

# TOEFL<sup>®</sup>

*1997 Edition*



# TEST & SCORE MANUAL



**T**he *TOEFL Test and Score Manual* has been prepared for deans, admissions officers and graduate department faculty, administrators of scholarship programs, ESL teachers, and foreign student advisers. The Manual provides information about the interpretation of TOEFL® scores, describes the test format, explains the operation of the testing program, and discusses program research activities and new testing developments.

This edition of the *Test and Score Manual* updates material in the 1995-96 edition, providing a description of revisions to the test introduced in July 1995, and other information of interest to score users. With the exception of “Program Developments” on page 10, information in this Manual refers specifically to the paper-and-pencil TOEFL test.

As this edition goes to press (summer 1997), a computer-based TOEFL test is under development and planned for introduction in the second half of 1998 (see page 11). More information about the computer-based test and score interpretation will appear on the TOEFL website at <http://www.toefl.org> and through new publications as it becomes available.

Add your name to the TOEFL web list (on “Educators” directory page) and receive e-mail announcements as they are released.

TOEFL Programs and Services  
International Language Programs  
Educational Testing Service



Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 1997 by Educational Testing Service. All rights reserved.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, GRADUATE RECORD EXAMINATIONS, GRE, SLEP, SPEAK, THE PRAXIS SERIES: PROFESSIONAL ASSESSMENTS FOR BEGINNING TEACHERS, TOEFL, the TOEFL logo, TSE, and TWE are registered trademarks of Educational Testing Service.

COLLEGE BOARD and SAT are registered trademarks of the College Entrance Examination Board.

GMAT and GRADUATE MANAGEMENT ADMISSION TEST are registered trademarks of the Graduate Management Admission Council.

SECONDARY SCHOOL ADMISSION TEST and SSAT are registered trademarks of the Secondary School Admission Test Board.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Violators will be prosecuted in accordance with both United States and international copyright and trademark laws.

Permissions requests may be made on-line at <http://www.toefl.org/copyrigh.html> or sent to:

Proprietary Rights Office  
Educational Testing Service  
Rosedale Road  
Princeton, NJ 08541-0001  
USA

Phone: 609-734-5032

# CONTENTS

<b>Overview</b> .....	7
TOEFL Policy Council .....	8
Committee of Examiners .....	8
Finance Committee .....	8
Research Committee .....	9
Outreach and Services Committee .....	9
TWE Test (Test of Written English) Committee .....	9
TSE Test (Test of Spoken English) Committee .....	9
<b>Program Developments</b> .....	10
TOEFL 2000 .....	10
The Computer-Based TOEFL Test .....	10
<b>Test of English as a Foreign Language</b> .....	11
Use of Scores .....	11
Description of the Paper-Based TOEFL Test .....	11
Development of TOEFL Test Questions .....	12
<b>TOEFL Testing Programs</b> .....	13
Friday and Saturday Testing Programs .....	13
TWE Test (Test of Written English) .....	13
TSE Test (Test of Spoken English) .....	13
Institutional Testing Program .....	13
<b>Procedures at Test Centers</b> .....	15
Measures to Protect Test Security .....	15
Identification Requirements .....	15
Photo File Records .....	16
Photo Score Reporting .....	16
Checking Names .....	16
Supervision of Examinees .....	17
Preventing Access to Test Materials .....	17
<b>TOEFL Test Results</b> .....	18
Release of Test Results .....	18
Test Score Data Retention .....	18
Image Score Reports .....	18
Official Score Reports from ETS .....	19
Information Printed on the Official Score Report .....	20
Examinee Score Records .....	21
Acceptance of Test Results Not Received from ETS .....	21
Additional Score Reports .....	22
Confidentiality of TOEFL Scores .....	22
Calculation of TOEFL Scores .....	22
Hand-Scoring Service .....	23
Scores of Questionable Validity .....	23
Examinees with Disabilities .....	24
Use of TOEFL Test Scores .....	25

(continued)

<b>Statistical Characteristics of the Test</b> .....	29
Level of Difficulty .....	29
Test Equating .....	29
Adequacy of Time Allowed .....	29
Reliabilities and the Standard Error of Measurement .....	29
Reliability of Gain Scores .....	32
Intercorrelations Among Scores .....	33
Validity .....	34
Content Validity .....	34
Criterion-Related Validity .....	35
Construct Validity .....	36
<b>Other TOEFL Programs and Services</b> .....	39
TWE Test (Test of Written English) .....	39
TSE Test (Test of Spoken English) .....	39
SPEAK Kit (Speaking Proficiency English Assessment Kit) .....	40
SLEP Test (Secondary Level English Proficiency Test) .....	40
Fee Voucher Service for TOEFL and TSE Score Users .....	40
TOEFL Fee Certificate Service .....	41
TOEFL Magnetic Score-Reporting Service .....	41
Examinee Identification Service for TOEFL and TSE Score Users .....	41
Support for External Research Studies .....	42
<b>Research Program</b> .....	43
TOEFL Research Report Series .....	43
TOEFL Technical Report Series .....	45
TOEFL Monograph Series .....	46
<b>Publications</b> .....	47
TOEFL Products and Services Catalog .....	47
Bulletin of Information for TOEFL, TWE, and TSE .....	47
Test Center Reference List .....	47
Test Forms Available to TOEFL Examinees .....	47
Guidelines for TOEFL Institutional Validity Studies .....	47
TOEFL Test and Score Data Summary .....	47
Institutional Testing Program Brochure .....	48
TOEFL Test of Written English Guide .....	48
TSE Score User's Manual .....	48
Secondary Level English Proficiency Test Brochure .....	48
The Researcher .....	48
<b>TOEFL Study Materials for the Paper-Based Testing Program</b> ..	49
<b>References</b> .....	50
<b>ETS Offices Serving TOEFL Candidates and Score Users</b> .....	54
<b>TOEFL Representatives</b> .....	55

# TABLES.....

<b>Table 1.</b> Minimum and Maximum Observed Section and Total Scores ..	23
<b>Table 2.</b> Reliabilities and Standard Errors of Measurement (SEM) .....	30
<b>Table 3.</b> Intercorrelations Among the Scores .....	33
<b>Table 4.</b> Correlations of Total TOEFL Scores with University Ratings ..	35
<b>Table 5.</b> Correlations of TOEFL Subscores with Interview and Essay Ratings .....	36
<b>Table 6.</b> TOEFL/GRE Verbal Score Comparisons .....	37
<b>Table 7.</b> TOEFL/SAT and TSWE Score Comparisons .....	37
<b>Table 8.</b> TOEFL, GRE, and GMAT Score Comparisons .....	37
<b>Table 9.</b> Correlations Between GMAT and TOEFL Scores .....	38





# OVERVIEW

The purpose of the Test of English as a Foreign Language (TOEFL®) is to evaluate the English proficiency of people whose native language is not English. The test was initially developed to measure the English proficiency of international students wishing to study at colleges and universities in the United States and Canada, and this continues to be its primary function. However, a number of academic institutions in other countries, as well as certain independent organizations, agencies, and foreign governments, have also found the test scores useful. The TOEFL test is recommended for students at the eleventh-grade level or above; the test content is considered too difficult for younger students.

The TOEFL test was developed for use starting in 1963-64 through the cooperative effort of more than 30 organizations, public and private. A National Council on the Testing of English as a Foreign Language was formed, composed of representatives of private organizations and government agencies concerned with testing the English proficiency of foreign nonnative speakers of English who wished to study at colleges and universities in the United States. The program was financed by grants from the Ford and Danforth Foundations and was, at first, attached administratively to the Modern Language Association. In 1965, the College Board® and Educational Testing Service (ETS®) assumed joint responsibility for the program.

In recognition of the fact that many who take the TOEFL test are potential graduate students, a cooperative arrangement for the operation of the program was entered into by Educational Testing Service, the College Entrance Examination Board, and the Graduate Record Examinations® (GRE®) Board in 1973. Under this arrangement, ETS is responsible for administering the TOEFL program according to policies determined by the TOEFL Policy Council.

**Educational Testing Service.** ETS is a nonprofit organization committed to the development and administration of responsible testing programs, the creation of advisory and instructional services, and research on techniques and uses of measurement, human learning and behavior, and educational development and policy formation. It develops and administers tests, registers examinees, and operates test centers for various sponsors. ETS also supplies related services; e.g., it scores tests; records, stores,

and reports test results; performs validity studies and other statistical studies; and undertakes program research. All ETS activities are governed by a 16-member board of trustees composed of persons from the fields of education and public service.

In addition to the Test of English as a Foreign Language and the Graduate Record Examinations, ETS develops and administers a number of other tests, including the Graduate Management Admission Test®, and The Praxis Series: Professional Assessments for Beginning Teachers® tests, as well as the College Board testing programs.

