

## Using TOEFL iBT® Scores for Determining English Proficiency of International Teaching Assistants (ITAs)

English-language proficiency is one among several important factors to consider when determining whether or not international graduate students are prepared to accomplish the various tasks associated with teaching at U.S. universities. The TOEFL iBT® test provides a trustworthy indication of a test taker's English-language proficiency in each of the four skills of listening, speaking, reading and writing. It was developed and validated specifically in relation to language use that is typical of university academic settings. As such, the TOEFL iBT test may serve as an appropriate tool for screening a teaching candidate's English-language abilities, and many U.S. universities use the test in this way.

This document provides suggestions for using TOEFL iBT scores in the ITA screening process, including:

- A. A sample of U.S. universities and their practices for using TOEFL iBT scores for ITA screening
- B. A set of research studies that have investigated TOEFL iBT cut scores for ITA screening
- C. An initial recommendation for using TOEFL iBT scores for ITA screening
- D. Approaches to monitoring and evaluating effectiveness and accuracy of the use of TOEFL iBT scores for ITA screening

### A. A sample of U.S. universities and their practices for using TOEFL iBT scores for ITA screening

Listed below are examples of six universities that refer to TOEFL iBT test scores during the process of determining ITA English proficiency in relation to readiness for teaching. Note that each university adopts distinct practices, in keeping with their own unique local situations and international student body. Also note that, in each of these examples, only TOEFL iBT Speaking section scores are utilized. For further details, see the websites provided for each.

1. University of Nebraska-Lincoln, <http://www.unl.edu/gradstudies/current/ita>  
Waiver (straight to teaching): iBT™ Speaking = 28  
Required to complete two-week course and conduct a teaching demo (judged by panel): iBT Speaking = 22-27  
Possible consider with caution: iBT Speaking = 18-21
2. University of Connecticut, <http://ita.uconn.edu/testing-english-proficiency-certification-effective-for-teaching-assistants-starting-in-fall-2014-or-after/>  
No additional testing (straight to teaching): iBT Speaking = 27  
Additional testing/micro-teaching: iBT speaking = 22-26
3. University of Illinois, <http://www.grad.illinois.edu/admissions/taengprof.htm>  
Minimum score to be allowed to teach: iBT Speaking = 24
4. University of San Francisco, [http://www.grad.usf.edu/inc/linked\\_files/ga\\_policies\\_spoken\\_english.pdf](http://www.grad.usf.edu/inc/linked_files/ga_policies_spoken_english.pdf)  
Unconditional appointment: iBT Speaking = 26  
Conditional appointment: iBT Speaking = 23-26

5. Bowling Green University, <https://www.bgsu.edu/arts-and-sciences/english/english-for-speakers-of-other-languages/international-teaching-assistant--ita--information.html>  
Waiver (straight to teaching): iBT Speaking = 24  
Probation (teach/tutor with monitoring): iBT Speaking = 21–23
6. University of Maryland, <http://marylandenglishinstitute.com/wkdir/english-programs/international-teaching-assistants/>  
Local assessment required, but it is recommended that students who have iBT Speaking = 25 or higher have a good chance of passing the local assessment

#### **B. A set of research studies that have investigated TOEFL iBT cut scores for ITA screening**

A number of studies have examined the use of Speaking scores on the TOEFL iBT test as a measure of speaking ability in an academic target language use domain for ITAs (Farnsworth, 2013; Wagner, 2016; Wylie & Tannenbaum, 2006; Xi, 2007, 2008). In Xi's (2007, 2008)<sup>1</sup> study, the relationship between TOEFL iBT speaking scores and local ITA screening test scores in three U.S. universities was moderately strong. The study also found that TOEFL iBT Speaking scores correlated more strongly with local assessment scores when these assessments had linguistically-driven scoring rubrics, as opposed to rubrics that emphasized nonlinguistic aspects of teaching. This was a desired result, given that the TOEFL iBT Speaking score is intended as an indicator of oral proficiency in academic context, not as an indicator of teaching ability per se. Wagner (2016)<sup>2</sup> also found a relationship between TOEFL iBT Listening scores and students' perceptions of their ITAs' oral proficiency; this was again a desired result, given that ITAs need to use their listening skills when they communicate with undergraduate students. The literature, therefore, suggests that the TOEFL iBT Speaking scores and, to some extent, Listening scores, are useful indicators of the language proficiency of prospective ITAs.

