

TOEFL[®] Research Report

TOEFL–RR-101

ETS Research Report No. RR–24-01

The Use of TOEFL iBT[®] in Admissions Decisions: Stakeholder Perceptions of Policies and Practices

Sara T. Cushing

Haoshan Ren

Yi Tan

January 2024

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 (2023–2024) members of the TOEFL COE are:

Lorena Llosa – Chair

Beverly Baker
Tineke Brunfaut
Bart Deygers
Atta Gebril
Yo In'Nami
Talia Isaacs
Gary Ockey
Anamaria Pinter
Koen Van Gorp
Wenxia Zhang

New York University, USA
University of Ottawa, Canada
Lancaster University, UK
Ghent University, Belgium
The American University in Cairo, Egypt
Chuo University, Japan
University College London, UK
Iowa State University, USA
University of Warwick, UK
Michigan State University, USA
Tsinghua University, China

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

The Use of TOEFL iBT® in Admissions Decisions: Stakeholder Perceptions of Policies and Practices

Sara T. Cushing, Haoshan Ren, & Yi Tan

Georgia State University,

This paper reports partial results from a larger study of how three different groups of stakeholders—university admissions officers, faculty in graduate programs involved in admissions decisions, and Intensive English Program (IEP) faculty—interpret and use TOEFL iBT® scores in making admissions decisions or preparing students to meet minimum test score requirements. Our overall goal was to gain a better understanding of the perceived role of English language proficiency in admissions decisions and the internal and external factors that inform decisions about acceptable ways to demonstrate proficiency and minimal standards. To that end, we designed surveys for each stakeholder group that contained questions for all groups and questions specific to each group. This report focuses on the questions that were common to all three groups across two areas: (1) understandings of and participation in institutional policy making around English language proficiency tests and (2) knowledge of and attitudes toward the TOEFL iBT test itself. Our results suggested that, as predicted, university admissions staff were the most aware of and involved in policy making but frequently consulted with ESL experts such as IEP faculty when setting policies. This stakeholder group was also the most knowledgeable about the TOEFL iBT test. Faculty in graduate programs varied in their understanding of and involvement in policy making and reported the least familiarity with the test. However, they reported that more information about many aspects of the test would help them make better admissions decisions. The results of the study add to the growing literature on language assessment literacy among various stakeholder groups, especially in terms of identifying aspects of assessment literacy that are important to different groups of stakeholders.

Keywords language assessment literacy; language testing policies; TOEFL iBT®; TOEFL® paper-based test; stakeholders; university admissions officers; graduate program faculty; Intensive English Program (IEP) faculty; admissions decisions; test preparation; proficiency; standards; surveys; policy making; Duolingo English Test; Pearson Test of English

doi:10.1002/ets2.12375

Introduction

Over the past few decades, as many colleges and universities in the United States have sought to increase their international student populations, providing evidence of English language proficiency has become an important component of a prospective student's application dossier. Over time, institutions have developed policies and procedures around English language proficiency assessment, including decisions about which tests to accept, what minimum scores to accept, and alternatives to standardized proficiency assessments such as completing a course of English study, in an effort to ensure that international students have sufficient language proficiency to be successful in a degree program.

Although the TOEFL® test, in its original paper form and now as the TOEFL iBT® test, has been the dominant proficiency test in the United States for decades, with IELTS as a second major option, other test developers including Pearson and Duolingo English Test have started promoting their tests (Pearson Test of English [PTE] and Duolingo English Test [DET]) as more convenient or less expensive alternatives to TOEFL iBT and IELTS. As Ginther and Elder (2014) noted, the entry into the market of new proficiency tests means that test score users need to be well equipped to understand the claims of test makers in deciding which tests to accept as evidence of language proficiency and what information is necessary to make decisions about minimum test scores. That is, decision makers and other stakeholders require a certain level of assessment literacy (Stiggins, 1991; Taylor, 2009, 2013) to revise and adjust their policies in light of new information. At the same time, test developers and researchers should “be sensitive to many different types of test stakeholder and the varying ways in which they find themselves engaging with and understanding assessment issues” (Taylor, 2013, p. 407).

Corresponding author: S. Cushing, email: stcushing@gsu.edu

In the United States, admissions decisions are typically made by different groups of stakeholders at the undergraduate and graduate levels. For undergraduates (both domestic and international), decisions are most often made centrally, within an office of admissions. Specialists in international admissions are thus an important stakeholder group, and this group is typically well trained in interpreting and using proficiency test scores.

In contrast, for many graduate programs, admissions decisions are made by department faculty themselves, sometimes following an initial screening process by a graduate admissions office at the college or university level. Given that faculty in graduate programs, particularly at the doctoral level, work closely with students and provide intensive mentoring to them, they may have more of a personal investment in knowing how test scores relate to academic performance than admissions professionals yet be less equipped with information about English proficiency testing that may be useful in their decision making.

One relatively common administrative structure is to designate a director of graduate studies (DGS) or graduate coordinator from within the faculty, who is tasked with leading the admissions process. A DGS may or may not have experience with, knowledge about, or interest in language proficiency tests or the testing policies of their institution but nevertheless needs to rely on test scores as one piece of evidence to take into consideration when deciding whether or not to admit a student. Thus, DGSs (or their equivalent) are another important stakeholder group, and a group that has been relatively underresearched.¹

A third stakeholder group whose voices are important in admissions policies, even if they are not directly involved in admissions decisions, is English language teaching professionals. At many institutions, an alternative to a language proficiency score for admissions is successful completion of a program of English study, typically through an Intensive English Program (IEP). There are many different models for how IEPs are administered and where they are housed in an institution, but IEPs generally are staffed by faculty with a master's degree in Teaching English as a Second Language (TESOL) or equivalent and several years of teaching experience in higher education.

There are two important reasons why IEP faculty perceptions of language tests and testing policy are relevant to an investigation of how language tests factor into admissions decision making. First, as Green (2018) noted in her study of assessment practices in IEPs, exit decisions (i.e., how instructors determine that students have sufficient English proficiency to matriculate into a degree program) are underresearched. IEPs are frequently autonomous units within a university and have great latitude for setting their own standards for successful completion and thus for certifying that students have attained a level of English proficiency that can be considered equivalent to the level represented by a minimum score on a language proficiency test. At present, there is little empirical evidence to back up the claim that these two paths to admission are equivalent.

Second, as experts in English language teaching, IEP faculty often serve as resources for other personnel at their institution who may have questions about minimum levels of proficiency recommended for academic study, specific language tests, and other factors that might impinge on admissions decisions. In addition, IEPs frequently offer test preparation courses to their students, and thus IEP faculty may have more up-to-date understanding of test structure and content than the test users (i.e., decision makers) themselves. As Deygers and Malone (2019, p. 364) noted, admissions personnel “seldom act independently and may rely on feedback from a variety of stakeholders from within and outside the university system.” Deygers and Malone thus recommended investigations of “beliefs of stakeholders who advise and influence the final policy makers. Such research could result in a better understanding not only of who the potential decision-makers are, but also how and when they can influence policy.” For these reasons, including the perceptions of IEP faculty in a study of stakeholders may provide a richer understanding of the factors that go into institutional policies on language tests and their implementation.

