

AS A FOREIGN LANGUAGE (TOEFL)

Jiddou Soukalo, Ph.D.
English Department
University of Nouakchott

The choice of The Test of English as a Foreign Language (TOEFL) for review is dictated by several reasons. First, I can relate to this test for having taken it twice (1975,1984) prior to undertaking studies in the USA. Subsequently, I feel it justified to know about this test based on the fact that many of my students and people from different academic levels and backgrounds wishing to further their studies will very probably have to take it. To talk knowledgeably about the test is very important especially if one has to prepare these people to take it. Additionally, because those students are English majors and prospective teachers, the TOEFL test is the kind of language proficiency test they need to be familiar with in constructing their own tests. Finally, my purpose in reviewing the TOEFL test is to verify the appropriateness/relevance or lack of it of some aspects of this test.

General description

Designed to evaluate the English proficiency of people for whom English is not a native language, the TOEFL test is group administered. It uses a multiple choice format and assesses the ability to understand North American English. The TOEFL test was first administered by the Modern Language Association. In 1965 The College Entrance Examination Board and Educational Testing Service became sponsors and were joined in 1973 by The Graduate Record Examination Board. Today the Policy Council that is responsible for the formulation of the TOEFL program policy has expanded to 15 members.

In its original form the TOEFL test consisted of 5 sections (1. listening comprehension, 2. English structure, 3. vocabulary, 4. reading comprehension, and 5. writing ability) and was made of 200 questions. Since the 1976 revised edition the questions have been reduced to 150, some of the sections have been combined and the test now is composed of three sections only: 1. listening comprehension, 2. structure and written composition, 3. vocabulary and reading comprehension. The test is administered in a single session of about three hours that include the time required for admitting examinees.

The TOEFL test is published by Educational Testing Service (ETS) which is based in Princeton, New Jersey. According to the 1988-89 TOEFL Bulletin, the test was administered to 452,000 people at more than 1,100 test centers in 170 countries and areas. More than 2,300 US and Canadian colleges and universities require non-native speakers of English to take the TOEFL. The TOEFL is administered six times a year under two types of programs which differ only in their testing days: The International Testing Program (Saturdays) and The Special Center Testing Program (Fridays). The test is administered in the USA and overseas. The cost of the test varies according to whether it is taken under The International Program (\$38) or The Special Program (\$45) (TOEFL Bulletin, 1995-96). The fee paid by the examinees entitles them to receiving their score

The first section of the TOEFL test (listening comprehension) which lasts 30 minutes assesses the examinee's ability to match 20 oral statements with the correct written options, to answer correctly 15 short conversation-based questions, and to answer correctly 15 questions based on a lecture. The second section (structure and written composition) which lasts 25 minutes, includes a completion part, and a part requiring judgment of grammaticality and appropriateness of words and expressions. Section three (vocabulary and reading comprehension) lasts 45 minutes. In the first part of this section the task is based on the context of sentences, and the examinee is asked to match 30 underlined words and phrases with the appropriate options. Part two includes 30 questions about comprehension of several reading passages.

Special procedures are used by the examiners to protect the security of the test before, during and after its administration. Only specially trained people are allowed to administer the TOEFL test. Training features include specially produced video taped and training workshops.

Practical evaluation

With respect to quality and appropriateness of materials, the literature seems to suggest that the TOEFL test is well constructed (Loyd, 1985) and that its accompanying manual for users is well written (Chase, 1972; Loyd, 1985). For each section of the test directions to the examinees are written in English and include sample exercises. The cassette that is provided for the administration of the listening comprehension section of the test is of high quality and presents the materials clearly (Loyd, 1985). Following the 1979 Federal legislation requiring 'institutions to judge U.S resident handicapped and nonhandicapped applicants by the same admissions criteria ... special administrations of the TOEFL occur on regular test dates' (Philips, 1985) and are adapted to the abilities of the examinees (eg. special testing facilities, longer and more frequent rest periods, omission of section 1 for hearing impaired).

The administration procedures of the TOEFL test put a lot of responsibility on the examiner. Such responsibility is usually based on good understanding of simple but crucial administrative rules. According to Philips (1985) these include strict procedures for admitting examinees in the room, seating of examinees in assigned seats, removal of extraneous materials from their desks, strict carrying out of instructions. Additionally, Philips points to the necessity of verifying the completeness of the materials, ensuring that the materials are safely stored, and preventing the breaking of the sealed test booklets.

Procedures during the actual administration of the test require the examiner to check the recording equipment, to see to it that all examinees are comfortable and ready, to monitor the allotted testing time for the written sections, and to check the names of the examinees' answer sheets against their identification documents as they leave the room. Irregularities during the test must also be reported and test materials sent to ETS TOEFL office.

Since the TOEFL test is based on multiple-choice items it is easily machine-scored. There are four scores reported, one on each section of the test and a total score. The examinees usually get their score reports within a month of the test administration.

conclude that the TOEFL test is a short, highly secure, and easily administered test. The administration procedures are essentially important to observe and the test scores are easily interpreted.