The Chauncey Group International Ltd., a wholly-owned subsidiary of ETS, provides assessment, training, and guidance products and services in the workplace, military, professional, and adult educational environments.

**College Board.** The College Board is a nonprofit, educational organization with a membership of more than 2,800 colleges and universities, schools, and educational associations and agencies. The College Board's board of trustees is elected from the membership, and institutional representatives serve on advisory councils and committees that review the programs of the College Board and participate in the determination of its policies and activities.

The College Board sponsors tests, publications, software, and professional conferences and training in the areas of guidance, admissions, financial aid, credit by examination, and curriculum improvement in order to increase student access to higher education. It also supports and publishes research studies about tests and measurement and conducts studies on education policy developments, financial aid need assessment, admissions planning, and related education management topics.

One major College Board service, the SAT® Program, includes the SAT I: Reasoning Test, and SAT II: Subject Tests. Subject Tests are available in such diverse content areas as writing, literature, languages, math, sciences, and history. The College Board contracts with ETS to develop these tests, operate test centers in the United States and other countries, score the answer sheets, and send score reports to examinees and to the institutions they designate as recipients.



### **Graduate Record Examinations Board.**

The GRE Board is an independent board affiliated with the Association of Graduate Schools and the Council of Graduate Schools in the United States and Canada. It is composed of 18 representatives of the graduate community. Standing committees of the board include the Research Committee, the Services Committee, and the Minority Graduate Education Committee.

ETS carries out the policies of the GRE Board and, under the auspices of the board, administers and operates the GRE program. Two types of tests are offered: a General Test and Subject Tests in 16 disciplines. ETS develops the tests, maintains test centers in the United States and other countries, scores the answer sheets, and sends score reports to the examinees and to the accredited institutions and approved fellowship sponsors the examinees designate as recipients. ETS also provides information, technical advice, and professional counsel, and develops proposals to achieve the goals formulated by the board.

In addition to its tests, the GRE program offers many services to graduate institutions and to prospective graduate students. Services to institutions include research, publications, and advisory services to assist graduate schools and departments in admissions, guidance, placement, and the selection of fellowship recipients. Services to students include test familiarization materials and services related to informing students about graduate education.

### **TOEFL Policy Council**

Policies governing the TOEFL program are formulated by the 15-member TOEFL Policy Council. The College Board and the GRE Board each appoint three members to the Council. These six members comprise the Executive Committee and elect the remaining nine members. Some of these members-at-large are affiliated with such institutions and agencies as graduate schools, junior and community colleges, nonprofit educational exchange organizations, and other public and private agencies with interest in international education. Others are specialists in the field of English as a foreign or second language.

There are six standing committees of the Council, each responsible for specific areas of program activity.

### **Committee of Examiners**

The TOEFL Committee of Examiners is composed of seven specialists in linguistics, language testing, or the teaching of English as a foreign or second language. Members are rotated on a regular basis to ensure the continued introduction of new ideas and philosophies related to second language teaching and testing.

The primary responsibility of this committee is to establish overall guidelines for the test content, thus assuring that the TOEFL test is a valid measure of English language proficiency reflecting current trends and methodologies in the field. The committee determines the skills to be tested, the kinds of questions to be asked, and the appropriateness of the test in terms of subject matter and cultural content. Committee members review and approve the policies and specifications that govern the test content.

The Committee of Examiners not only lends its own expertise to the test and the test development process but also makes suggestions for research and, on occasion, invites the collaboration of other authorities in the field, through invitational conferences and other activities, to contribute to the improvement of the test. The committee works with ETS test development specialists in the actual development and review of test materials.

### **Finance Committee**

The TOEFL Finance Committee consists of at least four members and is responsible to the TOEFL Executive Committee. The members develop fiscal guidelines, monitor and review budgets, and provide financial analysis for the program.



### **Research Committee**

An ongoing program of research related to the TOEFL program of tests is carried out under the direction of the Research Committee. Its six members include representatives of the Policy Council and the Committee of Examiners, as well as specialists from the academic community. The committee reviews and approves proposals for test-related research and sets guidelines for the entire scope of the TOEFL research program.

Because the studies involved are specific to the TOEFL testing programs, most of the actual research work is conducted by ETS staff members rather than by outside researchers. However, many projects require the cooperation of consultants and other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL office.

As research studies are completed, reports are published and made available to anyone interested in the TOEFL tests. A list of those in print at the time this *Manual* was published appears on pages 38-40.

### **Outreach and Services Committee**

This six-member committee is responsible for reviewing and making recommendations to improve and modify existing program outreach activities and services, especially as they relate to access and equity concerns; initiating proposals for the development of new program products and services; monitoring the Council bylaws; and carrying out additional tasks requested by the Executive Committee or the Council.

### **TWE® Test (Test of Written English) Committee**

This seven-member group consists of writing and ESL composition specialists with expertise in writing assessment and pedagogy.

The TWE Committee, with ETS test development specialists, is responsible for developing, reviewing, and approving test items for the TWE test. The committee also prepares item writer guidelines and may suggest research or make recommendations for improving the TWE test to ensure that the test is a valid measure of English writing proficiency.

### **TSE® Test (Test of Spoken English) Committee**

This committee has six members who have expertise in oral proficiency assessment and represent the TSE constituency.

The TSE Committee, with ETS test development specialists and program staff, oversees the TSE test content and scoring specifications, reviews test items and scoring procedures, and may make recommendations for research or test revisions to assure that the test is a valid measure of general speaking proficiency.



# PROGRAM DEVELOPMENTS.....

## TOEFL 2000

The TOEFL 2000 project is a broad effort under which language testing at ETS will evolve into the twenty-first century. The impetus for TOEFL 2000 came from the various constituencies, including TOEFL committees and score users. These groups have called for a new TOEFL test that (1) is more reflective of models of communicative competence; (2) includes more constructed-response items and direct measures of writing and speaking; (3) includes test tasks integrated across modalities; and (4) provides more information than current TOEFL scores about international students' ability to use English in an academic environment.

Changes to TOEFL introduced in 1995 (i.e., eliminating single-statement listening comprehension items, expanding the number of academic lectures and longer dialogs, and embedding vocabulary in reading comprehension passages) represented the first step toward a more integrative approach to language testing. The next major step will be the introduction of a computer-based TOEFL test in 1998. (See next column.)

TOEFL 2000 now continues with efforts that will lead to the next generation of computerized TOEFL tests. These include:

- the development of a conceptual framework that
  - takes into account models of communicative competence
  - identifies various task characteristics and how these will be used in the construction of language tasks
  - specifies a set of variables associated with each of these task components
- a research agenda that informs and supports this emerging framework
- a better understanding of the kinds of information test users need and want from the TOEFL test
- a better understanding of the technological capabilities for delivery of the TOEFL test into the next century

A series of TOEFL 2000 reports that are part of the foundation of the project are now available (see page 44). As future projects are completed, monographs will be released to the public in this new research publication series.

## The Computer-Based TOEFL Test

Testing on computer is an important advancement that enables the TOEFL program to take advantage of new forms of assessment made possible by the computer platform. This reflects ETS's commitment to create an improved English-language proficiency test that will

- better reflect the way in which people communicate effectively
- include more performance-based tasks
- provide more information than the current TOEFL test about the ability of international students to use English in an academic setting

The computer-based test is not just the paper test reformatted for the computer. While some questions will be similar to those on the current test, others will be quite different. For example, the Listening Comprehension and Reading Comprehension sections will include new question types designed specifically for the computer. In addition, the test will include an essay that can be handwritten or typed on the computer. The essay will measure an examinee's ability to generate and organize ideas and support those ideas using the conventions of standard written English.

Some sections of the test will be computer-adaptive. In computer-adaptive testing (CAT), the computer selects a unique set of test questions based on the test design and the test taker's ability level. Questions are chosen from a very large pool categorized by item content and difficulty. The test design ensures fairness because all examinees receive the same

- number of test questions
- amount of time (if they need it)
- directions
- question types
- distribution of content

The CAT begins with a question of medium difficulty. The next question is one that best fits the examinee's performance and the design of the test. The computer is programmed to make continuous adjustments in order to present questions of appropriate difficulty to test takers of all ability levels.

The TOEFL program has taken steps to assure that an individual's test performance is not influenced by a lack of computer experience. A computerized tutorial, designed especially for nonnative speakers of English, has been developed to teach the skills needed to take TOEFL on computer.

For periodic updates on the computer-based TOEFL test, visit TOEFL OnLine at <http://www.toefl.org>.

# TEST OF ENGLISH AS A FOREIGN LANGUAGE: The Paper-Based Testing Program.....

## Use of Scores

The TOEFL program encourages use of the test scores by an institution or organization to help make valid decisions concerning English language proficiency in terms of its own requirements. However, the institution or organization itself must determine whether the TOEFL test is appropriate, with respect to both the language skills it measures and its level of difficulty, and must establish its own levels of acceptable performance on the test. General guidelines for using TOEFL scores are given on pages 26-28.