Numerous scholars have called for more studies of the assessment literacy of various stakeholder groups so that language testing professionals can collaborate effectively with these groups. By the same token, some researchers (Deygers & Malone, 2019; Deygers & Vanbuel, 2022) have begun to note the importance of policy literacy (Lo Bianco, 2001) for language testing professionals. That is, for language testing research to have impact on policies, researchers need to understand how policies are created and implemented in light of real-world constraints and considerations. Similarly, for test development agencies to be able to provide useful and timely information to score users about their tests, it may be helpful to know more about the processes and people involved in setting policies and what other sources of information are being consulted. For example, admissions officers may consult with English language teaching professionals at their institution on issues related to proficiency tests, but these professionals may not have testing expertise or might not have kept up

with developments in language testing. In such cases, it may be as important for test developers to engage with English language teachers who are advising admissions officers as it is to engage with the admissions staff themselves. With regard to graduate admissions, it may be useful to know how graduate directors in various fields view the usefulness of English language tests in admissions decisions vis-à-vis other evidence of readiness for graduate study.

This paper presents selected results from a larger study of the perceptions of these three stakeholder groups—admissions professionals, graduate directors, and IEP faculty—on the use of TOEFL iBT in admissions decisions. We were interested in exploring four main topics: (a) how English language proficiency in general, and TOEFL iBT scores in particular, factor into admissions decisions; (b) the internal and external factors that inform decisions about accepting specific proficiency tests and the setting of minimum test scores; (c) the degree to which different stakeholder groups are aware of language testing policies at their institutions and how they are created; and (d) whether the three stakeholder groups have a shared understanding of the structure of the TOEFL iBT and the interpretation of test scores. In this paper, we focus on the third and fourth areas, reporting on survey results that explore what stakeholders understand about how language policies are created at their institutions, along with their knowledge and beliefs about the usefulness of the TOEFL iBT test scores in admissions decisions. In other words, we wanted to investigate both the assessment literacy and policy literacy of these stakeholder groups and understand how these areas of knowledge impact decisions about admissions.

Literature Review

As Deygers and Malone (2019) noted, most assessment literacy studies have focused on teachers, but researchers have begun to focus on what test users, particularly administrators responsible for admissions decisions, know about proficiency tests (Malone & Montee, 2014) or need to know about assessment to use test scores appropriately (Baker, 2016; Baker et al., 2014; Ginther & Elder, 2014; O’Loughlin, 2011, 2013). A few studies have taken a case study approach to the use of proficiency tests in admissions decisions. For example, O’Loughlin (2011) used institutional documents, surveys, and interviews to explore how administrative and academic staff at an Australian university used IELTS scores to select international students and what knowledge and beliefs about IELTS they had. They found that use of IELTS entry scores was based not on any understanding or implementation of a formal standard setting process but on “both a shared sense across Australian universities of the required English standards and prevailing market forces (p. 141);” that is, the entry score was set at a lower than optimal level, presumably because of competition for international students with other institutions. O’Loughlin’s (2011, 2013) research pointed out a number of areas in which respondents’ knowledge of IELTS was lacking and concluded that, in order for the interpretation and use of test scores to be valid and ethical, policy makers and other stakeholders “need to be better educated about the interpretation and use of test scores and other measures of proficiency” (p. 15).

Exploring university admissions in Flanders, Belgium, from a policy literacy perspective, Deygers and Vanbuel (2022) similarly found that admissions policies were not based on empirical evidence but rather informed by “common sense beliefs, institutional politics ... market dynamics ... and realpolitik” (p. 584).

In another case study approach, Ginther and Elder (2014) surveyed and interviewed faculty at one United States and one Australian university to explore levels of knowledge about test scores and their use in making decisions about graduate student admission. They found that the assessment literacy of participants was “limited” (Ginther & Elder, 2014, p. 26), particularly in terms of knowing how to use test scores beyond the minimum cutoff score and their understanding of current approaches to testing. Furthermore, they found that external factors such as economic and political exigencies often overrode considerations of test validity in influencing test use. Their study was limited to two institutions, however, and they conceded that additional confirmatory research was needed before their findings could be generalized.

Little if any research exists on the perceptions of international admissions officers regarding the interpretation and use of English proficiency tests in the United States, although at least one study (Newcomb, 2017) has delved into the experiences and training of international admissions officers. One recent dissertation (Wilson, 2021) explored the role of admissions officers in deciding whether to accept the DET for admission at the start of the COVID-19 pandemic and discussed the “potential consequences that may arise when test users lack a degree of language assessment literacy” (p. 5).

In the Australian context, O’Loughlin (2013) surveyed a sample of admissions staff at two Australian universities to find out what information about the IELTS they needed to do their jobs and found that most staff were not required to use test scores to make informed judgments about readiness for study; they simply needed to determine whether an applicant

met the minimum standard for admission. It is this situation that presumably prevails for most undergraduate admissions officers in the United States, but this is an empirical question.

In summary, research has suggested that stakeholders in admissions decisions may have limited knowledge about language tests beyond what is needed for their immediate needs and that aspects of test validity that are centrally important to language testing professionals often take a back seat to other concerns, such as enrollment management, when setting policies on admissions and using and interpreting language test scores. However, this research has been primarily conducted as case studies and thus further research is warranted to ultimately be able to make generalizations.

Our study contributes to this literature through surveys of three stakeholder groups described above: admissions officers responsible for international admissions at the undergraduate level; graduate directors who make admissions decisions for their own programs; and IEP faculty, whose expertise in English language teaching is frequently called upon in setting policy but who can also recommend students for matriculation in lieu of a proficiency test.

Our study explored several issues related to English proficiency in admissions decisions,² including questions specific to each stakeholder group. For the purposes of this report, we focus on the following specific research questions that address the factors of knowledge dissemination and beliefs:

1. What do participants know about language policies at their institution?
 - a. What tests are accepted, and what are the minimum test scores?
 - b. Who is responsible for setting these policies?
 - c. What factors are considered when setting policy, and who is consulted?
2. What knowledge, beliefs and attitudes do participants have toward the TOEFL iBT test?
 - a. What do participants report knowing and believing about the TOEFL iBT test?
 - b. To what extent are these areas of knowledge and belief consistent across the three stakeholder groups?
 - c. What do participants feel they need to know in order to make appropriate admissions decisions and/or prepare students for matriculating?

Methodology

Survey Development

Our original plan was to conduct face-to-face focus groups with our three stakeholder groups (admissions officers, graduate directors, and IEP faculty) at three different institutions as in-depth case studies of (a) how these stakeholder groups interacted with each other to establish, revise, and implement English language policies and (b) the degree to which there were shared understandings of the level of English proficiency needed for academic success within each institution. We planned to follow up these case studies with surveys targeted at larger samples from each group, investigating our original research questions along with any that emerged from the focus groups. However, the COVID-19 pandemic precluded travel for research, and we were unable to arrange for in-depth data collection at specific institutions. Instead, we concentrated our initial efforts on developing surveys to investigate our research questions, conducting a series of focus groups and interviews with individual members of each stakeholder group to gather information that would be useful in refining our research questions and survey items.