Content evaluation

Though the sample practice questions of the TOEFL bulletin of information give an idea of what the test items look like in every section, they are not really very informative as to the content of the TOEFL. In addition, sample tests available for \$ 3 are about half the length of a test that was actually administered. It is predictable that any evaluation of its content would not be valid because the test is not complete. While the present review cannot make complete and adequate evaluation of the content of the TOEFL test without a complete full length sample of the test, it is widely reported that the information provided about each section is usually general in nature and described as follows: 'common vocabulary, intonation, sounds, and other aspects of oral English that are difficult for nonnative speakers; basic English grammar; academic reading materials that ~~echnic~~ require no specialized subject matter knowledge.

Technical evaluation

With regards to the quality of norms, it is reported (Philips, 1985) that the norm group to which the examinees are compared is a 1964 group of foreign students (graduate and undergraduate) whose total standard score based on the five sections had a mean of 500.

For the 1976 revised edition of the TOEFL the sections raw scores are transformed to scaled scores ranging from 20 to 80 and the total score to standard scores ranging from 200 to 800. The following figures are figures relating to one million examinees who took the TOEFL between July 1980 and June 1982: the scaled scores ranged from 20 to 68, the total scores ranged from 217 to 677, and the means was 503 with a standard deviation of 66. Percentile ranks are also reported.

For the original form of the TOEFL test the total reliability based on KR.20 is .97 (Chase, 1972) whereas reliability coefficients for the test subparts range from .84 to .91 (Chase, 1972; Domino, 1972). For the present TOEFL test the reliabilities reported for the sections range from .87 to .89 and the reliabilities reported for the total score is .95 (Loyd, 1985). Therefore, the TOEFL test appears to have a very high reliability. However, according to Philips (1985) these reliability coefficients are limited to examinees tested in Canada and the USA and 'should have been calculated for other groups and for a representative sample of all nonnative English speakers who took the test during the period for which data are reported' (1985, p.664).

According to Chase (1972) and Domino (1972) pre 1976 validation studies on concurrent, predictive, and construct validity of the TOEFL scores can be summarized as follows: the TOEFL scores have substantial concurrent correlations (.70 to .80) with other proficiency tests (eg. Michigan Test of English Language Proficiency), and low correlations scores (.17 and .26) in predicting achievement scores (eg. grade point average). The studies suggest that construct validation of the TOEFL is scarcely permissible.

Though information on the validity of the present TOEFL test are supportive (Tatsuoka, 1985), there is reason to cast doubt on its content validity.

Comments from other reviewers

Several expert reviews of the TOEFL test point to strengths and weaknesses. In reporting evaluations, attempt will be made to select views shared by most reviewers. For example, the issue of selection of items to be included in the test and the issue of group selection for pretesting are recurrent ones. For Philips (1985) no rationale is given for inclusion of grammatical structures or vocabulary words. Concern for the harm that the selection of certain items may cause to the scores of examinees has also been expressed by Tatsuoka (1985). For Tatsuoka the use of proper nouns, for example, and especially proper nouns with sounds that are nonexistent in the language of the examinee, present enormous aural discrimination difficulty. Hence, the problem of selection of items is in order and requires some consideration. The fact that the TOEFL has a worldwide audience may hinder practical decisions on the part of the test publishers for matching item selection with characteristics of the examinee. It would appear that to solve this problem would require having to prepare several forms of the test based on several norm groups. Therefore, the pregnant question is how unbiased can a test such as the TOEFL be when it addresses so many different groups with varying experiences, language and cultural backgrounds?

With respect to selection of groups of students for pretesting, the representativeness of the norm group is questioned. Based on shortcomings such as item and group selection, some reviewers have denounced the misuse of TOEFL scores to predict academic achievement (Domino, 1972; Chase, 1972).

Generally speaking content validity, item selection, and information bias are closely related. They certainly constitute one of the unsolved issues of the TOEFL test. Test reviewers generally agree that apart from these selection issues and the bias they engender against some examinees, the TOEFL test is a 'useful measure in providing evidence about the ability of international applicants to handle instruction in English' (Philips, 1985; Loyd, 1985; Tatsuoka, 1985).

Overall evaluation

While the scores on a TOEFL test are indicative of an applicant's proficiency in English and his/her readiness or lack of it for engaging in academic studies, they can only be correlated with scores in similar proficiency tests and should not be used to predict an applicant's academic achievement in his/her field of study, for instance.

Academic institutions should always consider the effect of bias (cultural, racial, sex, content) of the test on the scores of the applicants. Many applicants have probably never heard spoken American English and may additionally not understand the concept of 'computer-dating', for instance.

TOEFL scores can be appropriately used in assigning new students the needed language instruction based on the scaled section scores. It should be borne in mind, however, that proficiency is to be viewed in terms of global or holistic performance rather than discrete criteria. Finally, it would be fair to describe the TOEFL test as a reliable instrument of English proficiency that need to improve on some aspects of its validity with respect to selection of items and pretesting groups.