The most directly relevant study for establishing appropriate standards for ITA screening is Wylie and Tannenbaum (2006)<sup>3</sup>. The purpose of this study was to establish cut scores for screening ITAs using TOEFL iBT Speaking section scores. A panel of 18 experts who work with ITAs was convened to judge the meaning and relevance of two different cut scores. First, they determined a cut score for minimally acceptable speaking skills for ITAs to be able to work with undergraduate students, which they set at a score of 23 out of 30 points on TOEFL iBT Speaking. Second, they determined a higher cut score based on relationship with a score of 50 on the *Test of Spoken English™ (TSE®)*<sup>4</sup>, and intended to indicate speaking proficiency fully adequate for teaching purposes. They set this higher cut score at 26 out of 30 points on iBT Speaking. The findings of this study corroborate and align closely with the findings of other studies, and they provide a useful basis for initial recommendations regarding appropriate cut scores for screening ITAs (see next section).

---

<sup>1</sup> <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2008.tb02088.x/abstract;jsessionid=B97948CCFC57086D2753990CA219B1D4.f01t04>

<sup>2</sup> <http://onlinelibrary.wiley.com/doi/10.1002/ets2.12104/full>

<sup>3</sup> [https://www.ets.org/research/policy\\_research\\_reports/publications/report/2006/icja](https://www.ets.org/research/policy_research_reports/publications/report/2006/icja)

<sup>4</sup> The TSE® is a discontinued test that was designed and used primarily as a measure of English speaking for employment purposes, including graduate teaching assistantships at U.S. universities.

The literature also points out that the same cut score might not be equally relevant across different ITA contexts. When making a decision about allowing an international student to teach, institutions will need to take into account a variety of factors such as:

- The type of classification error that is deemed more important to avoid, that is, false positive (allowing a student to teach when the student does not have sufficient language proficiency) or false negative (not allowing students with sufficient language proficiency to teach). Both classification errors have important consequences. False positive classifications are likely to result in frustration among undergraduate students taught by the prospective ITA, whereas false negative errors are likely to have consequences at the institutional level, as departments might not be allowed to employ qualified ITAs.
- The remedial measures that an institution has in place. For example, students who receive a score below the cut score might be asked to take an English-language course and teach upon successful completion of that course; in this case some false positive classifications can be rectified. If, however, the institution does not have enough resources to support such a language course, then a higher cut score might be required to minimize false positive classifications.
- The importance of distinguishing between sufficient language proficiency and adequate teaching skills. A language test such as the TOEFL iBT test should only be used as an indicator of the former, not the latter. In cases where undergraduate students complain about an ITA, for example, institutions need to examine whether such complaints are due to a lack of sufficient English-language skills versus teaching skills, intercultural competence, or related.

### C. An initial recommendation for using TOEFL iBT scores for ITA screening

Based on existing research, and reflecting commonalities in practices across U.S. institutions, an initial recommendation of English-language proficiency cut scores for ITA screening can be made in reference to TOEFL iBT Speaking section scores. While TOEFL iBT section scores for listening, reading, and writing are not typically used for ITA language proficiency screening purposes, a likely threshold of proficiency in each of these skills may be identified in reference to typical performance expectations for graduate students. The two approaches and recommended initial cut scores are explained below.

#### Listening, Reading and Writing

While all four English-language skills are important for teaching and teaching-related work typically done by ITAs, for the three skills of listening, reading, and writing it is likely that a single threshold proficiency level is sufficient for determining the extent to which ITAs are able to function successfully in English in the teaching environment. Two sources of information are useful in setting an initial cut score for these three skills:

- a. *A guide to understanding TOEFL iBT® scores*<sup>5</sup>, and
- b. *the Test and score data summary for TOEFL iBT® tests*<sup>6</sup>

---

<sup>5</sup> [https://www.ets.org/s/toefl/pdf/performance\\_feedback\\_brochure.pdf](https://www.ets.org/s/toefl/pdf/performance_feedback_brochure.pdf) and [https://www.ets.org/toefl/teachers\\_advisors/scores/interpret/](https://www.ets.org/toefl/teachers_advisors/scores/interpret/)

<sup>6</sup> [https://www.ets.org/s/toefl/pdf/94227\\_unlweb.pdf](https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf) and <https://www.ets.org/toefl/research>