Through personal contacts, we were able to arrange three focus groups with IEP faculty members at different institutions. It was more challenging to arrange focus groups with admissions professionals, in part because we had very few personal contacts we could call upon, and the international admissions staff at our own university declined to participate, citing a heavy workload. We were able to arrange one focus group of three admissions professionals in Georgia, but ultimately only two attended. We then interviewed an additional admissions professional from an institution in the Midwest using the same format. That respondent also provided us with insights into our draft survey and helped us revise our questions both at that interview and later through email.

Finally, we held three focus groups with graduate directors from a range of disciplines (business, biology, computer science, music, and applied linguistics) at our own institution. The faculty from the business department informed us that they relied heavily on international recruiters to screen prospective students and evaluate their language proficiency. We therefore scheduled an interview with two of these recruiters to explore their ideas about language proficiency assessments. We also interviewed the dean and one staff member in the graduate school of our institution for their insights because the

graduate admissions process begins in the graduate school. Although these interviews were fruitful and suggested that college or university-level graduate admissions specialists might be another stakeholder group, we ultimately decided not to include them as a separate group for this study. The focus groups and interviews are summarized in Table 1.

Based on our review of the relevant literature and information we received from participants from our main stakeholder groups as described above, we created, pilot tested, and distributed surveys targeted at each group (Surveys can be found here: <https://tinyurl.com/5xzb6ka>). The surveys contained several sections that were parallel, if not identical, across the three groups.

Surveys were administered via the Qualtrics platform. In addition to demographic information about the respondent's institution, each survey had the following sections: knowledge of English language testing policy, awareness of English language support for matriculated students, and understanding of TOEFL iBT test content and score interpretation. The graduate director and IEP faculty surveys also had specific questions relevant to specific issues that affected only them; results of these sections are reported elsewhere. Lastly, respondents were asked to evaluate and comment on one TOEFL iBT independent essay response in terms of the writer's perceived readiness to begin an academic program and estimated total TOEFL iBT score range. At the end of the survey, participants were invited to complete a second survey, in which they could evaluate up to eight additional essays for a small honorarium. However, due to space limitations, we are not reporting on the writing evaluations in this report.

Participants

As described previously, three groups of participants were recruited to participate in the study: faculty from IEPs where completion of the IEP satisfies the English language proficiency requirement, staff from undergraduate admissions office involved in international admissions decisions, and faculty who make decisions about graduate admissions (typically, graduate directors from individual programs). Participants were recruited through listservs, personal contacts, and examination of university websites for relevant names and contact information. Graduate directors were recruited primarily through individualized emails, while IEP faculty were recruited first through relevant listservs and later through emails to IEP directors who were asked to share the survey with their faculty. Admissions faculty were recruited mainly through bulk (i.e., nonindividualized) emails to a list of US-based admissions officers that had been shared with us by another researcher. Recruitment began in February 2022 and was completed in July 2022. Not everyone who began the survey completed it; however, when possible, we included data from incomplete surveys (e.g., data from those participants who completed the sections of the survey that were relevant to this report). Table 2 summarizes the survey responses from each participant group.

In total, 266 participants from 90 institutions began the surveys, but only 219 completed them (see Table 2 for a summary). Since respondents were recruited through a variety of means, it is impossible to calculate an exact response rate. However, we can estimate response rates from the number of individual emails sent out. For graduate directors,

Table 1 Focus Group and Interview Summary

Description	UA	GD	IEP	Graduate admissions and recruiters
Number of groups	2	3	3	2
Number of participants	3	7	15	4
Number of institutions	3	1	3	1

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs.

Table 2 Survey Respondents

Item	Main survey				Second survey (Essay evaluation)		
	UA	GD	IEP		UA	GD	IEP
Started	73	147	46	Started	30	42	22
Completed (>50%)	59	125	35	Completed (4+ essays)	26	32	16

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs.

Table 3 Institutions Represented in the Main Surveys

Trait	UA (<i>n</i> = 50)	IEP faculty (<i>n</i> = 23)	GD (<i>n</i> = 35)
Location			
New England & Mid East	10 (20%)	5 (22%)	8 (23%)
Great Lakes & Plains	13 (26%)	3 (13%)	9 (26%)
Southeast	11 (22%)	8 (35%)	9 (26%)
Southwest & Rocky Mts	3 (6%)	4 (17%)	4 (11%)
Far West	13 (26%)	4 (17%)	4 (11%)
Control			
Public	29 (48%)	16 (70%)	27 (77%)
Private	21 (42%)	7 (30%)	8 (23%)
Selectivity of UA program			
Inclusive	9 (18%)	5 (22%)	6 (17%)
Selective	19 (38%)	10 (44%)	9 (26%)
More selective	22 (44%)	8 (35%)	20 (57%)

Note. UA = undergraduate admissions; IEP = intensive English programs; GD = graduate directors.

374 emails were sent out and 125 usable responses were received for an approximate response rate of 33%. For admissions officers, we sent bulk emails to a list of 474 names. The response rate for this group was closer to 12%. Finally, since IEP faculty were recruited primarily through listservs, it is difficult to estimate a response rate. However, we were aware of 55 IEPs that met our criteria and received responses from at least one person at 23, or 42%, of these programs. For the second (optional) survey, 94 agreed to take the survey and 74 completed enough to qualify for an honorarium.

Institutions Represented

Respondents were asked to identify their institutions in the survey, but not all respondents chose to do so. A total of 90 institutions were identified by at least one participant in one or more of the three surveys. Table 3 below describes the demographic characteristics of these institutions based on the 2021 Carnegie Classifications of Institutions of Higher Education (<https://carnegieclassifications.acenet.edu/>). Note that this table refers to institutions, not individual respondents, as there were multiple respondents from several universities.

As the table shows, all regions of the continental United States were represented in all three groups. For IEPs and graduate directors, the majority of respondents worked at public universities, whereas the undergraduate admissions officers were more evenly balanced. As for selectivity, the sample leaned heavily toward more selective institutions, except for IEP faculty.

Data Analysis

Because this is an exploratory study, we provide summaries of the responses to the relevant survey items. However, in a few cases where it is relevant to our research questions, we also conducted one-way ANOVAs across the three stakeholder groups. Although we recognize that Likert-type rating scale scores are not interval data, we follow Norman (2010), who argued that ANOVA is robust to violations of nonnormality and thus can be used with Likert-type items.

