



February 24, 2004

TOEFL – the next generation

The venerable TOEFL exam has thousands of enthusiastic proponents, but it also has its share of detractors. Test developers hope to win them over with a state-of-the-art revision

Story by **TERRY FREDRICKSON**

One of the most familiar acronyms for students seeking admission to colleges or universities abroad has to be TOEFL[®], short for Test of English as a Foreign Language. Since its inception in 1964 nearly 20 million students have taken the test, many of them more than once.

Wildly popular among admissions directors almost from the beginning, it didn't take long for many of them to latch on to a cut-off score of 550 on the 677-point scale. That is still the requirement for most academic institutions accepting the TOEFL, although you will find some that are more (or less) demanding.



Susan Chyn

Interestingly, says Susan Chyn, a director at Educational Testing Service (ETS), the non-profit organisation which develops and administers the test, “we never told people how high the score needed to be for acceptance. When I started working with ETS in 1981 that 550 was already engrained.”

While this standard is still generally credible, there can be a problem in how a student attains it, a fact that Chyn readily admits. “The test-takers see the test-givers as somebody they've got to beat and so they'll find ways to outsmart them. It happens with every test I need to say – every test in the world.”

This is particularly true in the case of TOEFL where a huge industry has grown up to help students outsmart the test. Here, the emphasis is not so much on improving overall English, but mastering the specific structures and vocabulary most commonly tested. As a result, some test-wise students clearly do gain acceptance to foreign universities without the English proficiency necessary to succeed.

But TOEFL is a moving target. The test developers are acutely aware of the need to produce a valid test with reliable results, so test revision is a constant activity. There is a great deal of excitement about the next revision which is due to be implemented

worldwide in September of next year. If the test developers are right, the only way you can truly beat this “next-generation TOEFL” is to learn English to a very high standard.

The TOEFL evolves

As a former director of test development for both TOEFL and TOEIC, another popular ETS exam, Chyn has an insider’s perspective on the test and its evolution over the years.

“The original purpose of the test was to assess the ability of a non-native speaker of English to come to North America to study,” she explains. “It was designed to assess English language proficiency so that you could tell administrators, admissions officers, faculty deans in graduate schools, people giving scholarships or even visa officers whether or not that individual had sufficient English to learn and function in an academic setting. The initial purpose was quite specific, but over the years, the use of the test has broadened,” Chyn observes.

Although the biggest changes in the TOEFL to date are found in its computerised version, the test has never been static. “One of the first changes we made was to eliminate the section where vocabulary was tested in individual sentences,” Chyn recalls. “We felt that we could better measure vocabulary as part of contextualised language so vocabulary was incorporated into reading comprehension.”

The change to a computer-based test (CBT) was an eye-opener for the developers, Chyn says. “It was the first effort of its kind. No one had ever done it in large-scale testing. They were able to use graphics with reading comprehension. With the listening comprehension, they were able to use pictures for the first time.

“They kept the structure section, which also tested written expression, but they also incorporated an essay as a direct measure of writing. They distributed the essays to raters via the web, so scoring time was cut significantly.”

One of the most noticeable differences between the CBT and the traditional paper-based test – which, incidentally, is still widely used – lies in the scoring scale. The 550 paper-based standard has become 213 for the computerised version.

The discrepancy, says Chyn, stems, in part, from a fundamentally new method of testing which was applied to structure and listening comprehension. “Instead of the linear method of setting fixed questions and basing a score on those questions, the adaptive approach takes advantage of a huge question bank to provide questions of differing difficulties, according to the responses given by each individual,” Chyn explains.

“The theory is, that if a test-taker takes a hard item and gets it right, the computer gives you an even harder question. If you get that wrong, you will get an easier one. With each question, ETS gets increasingly precise measurement, as we are honing in on the true proficiency level of each individual.”

The good news, says Chyn, is this technique works very well for certain subjects, especially with discrete test questions. But in practice, especially where extended chunks of language are being measured, there have been mixed benefits.

“In early field tests, we found that we could measure X and Y really well with very few questions, but the students complained they didn’t feel they were getting a fair chance to show their ability,” Chyn relates. “So when we rolled out the actual adaptive test, we made the test longer. It was still shorter than the original, however.”

Even the modified adaptive design will soon become a thing of the past, Chyn says. “In the new test, there will be no adaptive modules whatsoever. When we looked at the various benefits and weaknesses, we realised we could do better. Some scientists had a great idea. We tried it and learned a lot. We are going back to linear model.”

Creating positive washback

One of the main raps against both versions of the current TOEFL is that they may fail to adequately capture a test-takers proficiency in the so-called productive skills of speaking and writing. Speaking is not tested at all and writing is only directly tested in the CBT through a 30-minute essay. Writing expression in the paper-based test is essentially error recognition.

A second complaint is that the exam tests skills separately with little attempt at integration. This, critics say, does not mirror the demands of college and university life where language skills are seldom used in isolation.

Except, that is, in many English language preparatory programmes where the overriding goal is a high TOEFL score. This, critics say, distorts the teaching process and students are ill served even if they do achieve a high enough score for university entry.

This negative “washback effect” has been a significant concern for Chyn and it is a big reason why she is so enthusiastic about the new-generation TOEFL.

“Not only do we have an increased focus on production, we are doing it in an integrated fashion in that we capture all your mental processing and language productive skills. We believe this can only have positive washback on classroom practices – whether it is a test prep course or a general English class. As a result, you will get individuals who have to learn better English.”

To better appreciate what she means, consider the following example question from the new-generation TOEFL website. Notice that the task requires reading, listening and speaking together with a good measure of mental agility.

Research plus technology

How would you prepare for such a test? Quick fixes and highly targeted study are dubious strategies at best. Indeed, it appears you will have little choice but to improve your overall English.

According to Chyn, the new-generation TOEFL is the result of ten years of research. Much of this research centred on the communication tasks common to an academic environment.

“They really researched this a lot,” Chyn says. “Some of us wanted the test a lot sooner. They did task analysis in subjects like biology and anthropology. They looked at language functions in social communication, in professional interactions, and in lectures. They did analyses of the type of words that were being used in lectures. They did corpora studies where they’re looking at frequency of lexicon.”

As a result, the language in the test is as realistic as possible, as are the tasks. And, for the first time in a TOEFL, test-takers will be able to take notes. This, says Chyn, should be a psychological boost if nothing else.

It is no accident that the advances in the TOEFL closely parallel advances in technology. Indeed, the Internet now plays a vital role in its delivery. “We called the current one the CBT. The new one, we’re calling iBT because it’s Internet based,” Chyn says.

Test-takers will be sitting in front of a computer with headsets containing earphones and a microphone. The full 3-½-hour test will be delivered over the Internet. Student spoken responses will be digitised and transmitted to the ETS centre for grading by human raters. Writing responses, including the essay, will be graded electronically.

There will be no excuses for prospective test-takers who say they were taken by surprise. Within days, in fact, anyone can go on line and see for themselves what the test will be like.

“In March, they will be able to go to <http://www.toefl.org/> and take part in what we’re calling the virtual tour,” Chyn relates. “It is really a prototype with an explanation, so you can just log on and there it is. That gives people about a year and a half to start getting ready.

“In July, one of the programme’s new babies, the English language on-line learning portal, will come on line. It’s an evolving product line, but in the beginning I think it will be a full-blown TOEFL test. You’ll actually be able to go on and take the test and for listening, reading and writing, you’ll get an immediate score. For speaking, you’ll have to take the TOEFL Academic Speaking Test (TAST) component. That will take a little longer,” Chyn explains.