TOEFL score users are invited to consult with the TOEFL program staff about their current or intended uses of the test results. The TOEFL office will assist institutions and organizations contemplating use of the test by providing information about its applicability and validity in particular situations. It also will investigate complaints or information obtained about questionable interpretation or use of reported TOEFL test scores.

## Description of the Paper-Based TOEFL Test

The TOEFL test originally contained five sections. As a result of extensive research (Pike, 1979; Pitcher and Ra, 1967; Swineford, 1971; *Test of English as a Foreign Language: Interpretive Information*, 1970), a three-section test was developed and introduced in 1976. In July 1995, the test item format was modified somewhat within the same three-section structure of the test.

Each form of the current (1997) TOEFL test consists of three separately timed sections delivered in a paper-and-pencil format; the questions in each section are multiple-choice, with four possible answers or options per question. All responses are gridded on answer sheets that are computer scored.

The total test time is approximately two and one-half hours; however, approximately three and one-half hours are needed for a test administration to admit examinees to the testing room, to allow them to enter identifying information on their answer sheets, and to distribute and collect the test materials. Brief descriptions of the three sections of the test follow.

## ■ Section 1, Listening Comprehension

Section 1 measures the ability to understand English as it is spoken in North America. The oral features of the language are stressed, and the problems tested include vocabulary and idiomatic expression as well as special grammatical constructions that are frequently used in spoken English. The stimulus material and oral questions are recorded in standard North American English; the response options are printed in the test books.

There are three parts in the Listening Comprehension section, each of which contains a specific type of comprehension task. The first part consists of a number of short conversations between two speakers, each followed by a single spoken question. The examinee must choose the best response to the question about the conversation from the four options printed in the test book. In the second and third parts of this section, the examinee hears conversations and short talks of up to two minutes in length. The conversations and talks are about a variety of subjects, and the factual content is general in nature. After each conversation or talk the examinee is asked several questions about what was heard and, for each, must choose the one best answer from the choices in the test book. Questions for all parts are spoken only one time.



## ■ Section 2, Structure and Written Expression

Section 2 measures recognition of selected structural and grammatical points in standard written English. The language tested is formal, rather than conversational. The topics of the sentences are of a general academic nature so that individuals in specific fields of study or from specific national or linguistic groups have no particular advantage. When topics have a national context, they refer to United States or Canadian history, culture, art, or literature. However, knowledge of these contexts is not needed to answer the structural or grammatical points being tested.

This section is divided into two parts. The first part tests an examinee's ability to identify the correct structure needed to complete a given sentence. The examinee reads incomplete sentences printed in the test book. From the four responses provided for each incomplete sentence, the examinee must choose the word or phrase that best completes the given sentence. Only one of the choices fits correctly into the particular sentence. The second part tests an examinee's ability to recognize correct grammar and to detect errors in standard written English. Here the examinee reads sentences in which some words or phrases are underlined. The examinee must identify the one underlined word or phrase in each sentence that would not be accepted in standard written English.

## ■ Section 3, Reading Comprehension

Section 3 measures the ability to read and understand short passages that are similar in topic and style to those that students are likely to encounter in North American colleges and universities. The examinee reads a variety of short passages on academic subjects and answers several questions about each passage. The questions test information that is stated in or implied by the passage, as well as knowledge of some of the specific words as they are used in the passage. To avoid creating an advantage to individuals in any one field of study, sufficient context is provided so that no subject-specific familiarity with the subject matter is required to answer the questions. Questions are asked about factual information presented in the passages, and examinees may also be asked to make inferences or recognize analogies. In all cases, the questions can be answered by reading and understanding the passages.

## Development of TOEFL Test Questions

Material for the TOEFL test is prepared by language specialists who are trained in writing questions for the test before they undertake actual item-writing assignments. Additional material is prepared by ETS test development specialists. The members of the TOEFL Committee of Examiners establish overall guidelines for the test content and specifications. All item specifications, questions, and final test forms are reviewed internally at ETS for cultural and racial bias and content appropriateness, according to established ETS procedures.

These reviews ensure that each final form of the test is free of any language, symbols, references, or content that might be considered potentially offensive or inappropriate for subgroups of the TOEFL test population, or that might serve to perpetuate negative stereotypes.

All questions are pretested on representative groups of international students who are not native speakers of English. Only after the results of the pretest questions have been analyzed for statistical and content appropriateness are questions selected for the final test forms.

Following the administration of each new form of the test, a statistical analysis of the responses to questions is conducted. On rare occasions, when a question does not function as expected, it will be reviewed again by test specialists. After this review, the question may be deleted from the final scoring of the test. The statistical analyses also provide continuous monitoring of the level of difficulty of the test, the reliability of the entire test and of each section, intercorrelations among the sections, and the adequacy of the time allowed for each section. (See "Statistical Characteristics of the Test," page 29.)

# TOEFL TESTING PROGRAMS.....

The TOEFL test is administered internationally on regularly scheduled test dates through the Friday and Saturday testing programs. It is also administered at local institutions around the world through the Institutional Testing Program (ITP). The ITP program does not provide official TOEFL score reports; scores are for use by the administering institution only.

## Friday and Saturday Testing Programs

The official TOEFL test is given at centers around the world one day each month – five Fridays and seven Saturdays.

The TOEFL office diligently attempts to make the test available to all individuals who require TOEFL scores. In 1996-97, more than 1,275 centers located in 180 countries and areas were established for the Saturday testing program to accommodate the more than 703,000 persons registered to take the test; 350 centers in more than 60 countries and areas were established for the more than 248,000 persons registered to take TOEFL under the Friday program.

Registration and administration procedures are identical for the Friday and Saturday programs. The test itself is also identical in terms of format and content. Score reports for administrations under both programs provide the same data. More information about these testing programs can be found in the *Bulletin of Information for TOEFL, TWE, and TSE*. (See page 47.)

As noted above, the TOEFL program provides 12 test dates a year. However, the actual number of administrations at any one center in a given country or area is scheduled according to demand and the availability of space and supervisory staff.

There are sometimes local scheduling conflicts with national or religious holidays. Although the TOEFL office makes every effort to avoid scheduling administrations of the test on such dates, it may be unavoidable in some cases.

Registration must be closed well in advance of each test date to ensure the delivery of test materials to the test centers. Registration deadline dates are about seven weeks before the test dates for centers outside the United States and Canada and five weeks before the test dates for centers within these two countries.

Almost all administrations are held as scheduled. On occasion, however, shipments of test materials may be impounded by customs officials or delayed by mail embargoes or transportation strikes. Other problems, ranging from political disturbances within countries, to power failures, to the last-minute illness of a test supervisor, may also force postponement of a TOEFL test administration.

If an administration must be postponed, a makeup administration is scheduled, usually on the next regularly scheduled test date. Occasionally it is necessary to arrange a makeup administration on another date.

Different forms of the test may be used at a single administration. Following each administration, the answer sheets are returned to ETS for scoring; test results are mailed to score recipients about one month after the answer sheets are received at ETS.

## TWE Test (Test of Written English)

In 1986, the TOEFL program introduced the Test of Written English. This direct assessment of writing proficiency was developed in response to requests from many colleges, universities, and agencies that use TOEFL scores. The TWE test is currently (1997) a required section of the TOEFL test at five administrations per year. For more information about the Test of Written English, see page 39.

## TSE Test (Test of Spoken English)

The Test of Spoken English measures the ability of nonnative speakers of English to communicate orally in English. It requires examinees to tape record spoken answers to a variety of questions. The TSE test is administered on all 12 Friday and Saturday TOEFL test dates. For more information about the Test of Spoken English, see page 39.

## Institutional Testing Program

The Institutional Testing Program permits approved institutions throughout the world to administer the TOEFL test to their own students on dates convenient for them (except for regularly scheduled TOEFL administration dates), using their own facilities and staff. Each year a number of forms of the TOEFL test previously used in the Friday and Saturday testing programs are made available for the Institutional Testing Program.

In addition to the regular TOEFL test, which is especially appropriate for use with students at the intermediate and higher levels of English language proficiency, ITP offers the Preliminary Test of English as a Foreign Language (Pre-TOEFL) for individuals at the beginning level. Pre-TOEFL measures the same components of English language skills as the TOEFL test. However, Pre-TOEFL is less difficult and shorter. Pre-TOEFL test results are based on a restricted scale that provides more discriminating measurement at the lower end of the TOEFL scale.

**Note:** There are minor differences in the number of questions and question types between the ITP TOEFL test and the Pre-TOEFL test.