Results

RQ1: Testing Policies

All three surveys included similar questions about policies related to language tests in admissions: which English language proficiency tests are accepted for admission, what the minimum scores are, and who is responsible for setting policies in these areas. Our main interest was not in making generalizations about specific policies but in the degree to which participants from the three groups were aware of policies and how they are set. Our assumption, based on our focus groups and literature review, is that international undergraduate admissions officers would be most familiar with such policies

Table 4 Percentage of Respondents Reporting Acceptance of Different English Language Proficiency Tests for Admissions

Test	UA (<i>n</i> = 60)	GD (<i>n</i> = 124)	IEP (<i>n</i> = 35)
TOEFL iBT test	100%	98%	100%
TOEFL paper-based test ^a	67%	69%	49%
IELTS	98%	84%	94%
Pearson Test of English (PTE)	62%	15%	46%
Duolingo English Test (DET)	93%	32%	57%
Other (please specify)	25%	7%	31%

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. ^aWe included TOEFL paper-based test based in our focus group results, even though this name is outdated.

and graduate directors the least. We therefore focus our discussion here not so much on the actual survey responses, but the degree to which responses differ across groups.

Our first question about testing policies was the following:

To the best of your knowledge, what tests does your institution accept as evidence of English language proficiency for undergraduate students (UA/IEP)/graduate students (GD)? Check all that apply.

Table 4 summarizes the responses to these questions. As the table shows, all three groups reported accepting the TOEFL iBT test, and the vast majority reported accepting IELTS. There was less consistency between groups, particularly between admissions officers and graduate directors, on some of the other tests. Specifically, the majority of admissions professionals reported accepting the PTE and DET, while fewer graduate directors reported accepting these tests. While it may well be that graduate admissions policies differ from undergraduate policies, a perhaps more likely explanation for this discrepancy is that graduate directors may be unfamiliar with these tests and were therefore disinclined to select them as options.

Even though we recognize that there are only very limited circumstances under which a paper format of TOEFL is used, we included the TOEFL paper-based test as an option on our surveys based on comments by our focus group participants, many of whom began their careers before the advent of the TOEFL iBT and have not followed the evolution of the TOEFL® products closely; in fact, some respondents referred to levels of proficiency in terms of the original TOEFL scale (e.g., “we look for a 550 TOEFL score”). In our survey results, graduate directors were somewhat more likely to state that TOEFL paper-based tests were accepted for admission than other participants, which may suggest less awareness of the changes in the TOEFL family of tests over the past three decades.

For those who selected “other” on this question, space was provided to name additional tests. The most frequent tests mentioned were SAT® or ACT (9), the CAE (now known as the Cambridge C1 Advanced Exam), and ITEP. The TOEFL® Essentials™ test was mentioned by two participants (one GD and one IEP) and the TOEFL iBT Home Edition by one UA respondent. UA and IEP participants were far more likely to provide names of additional tests than graduate directors, as the table shows.

Following this question, as well as others where we were not sure our list of options was comprehensive, participants were presented with an opportunity to clarify their responses; for example, the next survey question said, “Please clarify your answer to the previous question (optional).” Only four undergraduate admissions staff offered comments on this question. Two referred to the paper-based TOEFL (e.g., “I know the paper-based test is phasing out but results are valid 2 years.”) and two mentioned other ways of demonstrating proficiency (e.g., “If a student has taken an SAT or ACT exam and met a certain exam score, we would consider that as proficient as well.”).

There were 22 open-ended comments from graduate directors. Even though the question asked for clarification on specific tests, five participants commented that policies were set by the university or graduate school (e.g., “These are decided by the Graduate College, not our program. We do not have any specific requirements concerning English.”). Five expressed uncertainty about whether some of the tests would be accepted (e.g., “I’m actually not sure we *don’t* accept Pearson or Duolingo. But I’ve never seen them used.”).

Two participants took the opportunity here to express their personal opinions about the DET (in all, seven responses mentioned DET, by far the test most often mentioned). The first was “I argued for Duolingo and lost,” while the second

Table 5 Unit(s) Responsible for Setting Policies About Accepting Tests of English Proficiency for Admissions

Unit	UA (<i>n</i> = 64)	GD (<i>n</i> = 125)	IEP (<i>n</i> = 35)
State-level policy (Board of regents or equivalent)	6%	7%	11%
Chief academic officer (Provost) or equivalent	30%	12%	11%
Faculty/University senate	17%	14%	9%
Admissions office or equivalent	80%	38%	54%
Each college makes its own decision about what tests to accept	3 %	8%	6%
Graduate school ^a	-	28%	-
Each department makes its own decision about what tests to accept	8%	17%	6%
Other (please describe below)	8%	18%	6%
I don't know	2%	22%	34%

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. ^a This answer was volunteered by many participants in the graduate director survey.

stated: “Our institution made a policy of accepting Duolingo English Test scores during the pandemic and decided to continue this policy moving forward. We have been attempting to push back against this, and we’ve decided as a department not to list DET scores as an option.”

The next question was the following:

To the best of your knowledge, who is responsible for setting policies about which tests of English proficiency are accepted for undergraduate admission (UA/IEP)/graduate admission (GD) at your university? Check all that apply.

Participants were allowed to select more than one answer, so the reported percentages represent the percentage of respondents selecting each option and do not add up to 100%. Again, we were interested both in the variety of answers along with the degree of familiarity with the policies. As Table 5 shows, for all three groups, the admissions office or equivalent was most often mentioned. The second most frequent response for undergraduate admissions participants was the provost or equivalent, with others mentioning the faculty/university senate, suggesting some academic oversight of the process. Among graduate directors, the second most frequent response was not one of the options on the survey but was volunteered in the comments: 35 participants (28%) stated that the graduate school or equivalent at their university was responsible for these decisions. A substantial number of respondents, particularly among graduate directors and IEP faculty, selected “I don’t know” as a response.

Here again, participants were offered an opportunity to clarify their response. Forty-seven graduate directors commented, most of them stating that decisions were made at the level of the graduate school, as noted above. However, some respondents expressed uncertainty about this point. For example, one respondent wrote: “I assume it is the university’s Graduate School.” A few responses outlined the process in more detail as in the following comment: “These scores are in our university rules. They are recommended by Graduate Council and then voted upon by Graduate Faculty, Faculty Senate and Board of Trustees.”

A few respondents commented on the minimum scores rather than on the tests here, as in this response: “The Graduate School establishes which tests may be used and minimum scores, but colleges and departments may establish greater rigor.”

We then asked the following question: “Do you know how *minimum TOEFL iBT scores* are set at your institution? If you answer yes, we will ask you about it in the next question.” Participants who answered yes to this question were then asked: “Who is responsible for setting policies on *minimum TOEFL iBT scores* for undergraduate admission at your institution? Check all that apply.” Accordingly, the percentages in the first row of Table 6 refer to percentages of the total sample that claimed to have knowledge of how scores are set and thus being presented with the options. As the table shows, and as we expected, undergraduate admissions officers were much more likely than participants from the other two groups to report knowing how minimum scores were set. While 66% of undergraduate admissions officers responded yes to the knowledge question, only 40% of IEP faculty and 19.9% of graduate faculty did so, reflecting their relative lack of closeness to the issue.

