate that there is no minimum for that section.)
- a. Listening Comprehension (scores 31–68):
 - b. Structure and Written Expression (scores 31–68):
 - c. Reading Comprehension (scores 31–68):
 - d. Total scores (scores 310–677):
 - e. There are no absolute minimum required scores.
 - f. I do not know.²²
9. Overall, what do you think about the minimum score requirements? (Select one.)
- a. They are much too low.
 - b. They are somewhat too low.
 - c. They are about right.
 - d. They are somewhat too high.
 - e. They are much too high.
 - f. We do not have minimum score requirements.
 - g. I have no opinion.²³
10. In your opinion, how useful are TOEFL ITP scores as indicators of students' English-language proficiency in your institution? (Select one.)
- a. Very useful
 - b. Somewhat useful
 - c. Slightly useful
 - d. Of little or no use
 - e. I have no opinion

Please answer the questions that follow (11–12) while thinking about the various ways you use TOEFL ITP scores in your job (if you use the test for more than one purpose).

11. How USEFUL is the information you can currently access (for example, on the ETS website or in the Test Taker handbook) about each of the following aspects of the TOEFL ITP test? Please select the appropriate number on a scale from 0 to 4 in Table A1 on the following page.
12. What other information about the TOEFL ITP test do you think would be useful given the various ways in which your institution uses the test?

Survey Part C: About You

13. How confident are you that the information you provided about your use of the TOEFL ITP test is accurate?²⁵
- a. Not at all confident
 - b. Somewhat confident
 - c. Very confident

Table A1

	0 I do not use this information	1 Not useful	2 Slightly useful	3 Useful	4 Very useful
1. The content of the different sections of the TOEFL ITP test (Reading, Structure and Written Expression, Listening)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. How test takers can prepare for the TOEFL ITP test	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. How the TOEFL ITP test should be administered	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. How TOEFL ITP section and total scores are calculated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. How TOEFL ITP section and total scores are reported	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. The security of the TOEFL ITP test administration and score reports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. How long the TOEFL ITP test scores are valid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. The minimum TOEFL ITP test scores for entry into specific courses at your institution	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. The relationship between the TOEFL ITP test scores and scores on other English-language proficiency tests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. The validity ^a and reliability ^b of the TOEFL ITP test scores ²⁴	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. The recognition of the TOEFL ITP test locally and internationally	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

^aA language proficiency test is *valid* if it actually measures language proficiency as opposed to measuring something irrelevant to language proficiency (e.g., cultural knowledge). ^bA language proficiency test is *reliable* if it measures language proficiency in a consistent manner irrespective of when a student takes the test. For example, if somehow there was no change in the student's actual English-language proficiency, the student should receive similar scores if the test is taken twice.

14. Are you willing to participate in a follow-up interview?

- Yes
- No (please go to Question 16)

15. What is your preferred language for the interview? (Please respond only if you selected "Yes" in Question 14.)

- English
- Other (please specify):

16. Your personal information

- Name:
- Name of your institution:
- Job title:
- Department (or college/faculty/school):
- Country:
- E-mail:
- Phone number:

Appendix B

Sample Set of Interview Questions (#16)

- Can you tell me a little about your university and language center? Do you support just students at the university, or do you have other types of clients?
- Can you describe your role at the language center?
- Do you know how long your university has been using TOEFL ITP?
- You said that you use TOEFL ITP for a lot of different uses. Can you say more about what each use involves?

- a. To monitor students' progress
 - b. As an exit requirement from English-language programs
 - c. To admit students to short-term programs in an English-speaking country
 - d. To admit and place students in collaborative international degree programs
 - e. For scholarship programs
 - f. As a requirement for graduation from undergraduate and/or graduate programs
 - g. The selection of student exchanges between nations
5. You indicated that there is a minimum TOEFL ITP score of 450 in total. Can you say more about that?
 - a. Which use of the test of those you listed has the 450 minimum?
 - b. How are minimum scores chosen?
 - c. Who makes the decision?
 6. You said you believe the minimum scores are "about right." Can you say more about that? Why do you think they are "about right?"
 7. You said that TOEFL ITP scores are "very useful" for showing students' English-language ability. Can you talk about why you think so?
 8. How often does your language center administer the TOEFL ITP test and to how many students (approximately)?
 9. There was a question on the survey about if you use other tests in addition to TOEFL ITP. You said you use "test prediction of TOEFL." I am not sure what you meant by this. Can you say more please?
 10. You said on the survey that information about the minimum TOEFL ITP scores for entering specific courses at your university is "very useful" for you. Can you say how you use that information?
 11. You said that it is "useful" to know how TOEFL ITP test scores correlate to scores on other English-language proficiency tests. Can you say more about that? How do you use this information?
 12. You said that information about the recognition of the test locally and internationally is "very useful" for you. Can you talk about how you use this information?
 13. There was a question on the survey asking what other information about TOEFL ITP would be useful for you. You wrote something about TOEFL ITP being used for scholarships and student exchange. Can you say more about that? What kind of information is needed?
 14. Is there any other information or resources that you would like?
 15. Do you know what other teachers and students say about the test? Are their experiences and opinions about TOEFL ITP positive or negative?
 16. Overall, how well do you feel that TOEFL ITP scores meet your expectations? Do you expect your language center to continue using the test?
 17. Is there any other feedback you would like to share about your experience with TOEFL ITP?

Suggested citation:

Golubovich, J., Tolentino, F., & Papageorgiou, S. (2018). *Examining the applications and opinions of the TOEFL ITP® assessment series test scores in three countries* (TOEFL Research Report No. RR-84). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12231>

Action Editor: Donald Powers

Reviewers: Yeonsuk Cho and Margarita Olivera Aguilar

ETS, the ETS logo, MEASURING THE POWER OF LEARNING., TOEFL, TOEFL iBT, TOEFL ITP, and TOEIC are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>