### **How Institutional TOEFL Can Be Used**

The Institutional Testing Program is offered primarily to assist institutions in placing students in English courses at the appropriate level of difficulty, for determining whether additional work in English is necessary before an individual can undertake academic studies, or as preparation for an official Friday or Saturday TOEFL administration.

### **Institutional TOEFL Test Scores**

Scores earned under the Institutional Testing Program are comparable to scores earned under the worldwide Friday and Saturday testing programs. *However, ITP scores are for use by the administering institution only.*

ETS reports test results to the administering institution in roster form, listing the names and scores (section and total) of all students who took the test at that administration. Two copies of the score record for each student are provided to the administering institution: a file copy for the institution and a personal copy for the individual. Both copies indicate that the scores were obtained at an Institutional Testing Program administration.

ETS does not report scores obtained under this program to other institutions as it does for official scores obtained under the Friday and Saturday testing programs. ***To ensure score validity, scores obtained under the Institutional Testing Program should not be accepted by other institutions to evaluate an individual's readiness to begin academic studies in English.***



# PROCEDURES AT TEST CENTERS.....

Standard, uniform procedures are important in any testing program, but are essential for an examination that is given worldwide. Therefore, the TOEFL program provides detailed guidelines for test center supervisors to ensure uniform administrations.

*Preparing for a TOEFL/TWE or TSE Administration* is mailed to test supervisors well in advance of the test date. This publication describes the arrangements the supervisor must make to prepare for the test administration, including selecting testing rooms and the associate supervisors and proctors who will be needed on the day of the test.

The *Manual for Administering TOEFL*, included with every shipment of test materials, describes appropriate seating plans, the kind of equipment that should be used for the Listening Comprehension section, identification requirements, the priorities for admitting examinees to the testing room, and instructions for distributing and collecting test materials. It also contains detailed instructions for the actual administration of the test.

TOEFL program staff work with test center supervisors to ensure that the same practices are followed at all centers, and they conduct workshops during which supervisors can discuss procedures for administering the test. TOEFL staff respond to all inquiries from supervisors and examinees regarding circumstances or conditions associated with test administrations, and they investigate all complaints received about specific administrations.

## Measures to Protect Test Security

In administering a worldwide testing program at more than 1,275 test centers in 180 countries, the TOEFL program considers the maintenance of security at testing sites to be of paramount importance. The elimination of problems at test centers, including test-taker impersonations, is a continuing goal. To offer score users the most valid, reliable, and secure measurements of English language proficiency available, the TOEFL office continuously reviews and refines procedures to increase the security of the test before, during, and after its administration.

Because of the importance of TOEFL test scores to examinees and institutions, it is inevitable that some individuals will engage in practices designed to increase their reported scores. The careful selection of supervisors, a high proctor-to-examinee ratio, and carefully developed procedures for the administration of the test

(explained in the *Manual for Administering TOEFL*) are measures designed to prevent or discourage examinee attempts at impersonation, copying, theft of test materials, and the like, and thus to protect the integrity of the test for all examinees and score recipients.

## Identification Requirements

Strict admission procedures are followed at all test centers to prevent attempts by some examinees to have others with greater proficiency in English impersonate them at a TOEFL administration. To be admitted to a test center, every examinee must present an official document with a recognizable photograph and a completed photo file record with a recent photo attached. Although the passport is the basic document that is acceptable at all test centers, other *specific* photobearing documents may be acceptable for individuals who may not be expected to have passports or who are taking the test in their own countries.

Through embassies in the United States and TOEFL representatives and supervisors in other countries, the TOEFL office continually verifies the names of official, secure, photobearing identification documents used in each country, such as national identity cards, work permits, and registration certificates. In the Friday and Saturday testing programs, each admission ticket contains a statement specifying the documents that will be accepted at TOEFL test centers in the country in which the examinee is registered to take the test. This information is computer-printed on a red field to ensure that it will be seen. (The same information is printed on the attendance roster prepared for each center.) Following is a sample of the statement that appears on admission tickets for Venezuela.

**YOUR VALID PASSPORT. CITIZENS OF VENEZUELA MAY USE NATIONAL IDENTITY CARD OR LETTER AS DESCRIBED IN THE BULLETIN.**

Complete information about identification requirements is included in all editions of the *Bulletin of Information for TOEFL, TWE, and TSE*.



### Photo File Records

Every TOEFL examinee must present a completed photo file record to the test center supervisor before being admitted to the testing room. The photo file record contains the examinee's name, registration number, test center code, and signature, as well as a recent photo that clearly identifies the examinee (that is, the photo must look exactly like the examinee, with the same hairstyle, with or without a beard, and so forth). The photo file records are collected at the test center and returned to ETS, where the photos and identifying information are electronically captured and included on the examinee's score data file.

### Photo Score Reporting

As an additional procedure to help eliminate the possibility of impersonation at test centers, the official score reports that are *routinely* sent to institutions designated by the test taker, and the examinee's own copy of the score report, bear an electronically reproduced photo image of the examinee and his or her signature. (The score report also includes the

number of the passport or other identification document used to gain admission to the testing center and the name of the country issuing the document.) Examinees are advised in the *Bulletin of Information* that the score reports will contain these photo images. In addition to strengthening security through this deterrent to impersonation, the report form provides score users with the immediate information they may need to resolve any issues of examinee identity. Key features of the image score reports are highlighted on page 19.

### Checking Names

To prevent examinee attempts to exchange answer sheets or to grid another person's name (for whom he or she is taking the test) on the answer sheet, supervisors are asked to compare names on the identification document and the answer sheet and also to check the gridding of names on the answer sheet before examinees leave the room.

## Supervision of Examinees

Supervisors and proctors are instructed to exercise extreme vigilance during a test administration to prevent examinees from giving or receiving assistance in any way.

In addition, the *Manual for Administering TOEFL* advises supervisors about assigning seats to examinees. To prevent copying from notes or other aids, examinees may not have anything on their desks but their test books, answer sheets, pencils, and erasers. They are not permitted to make notes or marks of any kind in their test books. (Warning/Dismissal Notice forms are used to report examinees who violate procedures. An examinee is asked to sign the notice to document the violation and to indicate he or she understands that a violation of procedures has occurred and that the answer sheet may not be scored.)

If a supervisor is certain that someone has given or received assistance, the supervisor has the authority to dismiss the examinee from the testing room; scores for dismissed examinees will not be reported. If a supervisor **suspects** someone of cheating, the examinee is warned about the violation, is asked to sign a Warning/Dismissal Notice, and must move to another seat selected by the supervisor. A description of the incident is written on the Supervisor's Irregularity Report, which is returned to ETS with the answer sheet. Both suspected and confirmed cases of cheating are investigated by the Test Security Office at ETS. (See "Scores of Questionable Validity," page 23.)

Turning back to another section of the test, working on a section in advance, or continuing to work on a section after time is called are not permitted and are considered cheating. (To assist the supervisor, a large number identifying the section being worked on is printed at the top of each page of the test book.) Supervisors are instructed to warn anyone found working on the wrong section and to ask the examinee to sign a Warning/Dismissal Notice.

## Preventing Access to Test Materials

To ensure that examinees have not seen the test material in advance, a new form of the test is developed for each Friday and Saturday administration.

To prevent the theft of test materials, procedures have been devised for the distribution and handling of these materials. Test books are individually sealed, then packed and sealed in plastic bags. Test books, answer sheets, and Listening Comprehension recordings are sent to test centers in sealed boxes and are placed in secure, locked storage that is inaccessible to unauthorized persons. Supervisors are directed to count the test books several times — upon receipt, during the test administration, and after the test is over. No one is permitted to leave the testing room until the supervisor has accounted for all test materials. Except for "disclosed" administrations, when examinees may obtain the test book (see "Test Forms Available to TOEFL Examinees," page 47), supervisors must follow detailed directions for returning the test materials. Materials are counted upon receipt at ETS, and its Test Security Office investigates all cases of missing test materials.

# TOEFL TEST RESULTS.....

## Release of Test Results

About one month after a Friday or Saturday TOEFL administration, test results are mailed to the examinees and to the official score recipients they have specified, provided that the answer sheets are received at ETS promptly after the administration. Test results for examinees whose answer sheets are incomplete or whose answer sheets arrive late are usually sent two or three weeks later. All test results are mailed by the final deadline — 12 weeks after the test.

For the basic TOEFL test fee, each examinee is entitled to four copies of the test results: one copy is sent to the examinee, and up to three official score reports are sent directly by ETS to the institutions whose assigned code numbers the examinee has marked on the answer sheet.\* The institution code designates the recipient college, university, or agency. A list of the most frequently used institution and agency codes is printed in the *Bulletin of Information*. An institution whose code number is not listed should give applicants its code number before they take the test. (See page 20 for more information.)