In all three groups, the admissions office was identified most often as being responsible for setting minimum standards. Not surprisingly, graduate directors were most likely to respond that decisions were made at the college or department

Table 6 Unit(s) Responsible for Setting Policies About Minimum Test Scores

Unit	UA (<i>n</i> = 41, 66%)	GD (<i>n</i> = 22, 20%)	IEP (<i>n</i> = 14, 40%)
State-level policy (Board of regents or equivalent)	5%	5%	14%
Chief academic officer (Provost) or equivalent	22%	18%	36%
Faculty/University senate	12%	9%	7%
Admissions office or equivalent	95%	50%	64%
Each college makes its own decision about what tests to accept	10%	18%	0%
Each department makes its own decision about what tests to accept	5%	23%	0%
Other (please describe below)	15%	18%	14%
I don't know	5%	5%	14%

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs.

Table 7 Open-Ended Responses to the Question: Have you personally been involved in standard setting?

Group	Responses	Nature of comments
UA	Yes: 26 No: 7	Work with units on campus (6) Consult with other institutions (5) Review materials from testing companies (2) Look at concordances between tests (3) Use internal data (3)
GD	Yes: 8 No: 10	Only those with specific expertise responded
IEP	Yes: 10 No: 2	Consulted by other units on tests and cutoffs

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs.

level, with no IEP faculty choosing those options. Of course, the sample sizes are relatively small for these questions as they were presented only to those who claimed to have knowledge of the process, so these results should be viewed with caution.

Participants were next asked whether they had been involved in setting English proficiency standards: *Have you personally been involved in setting English language proficiency standards for any program at your university? If so, please describe your experience briefly.* Because this was an open-ended question rather than a yes/no choice, many participants (the vast majority, in the case of graduate directors) chose not to respond at all, perhaps implying a negative answer. A summary of the responses among the three groups are found in Table 7.

Among admissions professionals, 15 answered yes, either without specifying further or by naming a specific test that they had been involved in discussions about, while seven said no. Eleven others said yes, while providing more detail of various kinds. Six mentioned specific units that they worked with (departments, the English Language Center, the Office of Institutional Research) and five mentioned that they consulted with other institutions. In terms of data that they consulted, two mentioned information from testing companies, three mentioned correlations between tests, and three mentioned internal data.

Among IEP faculty, 12 provided responses to this question, 10 in the affirmative and two in the negative. The 10 who responded in the affirmative all stated that they were frequently consulted by other officials in the university. Example comments include the following:

- “Yes. Consult with departments to determine what level of English language proficiency is needed to successfully complete their program and be successful in their work upon graduation. Based on that information, cutoff scores are set.”
- “Yes, I am frequently asked about new tests - whether I would recommend to accept them (i.e., Duolingo, at home TOEFL, my best scores, etc.) and score recommendations.”

Among graduate directors, only 18 provided answers to this question, of which 10 simply responded “no.” Only three respondents provided specific answers describing their participation, as in this example: “As Associate Dean, I researched

Table 8 Factors for Setting Minimum Scores (Undergraduate Admissions [UA] and Intensive English Programs [IEP])

Factor	Group	I don't know	Not very important	Somewhat important	Very important	N(n)
State-level policy (Board of regents or equivalent)	UA	14%	66%	18%	16%	58(50)
	IEP	47%	47%	29%	24%	32(17)
Formal standard setting procedures	UA	9%	18%	45%	37%	56(51)
	IEP	42%	11%	50%	39%	31(18)
Information from test publishers	UA	5%	6%	51%	44%	58(55)
	IEP	34%	0%	52%	48%	31(20)
Advice from language teaching or testing expert(s) at your institution	UA	5%	9%	40%	51%	58(55)
	IEP	26%	9%	30%	61%	32(24)
Admissions practices at similar universities	UA	2%	0%	31%	69%	59(58)
	IEP	30%	0%	39%	61%	33(11)
Admissions practices at more selective universities	UA	5%	12%	50%	33%	58(55)
	IEP	34%	0%	52%	48%	32(21)
Enrollment management decisions (e.g., increasing or decreasing international student enrollment)	UA	5%	19%	42%	33%	57(54)
	IEP	33%	9%	41%	50%	33(22)
Pressure from applicants	UA	2%	67%	28%	4%	57 (56)
	IEP	44%	50%	44%	6%	32(18)
Other (please describe below)	UA	47%	12%	0%	88%	15(8)
	IEP	77%	50%	0%	50%	9(2)

Note. N represents the total of those responding to the item, and (n) represents those who did not answer “I don’t know.” Percentages in **bold** indicate the most frequent response.

the various tests, scores at dozens of universities, including our benchmarking institutions and presented the information to Graduate Council for them to make an informed decision.”

One question in this section was addressed to undergraduate admissions staff and IEP faculty but not to graduate directors: “*To the best of your knowledge, how important are the following factors when determining minimum English language proficiency standards and/or accepting new tests at your university?*” Table 8 displays the results, with the most frequent response in bold face in each row. Note that the percentages of respondents in the columns representing the relative importance of factors have been adjusted to exclude those answering, “I don’t know.” As the table shows, the factor that was judged the most important for both groups was “admissions decisions at similar universities,” with approximately two-thirds of respondents rating this as “very important.” For IEP faculty, this factor was tied with “Advice from language teaching or testing expert(s) at your institution.” IEP faculty, though a small sample, tended to rate the importance of admissions practices at more selective institutions, enrollment management decisions, and pressure from applicants higher than did admissions officers, reflecting perhaps some skepticism of the admissions process from some survey respondents.

Three admissions officers made clarifying comments to this question. In all three cases, the respondents stated that their standards were high due to faculty input or the selective nature of the institution. For example, one respondent wrote: “For ‘Other’ there is a lot of pressure from faculty to keep our TOEFL minimum quite high (95). Their reasoning as I understand it is that the College does not offer any language acquisition courses, nor any writing classes for language learners, so the student needs to be able to succeed pretty independently from day 1.”

Among four IEP faculty providing clarification comments, two took somewhat critical stances toward testing and/or admissions practices. One wrote:

“Assessments should have standards and institutional monetary concerns are irrelevant. Sure, it’s a business model but it is a poor one. Some ivy league schools do not have IEPs or lower levels of IEP and yet it does not deter applications. Having an appropriate bar ensures student success and reduces frustrations with having to take IEP courses. Those making decisions do not see the impact it has locally on students, faculty, and graduation success when standards are compromised to lower acceptance requirements (this applies to domestic students as well).”

Another had this to say:

“The school is now [SAT/ACT] test-optional (following others), so HS transcripts and references are more important than ever — even when considering English proficiency. It’s a liberal arts school, so we still look for well-rounded students who have used English in a purposeful way outside of the classroom (e.g., community service, civic engagement, athletics, etc.). Applicants requested/demanded that the college accept the DET because other schools were accepting it. TOEFL Essentials was also accepted, but it was not widely available. Cut-Scores at peer or aspirant institutions are/were very important. Any information from the test publisher needs to be written for non-experts who increasingly question the value of tests. How does the test address DEI and social justice issues? How is stereotype threat minimized? etc etc.”

To summarize briefly, the responses to this set of questions across the three stakeholder groups showed variations in knowledge about university policies regarding the acceptance of specific tests and setting minimal standards and how such policies are set. Not surprisingly, admissions professionals were the most informed about these policies, with graduate directors the least informed. IEP faculty reported being consulted frequently in their areas of expertise, though at least some of these respondents expressed some skepticism over how decisions on policies are made.