According to (a), a “high performance” on the Listening and Reading subsections of the TOEFL iBT test is registered at a section score ranging from 22–30. Similarly, in (b), a score of 23 on the Listening and Reading subsections indicates that the examinee has scored at around the 50th percentile of all graduate students taking the TOEFL iBT test in the year 2015. For Writing, a score of 22 indicates that the examinee has scored at around the 50th percentile of all graduate students, while a score of 23 suggests an incremental improvement to around the 60th percentile. Based on these sources, an initial cut score representing a threshold level of proficiency for these three skills might be set as:

**iBT Listening = 23 or higher**

**iBT Reading = 23 or higher**

**iBT Writing = 23 or higher**

### Speaking

Given a likely emphasis on English speaking as the primary medium of teacher activity, and based on research findings and current practices of other institutions, several possible cut scores might be set as initial decision points in the ITA screening process. Assuming that there is the possibility for institutional intervention and support to be provided, the following cut scores and decisions are recommended for initial use:

#### **TOEFL iBT Speaking = 27–30**

*Unconditional appointment:* No additional language testing or training required before teaching.

#### **TOEFL iBT Speaking = 23–26**

*Conditional appointment:* Additional requirements to be determined by institution, for example: (a) student delivers a teaching demonstration to a panel for arbitration regarding readiness to use English in the classroom; (b) student is placed into teaching but monitored closely during the initial X weeks of teaching; (c) student participates in additional language instruction such as a course on English for teaching purposes.

#### **TOEFL iBT Speaking = 22 or below**

*Non-appointment:* Teaching in English not permitted until speaking proficiency is improved (criterion for improvement set by institution).

**Note:** Given that all tests provide at best an estimate of ability, a procedure should be established for adjudicating and resolving borderline cases.

### **D. Approaches to monitoring and evaluating cut scores and decisions**

Any use of test scores for making important decisions, including screening of ITA candidates, should be monitored and evaluated for accuracy, effectiveness and consequences. A variety of possibilities exist for local validation of test use along these lines, several of which are presented here as examples of suggested practice.

- a. Analyze the proportion of ITAs screened into the different categories proposed in Section C, in comparison with TA placement needs of departments and existing practices. Is the use of the initial iBT cut scores approach similar, more severe, more lenient? What are the implications for departments, students and ESL support, based on initial screening decisions?

- b. Survey and/or interview ITAs placed into teaching positions at midpoint and end of semester in order to determine their perceptions regarding the English-language demands of teaching and their own readiness in terms of each of the four skills.
- c. Survey students in ITA-taught classes at end of semester in order to determine their perceptions regarding the English-language demands of teaching and their teachers' readiness in terms of general-language proficiency and/or each of the four skills (possibly in combination with regular end of semester course evaluation).
- d. Survey departmental decision makers regarding graduate students placed into teaching positions as well as non-appointed graduate students in order to determine their perceptions of:
  - Number of accurate screening decisions (how many candidates were accurately placed into teaching or non-appointed)
  - Number of false positives (how many candidates inaccurately placed into teaching)
  - Number of false negatives (how many candidates inaccurately excluded from teaching)
- e. Conduct regular classroom (and non-classroom instructional work, as applicable) observations of newly placed ITAs to assess 'on-the-job' use of English skills.
- f. Compile evidence and reflect on initial cut scores after one or several iterations of use; consider possible adjustments in cut scores and likely consequences.

## References

Farnsworth, T. (2013). An investigation into the validity of the TOEFL iBT Speaking test for International Teaching Assistant Certification. *Language Assessment Quarterly*, 10, 274–291.

Wagner, E. (2016). *A study of the use of TOEFL iBT Speaking and Listening scores for international teaching assistant screening (TOEFL iBT® Research Report No. 27)*. Princeton, NJ: Educational Testing Service.

Wylie, E., & Tannenbaum, R. (2006). *TOEFL academic speaking test: Setting a cut score for international teaching assistants (Research Memorandum No. RM-06-01)*. Princeton, NJ: Educational Testing Service.

Xi, X. (2007). Validating TOEFL® iBT Speaking and setting score requirements for ITA screening. *Language Assessment Quarterly*, 4(4), 318-351.

Xi, X. (2008). *Investigating the criterion-related validity of the TOEFL speaking scores for ITA screening and setting standards for ITAs (TOEFL iBT® Research Report No. 3)*. Princeton, NJ: Educational Testing Service.