The most common reason that institutions do not receive score reports following an administration is that examinees do not properly specify the institutions as score report recipients by marking the correct codes on the test answer sheet. (Examinees cannot write the names of recipients on the answer sheet.) An examinee who wants scores sent to an institution whose code number was not marked on the answer sheet must submit a Score Report Request Form naming the institution that is to receive the scores. There is a fee for this service.

## Test Score Data Retention

Language proficiency can change considerably in a relatively short period. Therefore, the TOEFL office will not report scores that are more than two years old. Individually identifiable TOEFL scores are retained on the TOEFL database for only two years from the date of the test. Individuals who took the TOEFL test more than two years ago must take it again if they want scores sent to an institution.\* After two years, all information that could be used to identify an individual is removed from the database. Score data and other information that may be used for research or statistical purposes do not include individual examinee identification information and are retained indefinitely.

## Image Score Reports

The image-processing technology used to produce the photo score reports allows ETS to electronically capture the image from the examinee's photograph, as well as the signature and other identifying data submitted by the examinee at the testing site, and to reproduce these with the examinee's test results directly on the score reports. The computerized electronic transfer of photo images permits a high-quality reproduction of the original photo on the score report. (If a photograph is too damaged or for other reasons cannot be accepted by the image-processing system, "Photo Not Available" will be printed on the score report.)

Steps have been taken to reduce the opportunities for tampering with examinee score records that institutions may receive directly from applicants. **However, to ensure that institutions receive valid score records, we urge that admissions officers and others responsible for the admissions process accept only official score reports sent directly by ETS.**

\* An institution or agency that is sponsoring an examinee and has made prior arrangements with the TOEFL office will also receive a copy of the examinee's official score report if the examinee has given permission to the TOEFL office.

\* A TOEFL score is measurement information and is subject to all the restrictions noted in this *Manual*. (These restrictions are also noted in the *Bulletin of Information*.) The test score is not the property of the examinee.

## Official Score Reports from ETS

TOEFL score reports give the score for each of the three sections of the test and the total score. Examinees who take the TOEFL test during an administration at which the Test of Written English is given also receive a TWE score printed in a separate field on the TOEFL score report. See page 20 for information about the score report codes.

### Features of the Image Reports:

- 1 The blue background color quickly identifies the report as being an official copy sent from ETS.
- 2 The examinee's name and scores are printed in red fields.
- 3 Reverse type is used for printing the name and scores.
- 4 The examinee's photo is taken from the photo file record given to the test center supervisor on the day of the test and reproduced on the score report.

- 5 The examinee's signature and ID number and the name of the country issuing identification are reproduced from the photo file record.
- 6 The word "copy" appears in the background color of score reports that are photocopied using either a black or color image copier.

Score reports are valid only if received directly from Educational Testing Service. TOEFL test scores are confidential and should not be released by the recipient without written permission from the examinee. All staff with access to score records should be advised of their confidential nature.

If you have any reason to believe that someone has tampered with a score report or would like to verify test scores, please call the following toll-free number between 8:30 AM and 4:30 PM New York time.

**800-257-9547**

TOEFL/TSE Services will verify the accuracy of the scores.

**TOEFL**® OFFICIAL SCORE REPORT

Test of English as a Foreign Language  
P.O. Box 6151 • Princeton, NJ 08541-6151 • USA

REGISTRATION NUMBER: 1234567 NAME (Family or Surname, Given, Middle): SALVAN KIM TORRES

JUL 97 Month Year TEST DATE		L340 CENTER NUMBER	11/04/78 Month/Day/Year DATE OF BIRTH	F SEX
0115 00 INST. CODE DEPT. CODE		1996-97 INTERPRETIVE INFORMATION	PHILIPPINES NATIVE COUNTRY	
UNDERGRADUATE DEPARTMENT		TAGALOG NATIVE LANGUAGE		
		1 DEGREE	2 REASON FOR TAKING TOEFL	1 TOEFL TAKEN BEFORE

TEST OF ENGLISH AS A FOREIGN LANGUAGE

41 SECTION 1	44 SECTION 2	47 SECTION 3	440 TOTAL SCORE
-----------------	-----------------	-----------------	--------------------

TOEFL SCALED SCORES

TEST OF WRITTEN ENGLISH

WRITING TEST NOT ADMINISTERED TWE SCORE

EXAMINEE'S ADDRESS:

SALVAN KIM TORRES  
465 QUEZON BLVD  
APARTADO 3-B  
MANILA, PHILIPPINES

YOUR SIGNATURE: *Kim Torres Salvan*  
NAME OF COUNTRY ISSUING PASSPORT OR IDENTIFICATION: *Philippines*  
NUMBER ON IDENTIFICATION DOCUMENT: *TS 2176239*

**COPY**

The face of this document has a multicolored background — not a white background.

Facsimile reduced. Actual size of entire form, 8<sup>1</sup>/<sub>2</sub>" x 11"; score report section 8<sup>1</sup>/<sub>2</sub>" x 3<sup>5</sup>/<sub>8</sub>".

## Information Printed on the Official Score Report

In addition to test scores, native country, native language, and birth date, the score report includes other pertinent data about the examinee and information about the test.

**INSTITUTION CODE.** The institution code designates the recipient college, university, or agency. A list of the most frequently used institution and agency codes is printed in the Bulletin of Information. An institution whose code number is not listed should give applicants its code number before they take the test. (This information should be included in application materials prepared for international students.)

**Note:** An institution that does not know its TOEFL code number or wishes to obtain one should call 609-771-7975 or write to ETS Code Control, P.O. Box 6666, Princeton, NJ 08541-6666, USA.

**DEPARTMENT CODE.** The department code number identifies the professional school, division, department, or field of study in which the graduate applicant plans to enroll. The department code list shown below is also included in the Bulletin of Information. The department code for all business schools is (02), for law schools (03), and for unlisted departments (99).

### Fields of Graduate Study Other Than Business or Law

#### HUMANITIES

- 11 Archaeology
- 12 Architecture
- 26 Art History
- 13 Classical Languages
- 28 Comparative Literature
- 53 Dramatic Arts
- 14 English
- 29 Far Eastern Languages and Literature
- 15 Fine Arts, Art, Design
- 16 French
- 17 German
- 04 Linguistics
- 19 Music
- 57 Near Eastern Languages and Literature
- 20 Philosophy
- 21 Religious Studies or Religion
- 22 Russian/Slavic Studies
- 23 Spanish
- 24 Speech
- 10 Other foreign languages
- 98 Other humanities

#### SOCIAL SCIENCES

- 27 American Studies
- 81 Anthropology
- 82 Business and Commerce
- 83 Communications
- 84 Economics
- 85 Education (including M.A. in Teaching)
- 01 Educational Administration
- 70 Geography
- 92 Government
- 86 History
- 87 Industrial Relations and Personnel
- 88 International Relations
- 18 Journalism
- 90 Library Science
- 91 Physical Education
- 97 Planning (City, Community, Regional, Urban)
- 92 Political Science
- 93 Psychology, Clinical
- 09 Psychology, Educational
- 58 Psychology, Experimental/Developmental
- 79 Psychology, Social
- 08 Psychology, other
- 94 Public Administration
- 50 Public Health
- 95 Social Work
- 96 Sociology
- 80 Other social sciences

#### BIOLOGICAL SCIENCES

- 31 Agriculture
- 32 Anatomy
- 05 Audiology
- 33 Bacteriology
- 34 Biochemistry
- 35 Biology
- 45 Biomedical Sciences
- 36 Biophysics
- 37 Botany
- 38 Dentistry
- 39 Entomology
- 46 Environmental Science
- 40 Forestry
- 06 Genetics
- 41 Home Economics
- 25 Hospital and Health Services Administration
- 42 Medicine
- 07 Microbiology
- 74 Molecular and Cellular Biology
- 43 Nursing
- 77 Nutrition
- 44 Occupational Therapy
- 56 Pathology
- 47 Pharmacy
- 48 Physical Therapy
- 49 Physiology
- 55 Speech-Language Pathology
- 51 Veterinary Medicine
- 52 Zoology
- 30 Other biological sciences

#### PHYSICAL SCIENCES

- 54 Applied Mathematics
- 61 Astronomy
- 62 Chemistry
- 78 Computer Sciences
- 63 Engineering, Aeronautical
- 64 Engineering, Chemical
- 65 Engineering, Civil
- 66 Engineering, Electrical
- 67 Engineering, Industrial
- 68 Engineering, Mechanical
- 69 Engineering, other
- 71 Geology
- 72 Mathematics
- 73 Metallurgy
- 75 Oceanography
- 76 Physics
- 59 Statistics
- 60 Other physical sciences

**Use 99 for any department not listed.**

**TOEFL SCORES:** Three section scores and a total score are reported for the TOEFL test. The three sections are:

- Section 1 — Listening Comprehension
- Section 2 — Structure and Written Expression
- Section 3 — Reading Comprehension

**TEST OF WRITTEN ENGLISH (TWE):** Effective July 1995, the TWE test is administered in August, October, December, February, and May.