RQ2: Familiarity With the TOEFL iBT Test

Our second research question addresses the degree to which stakeholder groups have similar understandings and beliefs about the TOEFL iBT test. To address this question, we first asked about participants’ familiarity with various aspects of the TOEFL iBT test. They were asked to rate their familiarity with these aspects of the test on a scale from 1 to 5, 1 being “not familiar at all” and 5 being “extremely familiar.” Following this question, we asked about the importance of this knowledge for admissions decisions (in the case of undergraduate admissions and graduate directors); for IEP faculty, we asked more generally how important the knowledge was for IEP faculty, also using a 5-point scale, with 1 labeled as “not at all important” and 5 as “extremely important.” These questions were worded as follows:

- “How familiar are you with the following aspects of the TOEFL iBT test?” [identical for all three groups]
- “For you to make informed admissions decisions, how important is it to be familiar with the following aspects of the TOEFL iBT test?” [Undergraduate admissions]
- “For you to make the best admissions decisions, how important is it to be familiar with the following aspects of the TOEFL iBT test?” [Graduate directors]
- “In your opinion, how important is it for your IEP faculty to be familiar with the following aspects of the TOEFL iBT test?” [IEP faculty]

Table 9 summarizes participants’ familiarity with aspects of the TOEFL iBT test, and Table 10 summarizes their views on how important it is to be familiar with them. Table 9 also includes the results of a one-way ANOVA comparing the familiarity of each aspect of TOEFL iBT across the groups.

As Table 9 shows, undergraduate admissions officers reported being familiar with the components of the test, the information in the score report, how to interpret total and part scores, the validity and reliability of the TOEFL iBT, and how long scores are valid, this last point being the highest scores. For ease of interpretation, mean scores of 3.5 or higher are presented in bold face. IEP faculty reported being familiar with the different components of the test, how students can prepare for the test, and how to interpret test scores. Graduate directors, in contrast, were not very familiar with any aspects of the test, exceeding mean scores of 2.5 only in knowing the information included in the score report, interpreting total and part scores, and knowing how long scores are valid. For the rest of the items, the mean scores were lower than 2.5, indicating that the majority of scores were 1 or 2. The mean scores for graduate directors were statistically significantly lower than those for admissions officers on all items, and lower than those of IEP faculty for all items except for knowledge about when TOEFL scores expire. IEP and admissions scores were similar on all items except the differences between the paper-based TOEFL test and the TOEFL iBT test and how students can prepare for the test (IEP faculty reported more knowledge on these items) and, similar to graduate directors, when test scores expire (IEP faculty knew less).

As for the importance of these aspects of knowledge, only admissions officers had any mean scores greater than 3.5, as shown in Table 10. IEP faculty, in particular, did have any mean scores over 3.5, in contrast to the previous table. Because the items were not identical across the three surveys, we could not perform any statistical tests to compare across groups.

Table 9 Familiarity with Aspects of TOEFL iBT

Structure	UA		GD		IEP		Sig. Diff.
	Mean	SD	Mean	SD	Mean	SD	
<i>Test structure</i>							
The different components of the test	3.5	0.9	2.5	1.3	3.8	1.2	GD < UA & IEP
How the TOEFL iBT test differs from the (old) TOEFL paper-based test	2.8	1.1	1.8	1.2	3.4	1.4	GD < UA < IEP
How the TOEFL iBT test differs from other standardized proficiency tests	3.1	1.0	1.8	1.2	3.3	1.2	GD < UA & IEP
How students can prepare to take the TOEFL iBT test	2.9	1.2	2.1	1.3	3.9	1.0	GD < UA < IEP
<i>Scoring and score reporting</i>							
How the test is scored	3.2	1.0	2.1	1.3	3.4	1.2	GD < UA & IEP
The information included in a score report	3.7	1.0	2.6	1.3	3.4	1.3	GD < UA & IEP
How to interpret total test score	3.7	1.0	2.7	1.4	3.6	1.2	GD < UA & IEP
How to interpret section scores	3.6	1.1	2.6	1.4	3.5	1.3	GD < UA & IEP
<i>Validity of test scores</i>							
Validity and reliability of the TOEFL iBT scores	3.6	1.0	2.3	1.3	3.4	3.1	GD < UA & IEP
How long scores are valid (i.e., when they expire)	3.8	1.2	2.7	1.5	3.1	1.3	UA > GD & IEP

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. Mean scores of 3.5 or higher are noted in bold.

Table 10 Perception of Importance of Aspects of the TOEFL iBT Test

Aspect	UA		GD		IEP	
	Mean	SD	Mean	SD	Mean	SD
<i>Test structure</i>						
The different components of the test	3.5	0.9	2.7	1.2	3.3	1.2
How the TOEFL iBT test differs from the (old) TOEFL paper-based test	2.8	1.1	2.2	1.1	2.5	1.2
How the TOEFL iBT test differs from other standardized proficiency tests	3.1	1.0	2.2	1.1	2.9	1.0
How students can prepare to take the TOEFL iBT test	2.7	1.1	2.2	1.1	3.5	1.2
<i>Scoring and score reporting</i>						
How the test is scored	3.5	1.0	2.6	1.2	2.9	1.0
The information included in a score report	3.7	0.9	2.9	1.2	2.9	0.9
How to interpret total test score	3.8	1.0	3.1	1.2	3.1	1.1
How to interpret section scores	3.7	1.0	3.0	1.3	3.0	1.2
<i>Validity of test scores</i>						
Validity and reliability of the TOEFL iBT scores	3.9	0.9	2.9	1.3	3.0	1.1
How long scores are valid (i.e., when they expire)	3.9	0.9	2.8	1.3	2.7	1.2

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. Mean scores of 3.5 or higher are noted in bold.

In order to compare the perceived importance of each factor with its perceived familiarity, we subtracted the mean familiarity score from the mean importance score for each group. Results are presented in Table 11; note that a positive difference indicates that the importance score is higher than the familiarity score, suggesting that more information in this area would be useful. Here there are striking differences among the groups. For admissions officers, the differences were small in most areas, though they felt it might be important to know more about how the test was scored (difference = .32), validity and reliability of the test (difference = .25), and how to interpret total scores (difference = .18). The only area in which their knowledge substantially exceeded their perception of its importance was in how to prepare for the test (difference = −.15).

For IEP faculty, perceived familiarity exceeded importance in every category, sometimes substantially, as in the difference between the old TOEFL paper-based test and the TOEFL iBT (difference = −.76). Other difference scores ranged from −.32 (similarity to other tests) to −.53 (information on score reports). This suggests that IEP faculty who participated in the study felt they knew more about the test than was necessary to fulfill their job requirements.

