

The New Test of English as a Foreign Language (TOEFL)

Eileen Tyson

A
B C D
E F G H I
J K L M N O P
Q R S T U V W X Y Z


Forty-one years ago, 765 students took the TOEFL (Test of English as a Foreign Language) test. Over the years, that number has increased dramatically, to 720,000 in 2004. After ten years of development, TOEFL's new Internet-based test became available in late 2005. Until then, there had been no speaking component of TOEFL. Instead, students were required to pay for and take a separate speaking test. The Educational Testing Service (ETS) knew it needed to develop a method that met its very high standards for objectivity, reliability and validity—and the face to face interview was not an option because there's simply too much opportunity for bias between the interviewer and the test taker.

ETS measures not just how much a test taker knows about English, but more importantly, how well he or she can use English for communication. With something brand new called “integrated tasks,” ETS uses questions that ask test takers to combine or integrate abilities. For example, ETS asks test takers to read, listen, and then speak and write about what they heard. This is a huge change for the language-testing community. Test takers are not accustomed to being asked to speak in terms of integrating abilities. This is a much more complex assessment and couldn't occur without Internet-based delivery.

ETS relies on the concept of minimally acceptable language skills. It is searching for the minimal English competency a student needs to understand his or her courses and the minimal English competency that the physical therapist needs to do his or her job. Think about English proficiency as a continuum. Some non native speakers have no trouble reading or communicating. Other can't communicate. Standard-setting looks for folks in the middle, who have good enough English to manage. They may not be eloquent, but they are understandable.

Test takers must go to a secure testing center where identification is required. The center maintains the same standard of security as the computer based and paper based TOEFL. ETS makes test forms available on Greenwich Mean Time. This ensures that everyone in the world for whom it is not nighttime or middle of the night will be taking the test at literally the same time. This avoids time-zone cheating. Each administration uses a different test form to maintain security. Test takers have up to four hours to complete the test.

After operational trials globally during the summer of 2005, testing began in the United States in September and in Canada, Italy, France and Germany in October. The roll-out for the rest of the world began a month later. Anywhere TOEFL IBT could not be introduced, ETS kept the old tests temporarily.



[English language proficiency] standard-setting looks for folks in the middle, who have good enough English to manage. They may not be eloquent, but they are understandable.

A new score scale was also required. There are four scores, one for each test section, on a 0 to 30 scale. Total score is on a 0 to 120 scale. To determine what new TOEFL score requirements should be, check the score comparison chart on the Educational Testing Service website. Current TOEFL score requirements can be translated into the new IBT score. The new test measures four skills, whereas the old one measured three.

Organizations may also conduct a standard-setting process, or adopt standards already suggested. Last May, 22 participants from six healthcare organizations – including physical therapy – reviewed the entire test and set scores for each section. Two different processes were used in standard-setting. Participants reviewed the sections where the test taker had to construct and produce something written or spoken. Judgment was based upon actual responses from field studies scoring at the different levels. For example, in the speaking section, people would listen to a level four response as well as a three, a two, and a one. Then participants reviewed selected-response sections where test takers selected the right answer. Judging was based on whether someone with minimally acceptable English would be able to answer this question correctly.

ETS gave the panel practice in making judgments. Individuals would listen, for example, to sample responses and make judgments. Then there would be a discussion: Why did you rate this response a level three and so on? In a second round of voting, the panel made the recommendation for each spoken or written response.

The group identified ahead of time the characteristics of a strong and weak writer and described the characteristics of a minimally acceptable physical therapist writer by answering the following questions.

- Is the writing understandable?
- Do errors in syntax, grammar and vocabulary obscure meaning?
- Are ideas and issues understandable?
- Do they write quickly enough to complete tasks in a reasonable length of time?

Then the group reviewed two writing tasks. There was an integrated test question (20 minutes), where test takers

listened, read, and then wrote in response. The second task was an essay (30 minutes) that asked test takers about their experience or opinion. For example, in one integrated question, test takers read a passage about the benefits of working in groups, and then heard a short lecture describing problems with working in groups. They were given 20 minutes to respond in writing to what they read and what they heard. They were asked to summarize how the lecture casts doubt on points made in the reading.

The other form of standard-setting was for listening. For students, we added lectures where the professor spoke and asked questions. We tried to simulate what takes place in a classroom or in communications situations. How much listening skill did a student need? ETS considered each separate question. Would a student who is just good enough get this answer correct or wrong?

ETS also identified the listening tasks that a physical therapist or a medical health professional faces, including understanding directions, questions, verbal reports from support staff, patients, families, and other medical professionals. The group described a minimally acceptable health professional listener who could receive and respond appropriately to information based on vocabulary and intonation; they understand subtleties and have a fairly broad vocabulary. ETS asked whether a minimally acceptable listener would get a particular question correct. It made first-round judgments, and then returned to discuss the questions where there was the most disagreement. Then the panel made a second recommendation and raw scores were converted to scaled scores.

While the panel made its recommendations, each university, professional association or state board must decide whether to take those recommendations. ETS encourages organizations to use standardized test scores responsibly. Conduct validation studies and review score standards periodically. After an interval of two years, for example, ask, are those scores measuring what needs to be measured? Are they giving the desired results? Do scores need to be higher or lower? As universities set and report their score requirements, ETS is posting them on the TOEFL website, www.toefl.org. The Educational Testing Service website is www.ets.org. ■