Scores	Explanations of TWE Scores
<b>6.0</b>	Demonstrates clear competence in writing on both the rhetorical and syntactic levels, though the essay may have occasional errors.
<b>5.5</b>	
<b>5.0</b>	Demonstrates competence in writing on both the rhetorical and syntactic levels, though the essay will probably have occasional errors.
<b>4.5</b>	
<b>4.0</b>	Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.
<b>3.5</b>	
<b>3.0</b>	Demonstrates some developing competence in writing, but the essay remains flawed on either the rhetorical or syntactic level, or both.
<b>2.5</b>	
<b>2.0</b>	Suggests incompetence in writing.
<b>1.5</b>	
<b>1.0</b>	Demonstrates incompetence in writing.
<b>1NR</b>	Examinee did not write an essay.
<b>OFF</b>	Examinee did not write on the assigned topic.

**INTERPRETIVE INFORMATION:** The date of the most current edition of the TOEFL Test and Score Manual is printed here. (This date is printed only on the official score report.)

**TEST DATE:** Because English proficiency can change considerably in a relatively short period, please note the date on which the test was taken. Scores more than two years old cannot be reported, nor can they be verified.

#### PLANS TO WORK FOR DEGREE:

1 = Yes      2 = No      0 = Not answered

#### REASON FOR TAKING TOEFL:

- 1 = To enter a college or university as an undergraduate student
- 2 = To enter a college or university as a graduate student
- 3 = To enter a school other than a college or university
- 4 = To become licensed to practice a profession
- 5 = To demonstrate proficiency in English to the company for which the examinee works or expects to work
- 6 = Other than above
- 0 = Not answered

#### NUMBER OF TIMES TOEFL TAKEN BEFORE:

1 = One                      3 = Three                      0 = None or not  
2 = Two                      4 = Four or more                      answered

## Examinee Score Records

Examinees receive their test results on a form titled Examinee's Score Record. **These are NOT official TOEFL score reports and should not be accepted by institutions.**

### Acceptance of Test Results Not Received from ETS

Bear in mind that examinees may attempt to alter score records. Institution and agency officials are urged to verify all TOEFL scores supplied by examinees. TOEFL/TSE Services will either confirm or deny the accuracy of the scores submitted by examinees.

If there is a discrepancy between the official scores recorded at ETS and those submitted in any form by an examinee, the institution will be requested to send ETS a copy of the score record supplied by the examinee. At the written request of an official of the institution, ETS will report the official scores, as well as all previous scores recorded for the examinee within the last two years. Examinees are advised of this policy in the *Bulletin*, and, in signing their completed registration forms, they accept these conditions. (Also see "Test Score Data Retention" on page 18.)

### How to Recognize an Unofficial Score Report:

- 1 ★★ ★ Examinee's Original Score Record ★★ ★ is printed at the bottom of the score record.
- 2 The Examinee's Score Record is printed on white paper.

### How to Recognize If a Score Report Has Been Altered:

- 3 The last digit of the total score should end in "0," "3," or "7."
- 4 There should be no erasures. Do the shaded areas seem lighter than others, or are any of these areas blurred?
- 5 The typeface should be the same in all areas.

<b>TOEFL</b> <sup>®</sup> EXAMINEE'S SCORE RECORD	REGISTRATION NUMBER	1234567			NAME (Family or Surname, Given, Middle)			SALUAN KIM TORRES			
	TEST OF ENGLISH AS A FOREIGN LANGUAGE		JUL 97		2034	02	11/04/78	F	DATE OF BIRTH SEX		
	SECTION 1	SECTION 2	SECTION 3	TOTAL SCORE		PHILIPPINES					
	45	51	51	490		NATIVE COUNTRY					
TEST OF WRITTEN ENGLISH				L340		INST. CODE		DEPT. CODE		TAGALOG	
TWE NOT OFFERED ON THIS TEST DATE				9.9		SPONSOR CODE		NATIVE LANGUAGE		1 2 1	
EXAMINEE'S ADDRESS:				CENTER NUMBER		REASON FOR TAKING TOEFL		TOEFL TAKEN BEFORE		DEGREE	
SALUAN KIM TORRES 465 QUEZON BLVD APARTADO 3-B MANILA, PHILIPPINES				1		2		1		YOUR SIGNATURE NAME OF COUNTRY ISSUING PASSPORT OR IDENTIFICATION NUMBER ON IDENTIFICATION DOCUMENT	
Test of English as a Foreign Language • P. O. Box 6151 • Princeton, NJ 08541-6151 • USA				★ ★ ★ EXAMINEE'S ORIGINAL SCORE RECORD ★ ★ ★		Kim Torres Saluan		Philippines		TS 2176239	

Facsimile reduced

### DOs and DON'Ts

**Do** verify the information on an examinee's score record by calling TOEFL/TSE Services:

**800-257-9547**

**Don't** accept scores that are more than two years old.

**Don't** accept score reports from another institution that were obtained under the TOEFL Institutional Testing Program.

**Don't** accept photocopies of score reports.

### Additional Score Reports

Individuals who have taken the TOEFL test at scheduled Friday or Saturday test administrations may request that official score reports be sent to additional institutions at any time up to two years after the date on which they took the test.

There are two score reporting services: (1) **regular** and (2) **rush** reporting. The regular service mails additional score reports within two weeks after receipt of an examinee's Score Report Request Form. The rush reporting service mails score reports to institutions within four working days after a request form has been received. There is an additional fee for the rush service.

### Confidentiality of TOEFL Scores

Information retained in TOEFL test files about an examinee's native country, native language, and the institutions to which the test scores have been sent, as well as the actual scores, is the same as the information printed on the examinee's score record and on the official score reports. An official score report will be sent only at the written consent of the examinee to those institutions or agencies designated on the answer sheet by the examinee on the day of the test, on a Score Report Request Form submitted at a later date, or otherwise specifically authorized by the examinee.\*

To ensure the authenticity of scores, the TOEFL program office urges that institutions accept only official copies of TOEFL scores received directly from ETS.

Score users are responsible for maintaining the confidentiality of an individual's score information. Scores are not to be released by the institutional recipient without the explicit permission of the examinee. Dissemination of score records should be kept to a minimum, and all staff with access to them should be informed of their confidential nature.

The TOEFL program recognizes the right of institutions as well as individuals to privacy with regard to information supplied by and about them that is stored in data or research files held by ETS and the concomitant responsibility to safeguard information in its files from unauthorized disclosure. As a consequence, information about an institution (identified by name) will be released only in a manner consistent with a prior agreement, or with the explicit consent of the institution.

### Calculation of TOEFL Scores

The raw scores for the three sections of the TOEFL test are the number of questions answered correctly. No penalty points are subtracted for wrong answers. Although each new form of the test is constructed to match previous forms in terms of content and difficulty, the level of difficulty may vary slightly from one form to another. Raw scores from each new TOEFL test are statistically adjusted, or equated, to account for relatively minor differences in difficulty across forms, thereby allowing scores from different forms of the test to be used interchangeably.

At the time of the first administration of the three-section TOEFL test (1976), the scale for reporting the total score was linked to the scale that was then in use for the original five-section test. Since April 1996 the scale has been maintained by linking current tests to the scale of the July 1995 initial revised TOEFL test.

The three separate sections are scaled so the mean scaled score for each section equals one-tenth of the total scaled score mean (the standard deviations of the scaled scores for the three sections are equal) and the total score equals ten-thirds times the sum of the three section scaled scores.

\* See footnote on page 18.

This method of scaling results in rounded scores for which the last digit can take on only three values: zero, three, or seven.

**Example:**

$$\begin{array}{rccccccc} \text{Section 1} & & \text{Section 2} & & \text{Section 3} & & \text{Sum} \\ 46 & + & 54 & + & 50 & = & 150 \\ & & & & & & (150 \times 10) \div 3 = 500 \end{array}$$

TOEFL scores for Sections 1 and 2 are reported on a scale that can range from 20 to 68. Section 3 scores range from 20 to 67. TOEFL total scores are reported on a scale that can range from 200 to 677.

Scores for each new test form are converted to the same scale by a statistical equating procedure known as item response theory (IRT) true score equating, which determines equivalent scaled scores for persons of equal ability regardless of the difficulty level of the particular form of the test and the average ability level of the group taking the test.\*

The reported scores are not based on either the number or the percentage of questions answered correctly. Nor are they related to the distribution of scores on any other test, such as the SAT or the GRE tests.

Actual ranges of observed scores for the period from July 1995 through June 1996 are shown in Table 1. Note that for the section and total scores, all minimum observed section and total scores are higher than the lowest possible scores.