The opposite was true of graduate directors, whose ratings of importance were lower than those of familiarity, suggesting that they felt they would benefit from additional knowledge in every area (even if the absolute importance was not

Table 11 Differences in Mean Scores Between Familiarity and Importance

Trait	UA	GD	IEP
<i>Test structure</i>			
The different components of the test	−0.03	0.23	−0.47
How the TOEFL iBT test differs from the (old) TOEFL PBT test	0.05	0.36	−0.76
How the TOEFL iBT test differs from other standardized proficiency tests	0.00	0.44	−0.32
How students can prepare to take the TOEFL iBT test	−0.15	0.09	−0.44
<i>Scoring and score reporting</i>			
How the test is scored	0.32	0.57	−0.44
The information included in a score report	−0.02	0.31	−0.53
How to interpret total test score	0.18	0.39	−0.50
How to interpret section score	0.05	0.40	−0.44
<i>Validity of test scores</i>			
Validity and reliability of the TOEFL iBT scores	0.25	0.59	−0.35
How long scores are valid (i.e., when they expire)	0.07	0.13	−0.35

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. Positive numbers indicate that participants' ratings of familiarity are lower than their ratings of importance.

Table 12 Belief Statements About TOEFL iBT

Statement	UA (<i>n</i> = 54-55)	GD (<i>n</i> = 119)	IEP (<i>n</i> = 34-35)	Sig. Diff.
TOEFL [iBT] is trustworthy	4.49 (.57)	3.67 (.91)	3.94 (.94)	UG > GD,IEP
TOEFL [iBT] is useful for making admission decisions	4.56 (.57)	3.94 (.92)	3.74 (1.01)	UG > GD,IEP
TOEFL[iBT] is secure	4.26 (.81)	3.56 (.86)	3.86 (1.03)	UG > GD
TOEFL [iBT] measures proficiency accurately	4.25 (.62)	3.50 (.86)	3.72 (.99)	UG > GD,IEP
TOEFL [iBT] is more valid and reliable than other proficiency tests	3.33 (.86)	3.16 (.86)	3.13 (.86)	N/A

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs. The survey itself did not specify TOEFL iBT, but our data from focus groups and qualitative comments suggest that participants use the term "TOEFL" generically to refer to the TOEFL iBT.

as high as it was for the other two groups). The two areas with the largest difference between perceived importance and familiarity were validity and reliability (difference = .59) and how the test is scored (difference = .57).

Next, we asked participants to indicate the degree of their agreement with five statements about the TOEFL iBT test. Their responses are summarized in Table 12. One-way ANOVAs showed significant differences at $p < .001$ for all statements except one, with medium to large effect sizes. As the table shows, admissions officers had the highest rate of agreement, with averages above four on all questions apart from the last one. Graduate directors were most positive about the statement that TOEFL is useful for making admissions decisions, while IEP faculty had the highest mean scores for the statement that TOEFL is trustworthy. All three groups were relatively neutral on the statement that TOEFL is more reliable and valid than other proficiency tests; this was also the only item that did not differ significantly among the three groups, based on one-way ANOVAs. It may be that respondents (particularly graduate directors) may be unfamiliar with other tests and thus were not sure how to answer this question. Based on post hoc tests, undergraduate admissions officers had significantly higher levels of agreement than graduate directors on all the other statements and significantly higher levels of agreement than IEP faculty on all except the security of the TOEFL.

Finally, we asked participants where they got their information about the TOEFL iBT test. We offered five options and invited participants to comment on additional sources. The results are summarized in Table 13.

As the table shows, for UG admissions officers, the most frequent source of information by far was ETS materials or training (78%) with discussions with language specialists the second highest (48%). For IEP faculty, the main source of their knowledge was having taught TOEFL preparation courses (79.4), followed closely by materials from ETS (76.5%). Participants from both of these groups frequently reported having information from multiple sources.

For graduate directors, in contrast, the top source of knowledge (24%) was having taken the old TOEFL paper-based test themselves. This may not be surprising, given that faculty members advanced enough in their academic careers to

Table 13 Sources of Information About the TOEFL iBT Test

Source	UA (<i>n</i> = 50)	GD (<i>n</i> = 82)	IEP (<i>n</i> = 35)
Taken TOEFL iBT test	12%	13%	32%
Taken [old] TOEFL paper-based test	6%	24%	38%
Taught a TOEFL preparation class	12%	9%	79%
Received materials or training from ETS	78%	23%	76%
Consulted with language assessment specialist	48%	23%	59%
Other	26%	29%	18%

Note. UA = undergraduate admissions; GD = graduate directors; IEP = intensive English programs.

take on administrative roles such as graduate director might easily have entered their fields as international students prior to the introduction of the TOEFL iBT test in 2005. We return to this finding in the discussion section.

Discussion

To summarize the results of our study, our first research question asked what participants know about language policies at their institutions. Our hypothesis was supported that, among the three groups, undergraduate admissions officers would be most knowledgeable about language testing policies and who is responsible for setting them. As for the factors that influence these policies, admissions staff most frequently mentioned admissions standards from similar universities, followed by consulting with language teaching and testing experts at their institutions, and then information from test publishers. This finding is consistent with previous research, which has acknowledged the role of market factors and other externalities in setting admissions standards (Deygers & Vanbuel, 2022; Ginther & Elder, 2014; O'Loughlin, 2011). However, IEP faculty were somewhat more likely to attribute policies to factors not related to language proficiency than were admissions officers, suggesting some skepticism of the process.

Graduate directors were, as expected, less knowledgeable than admissions officers about language policies. They were frequently unable to state why their program had a certain preferred TOEFL iBT score. "The score was in place when I began" was a common theme. Several graduate directors simply stated that the graduate school (or equivalent) set the standards and they went along with whatever was recommended.

Our second research question addressed the knowledge and beliefs that participants held about the TOEFL iBT in particular. Again, we found that admissions staff were the most familiar and graduate directors the least. However, our results suggest that graduate directors feel that they would like to know more about the TOEFL iBT and testing in general to help them make informed admissions decisions. In fact, we heard from quite a few participants who expressed an interest to know more after taking the survey. For example, one participant wrote: "I just took the survey – and it made me feel like I really need to learn more about this topic, including because we are reviewing our use of standardized tests for admissions in general. This is also my first year in this position. I am going to share the survey with our graduate committee, including to get them to begin thinking about this topic more deeply, too."

Every study has limitations, and ours is no exception. We had originally hoped to do more in-depth focus groups of stakeholders at a small number of institutions to be able to compare the perspectives of IEP faculty, graduate directors, and admissions professionals working in the same context. However, we were unable to travel and thus this plan was not feasible. We thus had to rely on surveys, which were necessarily limited by the need to keep them short enough to be feasible. Some of our original research questions were not amenable to the kinds of questions that can be asked on a survey, and we were unable to explore these connections within institutions. Even so, it may be that our study can provide an impetus for future research along these lines as a way to go deeper into some of the issues raised in the study.

We also did not reach our recruiting target of 300 survey respondents, despite expending a great deal of effort attempting to recruit participants. Our sample of IEP faculty was rather smaller than we had hoped, but we were collecting data at a time when IEPs across the country were having difficulties and many were drastically reducing their size. Furthermore, we restricted our sample to faculty from IEPs where completion of the IEP suffices as evidence of English proficiency, which reduced the population of potential respondents even further. However, we are not convinced that a larger sample would have led to vastly different results, except that we would have been able to make more confident statements about differences across disciplines in graduate programs or across different types of institutions. Future research might be useful to be able to make this kind of comparison.