**Table 1. Minimum and Maximum Observed Section and Total Scores, July 1995 - June 1996**

Section	Min.	Max.
1. Listening Comprehension	25	68
2. Structure and Written Expression	21	68
3. Reading Comprehension	22	67
Total Score	263	677

\* See Cook and Eignor (1991) for further information about IRT true score equating.

## Hand-Scoring Service

Examinees are responsible for properly completing their answer sheets to ensure accurate scoring. They are instructed to use a medium-soft black lead pencil, to mark only one answer to each question, to fill in the answer space completely so the letter inside the space cannot be seen, and to erase all extra marks thoroughly. Failure to follow any of these instructions may result in the reporting of an inaccurate score.

Examinees who question whether their reported scores are accurate may request that their answer sheets be hand scored. There is a fee for this service. A request for hand scoring must be received within six months of the test date; later requests cannot be honored.

The TOEFL office has established the following hand-scoring procedures: the answer sheet to be hand scored is first confirmed as being the one completed by the person requesting the service; the answer sheet is then hand scored twice by trained ETS staff working independently. If there is a discrepancy between the hand-scored and computer-scored results, the hand-scored results, *which may be higher or lower than those originally reported*, will be reported to all recipients of the earlier scores, and the hand-scoring fee will be refunded to the examinee. The results of the hand scoring are available about three weeks after receipt of the examinee's request. Experience has shown that very few score changes result from hand-scoring requests.

## Scores of Questionable Validity

Improved scores over time are to be expected if a person is studying English; they may not indicate irregularities. However, institutions and other TOEFL score recipients that note inconsistencies between test scores and English performance, especially in cases where there is reason to suspect an inconsistency between a high TOEFL score and relatively weak English proficiency, are encouraged to refer to the official photo score report for the possibility of impersonation. Institutions should notify the TOEFL office if they find any evidence of impersonation. ETS reports TOEFL scores for a period of two years after the date the test was administered.



Irregularities uncovered by institutions and reported to ETS, as well as those brought to the attention of the TOEFL office by examinees or supervisors who believe that misconduct may have taken place, are investigated.

Misconduct irregularities are reviewed, statistical analyses are conducted, and scores may be canceled by ETS. For other irregularities, the ETS Test Security Office assembles relevant documents, such as previous score reports, registration forms, and answer sheets. When handwriting differences or evidence of possible copying or exchange of answer sheets is found, the case is referred to the ETS Board of Review, a group of senior professional staff members. Based on its independent examination of the evidence, the Board of Review directs appropriate action.

ETS policy and procedures are designed to provide reasonable assurance of fairness to examinees in both the identification of suspect scores and the weighing of information leading to possible score cancellation. These procedures are intended to protect both score users and examinees from inequities that could result from decisions based on fraudulent scores and to maintain the integrity of the test.

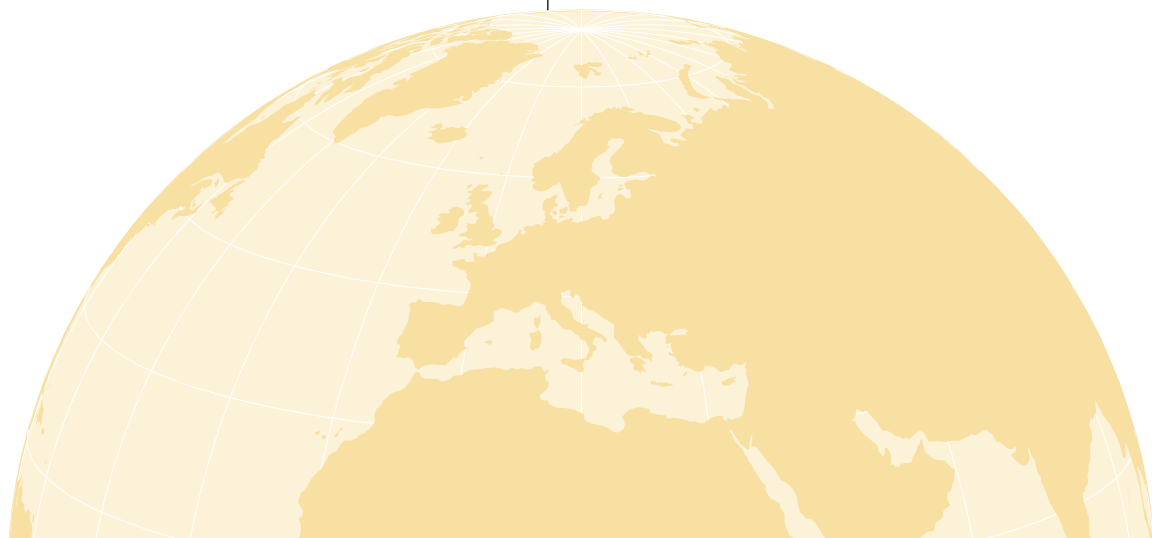
### **Examinees with Disabilities**

Nonstandard testing arrangements may include special editions of the test, the use of a reader and/or amanuensis, a separate testing room, and extended time and/or rest breaks during the test administration.

Nonstandard administrations are given on regularly scheduled test dates, and security procedures are the same as those followed for standard administrations.

The TOEFL office advises institutions that the test may not provide a valid measure of the examinee's proficiency, even though the conditions were designed to minimize any adverse effects of the examinee's disability upon test performance. The TOEFL office continues to recommend that alternative methods of evaluating English proficiency be used for individuals who cannot take the test under standard conditions. Criteria such as past academic record (especially if English has been the language of instruction), recommendations from language teachers or others familiar with the applicant's English proficiency, and/or a personal interview or evaluation are suggested in lieu of TOEFL scores. Because the individual circumstances of nonstandard administrations vary so widely and the number of examinees tested under nonstandard conditions is still quite small, the TOEFL program cannot provide normative data for interpreting scores obtained in such administrations.

A statement that the scores were obtained under nonstandard conditions is printed on the official score report (and on the Examinee's Score Record) of an examinee for whom special arrangements were made. Each score recipient is also sent an explanatory notice emphasizing that there are no normative data for scores obtained under nonstandard testing conditions and, therefore, that such scores should be used within these parameters.



# USE OF TOEFL TEST SCORES.....

**The TOEFL test is a measure of general English proficiency. It is not a test of academic aptitude or of subject matter competence, nor is it a direct test of English speaking or writing ability.**

TOEFL test scores can assist in determining whether an applicant has attained sufficient proficiency in English to study at a college or university. However, even though an applicant may achieve a high TOEFL score, the student who is not academically prepared may not easily succeed in a given program of study. Therefore, determination of academic admissibility of nonnative English speakers is dependent upon numerous additional factors, such as previous academic record, other institution(s) attended, level and field of study, and motivation.

If a nonnative English speaker meets academic requirements, official TOEFL test scores may be used in making the following kinds of decisions:

- The applicant may begin academic work with no restrictions.
- The applicant may begin academic work with some restrictions on academic load and in combination with concurrent work in English language classes. (This implies that the institution can provide the appropriate English courses to complement the applicant's part-time academic schedule.)
- The applicant is declared eligible to begin an academic program within a stipulated period of time but is assigned to a full-time program of English study. (Normally, such a decision is made when an institution has its own intensive English-as-a-second-language program.)
- The applicant's official status will not be determined until he or she reaches a satisfactory level of English proficiency. (Such a decision will require that the applicant pursue full-time English training, at the same institution or elsewhere.)

All of the above decisions require the institution to judge whether the applicant has sufficient command of English to meet the demands of a regular or modified program of study. Such decisions should never be based on TOEFL scores alone; they should be based on all relevant information available.

## Who Should Take the TOEFL Test?

All international applicants who are nonnative speakers of English should provide evidence of their level of English proficiency prior to beginning academic work at an institution where English is the language of instruction. TOEFL scores are frequently required for the following categories of applicants:

- Individuals from countries in which English is one of the official languages, but not necessarily the first language of the majority of the population or the language of instruction at all levels of schooling. Such countries may include, but are not limited to, the British Commonwealth countries and US territories and possessions.
- Persons from countries where English is not the native language, even though there may be schools or universities in which English is the language of instruction.

Many institutions report that they frequently do not require TOEFL test scores of certain kinds of international applicants. These include:

- Nonnative speakers who hold degrees or diplomas from postsecondary institutions in English-speaking countries (e.g., the United States, Canada, England, Ireland, Australia, New Zealand), provided they have spent a specified minimum period of time in successful full-time study (generally two years) with English as the language of instruction.
- Transfer students from other institutions in the United States or Canada after favorable evaluation of previous academic course work and course load and length of time at the previous institution.
- Nonnative speakers who have taken the TOEFL test within the past two years and who have successfully pursued academic work in an English-speaking country for a specified minimum period of time (generally two years) with English as the language of instruction.