Conclusion

Taken as a whole, the study adds to the growing literature on language assessment literacy among different groups of stakeholders (Taylor, 2013) in terms of helping to identify which aspects of assessment literacy are important to different test taker groups. In particular, the study complements recent work by Malone and Montee (2014) in shedding light on what decision makers know/believe about the TOEFL iBT test and what they feel they need to know about language proficiency tests to do their job well. By including the perspectives of English language professionals in IEPs, our study also addresses in part Deygers and Malone's (2019) call to understand the perspectives of those who influence policy makers, in that admissions officers frequently consulted with the IEP faculty when contemplating policy changes.

Undergraduate admissions staff may not always be aware that achieving a minimum score on a language proficiency test like the TOEFL iBT test does not necessarily mean that students cannot benefit from additional ESL support (see ETS, 2020b, for a discussion of this issue). Policies that direct resources toward continuing to support students with minimally acceptable test scores may lead to higher retention and graduation rates for these students. Our study found that admissions staff often relied on experts at their own institutions for advice on accepting tests and setting cut scores. Such collaboration can benefit both the admissions staff and the ESL professionals as they seek to best serve students preparing to matriculate into a degree program.

In the absence of collaboration with language experts, or under enrollment management pressure in a competitive environment, it is not unrealistic for admissions professionals to base their decisions on what other institutions are doing. It might be beneficial to collect and disseminate data on the kinds of problems that students with adequate but low test scores may have in academic listening, speaking, reading, and writing, so that resources can be made available to such students, or in the absence of such resources, standards can be set at a higher level. As Deygers and Vanbuel (2022) noted, policy changes such as adding support programs for international students can be accomplished through more empirical studies of the level of proficiency needed for students to participate fully in their academic programs combined with an understanding on the part of language researchers of how policies are made and which key stakeholders need to be involved in communicating research results.

Many IEP faculty and graduate directors, particularly those in leadership positions, began their academic careers before the advent of the TOEFL iBT test, and thus their knowledge about the test may be out of date. We found that a number of IEP faculty still default to the old TOEFL paper-based test score scale when discussing admissions and exit criteria, and graduate directors who took the TOEFL paper-based test more than two decades ago would have little reason to know about how the test has changed. Given that the emphasis in the test has shifted from language knowledge to language use, it may be beneficial for test providers to provide some professional development around this topic to these stakeholder groups. The TOEFL Research Insights series is a good example of the kind of information that could benefit stakeholders. Volume 6, for example, provides useful information about the history of the TOEFL program (ETS, 2020a).

Of all the groups we surveyed, graduate directors were the ones who expressed the least satisfaction with their level of understanding of the TOEFL iBT test. This is a stakeholder group where test developers might focus their outreach efforts. However, given the great variation in disciplines, and the nature of the graduate director job, which tends to rotate frequently, it might be more productive to direct communications to deans of graduate colleges within universities, perhaps with materials that can easily be shared across departments, such as the TOEFL Research Insights series noted above.

Finally, as Deygers and Malone (2019) argued, it is important for language assessment professionals to understand the perspectives of policy makers so that their work can have “real-world impact” (p. 349). By highlighting the commonalities and differences across our three stakeholder groups, we see our study as contributing to what Taylor (2013, p. 411) characterizes as “collaborative ongoing dialogue taking place at the interface between language testing experts and non-specialist test stakeholders.”

Notes

- 1 Another stakeholder group not included in this study was recruiters for graduate programs who focus on recruitment of international students for specific disciplines, such as computer science or business.
- 2 Three additional research questions reported elsewhere include: (a) For graduate programs, what is the role of language proficiency vis-à-vis other considerations in making admissions decisions? (b) 2. For IEP faculty only, how do university language proficiency requirements influence curricular and assessment decisions? (c) Do participants from the three stakeholder

groups evaluate TOEFL iBT essays similarly, in terms of their perceptions of the writer's readiness for academic study and range of probable total TOEFL iBT scores.

References

- Baker, B. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *Canadian Journal of Applied Linguistics*, 19(1), 63–83. <https://journals.lib.unb.ca/index.php/CJAL/article/view/23033>
- Baker, B. A., Tsushima, R., & Wang, S. (2014). Investigating language assessment literacy: Collaboration between assessment specialists and Canadian university admissions officers. *Language Learning in Higher Education*, 4(1), 137–157. <https://doi.org/10.1515/cercles-2014-0009>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347–368. <https://doi.org/10.1177/0265532219826390>
- Deygers, B., & Vanbuel, M. (2022). Advocating an empirically-founded university admission policy. *Language Policy*, 21(4), 575–596. <https://doi.org/10.1007/s10993-022-09615-6>
- ETS. (2020a). TOEFL® research insight series: Vol. 6. TOEFL® program history. <https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v6.pdf>
- ETS. (2020b). TOEFL® research insight series: Vol. 8. Guidelines for setting useful score requirements for the TOEFL iBT® test. <https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v9.pdf>
- 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. TOEFL iBT-24). ETS. <https://doi.org/10.1002/ets2.12037>
- Green, A. (2018, October 12–13). *Assessment practices in US intensive English programs* [Paper presentation]. East Coast Organization of Language Testers, Princeton, New Jersey, United States.
- Lo Bianco, J. (2001). Policy literacy. *Language and Education*, 15(2–3), 212–227. <https://doi.org/10.1080/09500780108666811>
- Malone, M. E., & Montee, M. (2014). *Stakeholders' beliefs about the TOEFL iBT® test as a measure of academic language ability* (Research Report No. TOEFL iBT-22). ETS. <https://doi.org/10.1002/ets2.12039>
- Newcomb, J. T. (2017). *Understanding the lived experiences and training of undergraduate admission officers in detecting fraudulent application materials of international applicants at highly selective institutions* [Doctoral dissertation]. Northeastern University.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8(2), 146–160. <https://doi.org/10.1080/15434303.2011.564698>
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. <https://doi.org/10.1177/0265532213480336>
- Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7), 534–539. <https://www.jstor.org/stable/20404455>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Wilson, M. E. (2021). *When professional roles take center stage: How international enrollment managers make sense of their decision to accept Duolingo English test* [Unpublished doctoral dissertation]. Northeastern University.

Suggested citation:

Cushing, S. T., Ren, H., & Tan, Y. (2024). *The use of TOEFL iBT in admissions decisions: Stakeholder perceptions of policies and practices* (TOEFL Research Report No. RR-101). ETS. <https://doi.org/10.1002/ets2.12375>

Action Editor: Spiros Papageorgiou

Reviewers: This report was reviewed by the Research Subcommittee of the TOEFL Committee of Examiners.

ETS, the ETS logo, TOEFL, TOEFL iBT, and the TOEFL logo are registered trademarks of Educational Testing Service (ETS). TOEFL ESSENTIALS is a trademark of ETS. SAT is a trademark of the College Board. All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the [ETS ReSEARCHER](#) database.