## Guidelines for Using TOEFL Test Scores

As part of its general responsibility for the tests it produces, the TOEFL program is concerned about the use of TOEFL test scores by recipient institutions. The program office makes every effort to ensure that institutions use TOEFL scores properly — for example, by providing this *Manual* to all institutions that are interested in using the scores and by regularly advising institutions of any program changes that may affect the interpretation of TOEFL test scores. The TOEFL office encourages individual institutions to request assistance of TOEFL professional staff relating to the proper use of scores.

An institution that uses TOEFL test scores should consider certain factors to evaluate an individual's performance on the test and to determine appropriate score requirements. The following guidelines are presented to assist institutions in arriving at reasonable decisions.

- **Base the evaluation of an applicant's readiness to begin academic work on all available relevant information, not solely on TOEFL test scores.**

The TOEFL test measures an individual's ability in several areas of English language proficiency. It is not designed to provide information about scholastic aptitude, motivation, language-learning aptitude, or cultural adaptability. The eligibility of a foreign applicant should be fully established on the basis of all relevant academic and other criteria, including sufficient proficiency in English to undertake the academic program at that institution.

- **Do not use rigid cut-off scores to evaluate an applicant's performance on the TOEFL test.**

Because test scores are not perfect measures of ability, the use of rigid cut-off scores should be avoided. The standard error of measurement should be understood and taken into consideration in making decisions about an individual's test performance or in establishing appropriate critical score ranges for the institution's academic demands (see "Reliabilities and the Standard Error of Measurement," page 29).

- **Consider TOEFL section scores as well as total scores.**

The total score on the multiple-choice TOEFL test is based on the scores of the three sections of the test. Although a number of applicants may achieve the same total score, they may have different section score profiles, which could significantly affect subsequent academic performance. For example, an applicant with a low score on the Listening Comprehension section but relatively high scores on the other sections might have greater initial difficulty in lecture courses.\* This information could be used in advising and placing applicants.

If an applicant's score on the Structure and Written Expression section is considerably lower than the scores on the other sections or if the applicant's score on the TWE test is low, it may be that the individual should take a reduced academic load or be placed in a course designed to improve composition skills and knowledge of English grammar. An applicant whose score on the Reading Comprehension section is much lower than the scores on the other two sections might be advised to take a reduced academic load or to postpone enrollment in courses that involve a significant amount of reading.\*

- **Consider the kinds and levels of English proficiency required in different fields and levels of study and the resources available at the institution for improving the English language skills of nonnative speakers.**

An applicant's field of study can affect the kind and level of language proficiency that are appropriate. Students pursuing studies in fields requiring high verbal ability (such as journalism) will need a greater command of English, particularly structure and written expression and writing, than will those in fields that are not so dependent upon reading and writing abilities. Many institutions require a higher range of TOEFL test scores for graduate applicants than for undergraduates.

Institutions offering courses in English for nonnative speakers of English can modify academic course loads to allow for additional concurrent language training, and thus may be able to consider applicants with a lower range of scores than can institutions that do not offer additional language training.

\* See page 39 for information about the Test of Spoken English and oral proficiency.

\* See page 39 for information about TSE.

■ **Consider TOEFL test scores to help interpret an applicant's performance on other standardized tests.**

International applicants are frequently required to take standardized admission tests in addition to the TOEFL test. In such cases, TOEFL scores may prove useful in interpreting the scores obtained on the other tests. For example, if an applicant's TOEFL scores are low and the scores on another test are also low (particularly one that is primarily a measure of aptitude or achievement in verbal areas), one can legitimately infer that the applicant's performance on the other test was impaired because of deficiencies in English. On the other hand, application records of students with high verbal aptitude scores but low TOEFL scores should be reviewed carefully. The scores may not be valid.

Interpreting the relationship between the TOEFL test and aptitude and achievement tests in verbal areas can be complex. Few of even the most qualified foreign applicants approach native proficiency in English. Factors such as cultural differences in educational programs may also affect performance on tests of verbal ability.

The TOEFL program has published four research reports that can assist in evaluating the effect of language proficiency on an applicant's performance on specific standardized tests.

*The Performance of Nonnative Speakers of English on TOEFL and Verbal Aptitude Tests* (Angelis, Swinton, and Cowell, 1979) gives comparative data about foreign student performance on TOEFL and either the GRE verbal or the SAT verbal and the Test of Standard Written English (TSWE). It provides interpretive information about how combined test results might best be evaluated by institutions that are considering foreign students. *The Relationship Between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language* (Powers, 1980) provides a similar comparison of performance on the GMAT and TOEFL tests. Finally, *Language Proficiency as a Moderator Variable in Testing Academic Aptitude* (Alderman, 1981) and *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL* (Wilson, 1982) contain information supplementing that provided in the other two studies. (See "Validity," page 34.)

■ **Do not use TOEFL test scores to predict academic performance.**

The TOEFL test is designed to be a measure of English language proficiency, not of academic aptitude. Although there may be some unintended overlap between language proficiency and academic aptitude, other tests have been designed to measure academic aptitude more precisely and are available for that purpose. Use of TOEFL scores to predict academic performance is inappropriate. Numerous predictive validity studies,\* using grade-point averages as criteria, have been conducted in the past. These studies have shown that correlations between TOEFL test scores and grade-point averages are often too low to be of any practical significance. Moreover, low correlations are to be expected when TOEFL scores are used properly. If an institution admits those international applicants who have demonstrated a high level of language competence, one would expect that English proficiency would no longer be highly correlated with academic success.

The English proficiency of an international applicant is not as stable a characteristic as verbal or mathematical aptitude. Proficiency in a language is subject to change over relatively short periods of time. If considerable time has passed between the date on which an applicant took the TOEFL test and the date on which he or she actually begins academic studies, there may be a greater impact on academic performance due to language loss than had been anticipated. On the other hand, a student who might be disadvantaged because of language problems during the first term of study might not be disadvantaged in subsequent terms.

■ **Assemble information about the validity of TOEFL test score requirements at the institution.**

The TOEFL program strongly encourages users to design and carry out institutional validity studies.\*\* Because it is important to establish appropriate standards of language proficiency, validity evidence may provide support for raising or lowering a particular standard as necessary. It may also be used to defend the standard should its legitimacy be challenged.

\* Chase and Stallings, 1966; Heil and Aleamoni, 1974; Homburg, 1979; Hwang and Dizney, 1970; Odunze, 1980; Schrader and Pitcher, 1970; Sharon, 1972.

\*\* A separate publication, "Guidelines for TOEFL Institutional Validity Studies," provides information to assist institutions in the planning of local validity studies. This publication is available without charge from the TOEFL program office upon request.

An important source of validity evidence for TOEFL scores is contained in information about subsequent performance by applicants who are admitted. Student scores may be compared to a variety of criterion measures, such as teacher (or adviser) ratings of English proficiency, graded written presentations, grades in ESL courses, and self-ratings of English proficiency. However, when evaluating a standard with data obtained solely from individuals who have met the standard (that is, only students who have been admitted), an interesting phenomenon may occur. If the current standard is set at a high level, so that only those with a high degree of language proficiency are admitted, there may be no relationship between the TOEFL scores and any of the criterion measures. Because there will be no important variability in English proficiency among the group members, variations in success on the criterion variable will likely be due to other causes, such as knowledge of the subject matter, academic aptitude, study skills, cultural adaptability, and financial security.

On the other hand, if the language proficiency standard is set at a low level, a large number of applicants selected with TOEFL scores may be unsuccessful in the academic program because of inadequate command of English, and there will be a relatively high correlation between their TOEFL scores and its criterion measure. Also, with a standard that is neither too high nor too low, the correlation between TOEFL scores and subsequent success will be only moderate. The magnitude of the correlation will depend on other factors as well. These factors may include variability in scores on the criterion measure and/or the reliability of the raters, if raters are used. Expectancy tables can be used to show the distribution of performance on the criterion variables for students with given TOEFL scores. Thus, it may be possible to depict the number or percentage of students at each score level who attain a certain language proficiency rating as assigned by an instructor, or who rate themselves as not being hampered by lack of English skills while pursuing college-level studies.

Another approach is to use a regression equation to support a score standard. Additional information about the setting and validation of test score standards is available in a manual by Livingston and Zieky (1982).

Several other methodological issues should be considered when conducting a standard-setting or validation study. Because language proficiency can change within a relatively short time, student performance on a criterion variable should be assessed during the first term of enrollment. However, if TOEFL scores are not obtained immediately prior to admission, gains or losses in language skills may reduce the relationship between the TOEFL test and the criterion.

Another issue that should be addressed is the relationship between subject matter or level of study and language proficiency. All subjects may not require the same level of language proficiency for the student to perform acceptably. For instance, the study of mathematics normally requires a lesser degree of English language proficiency than the study of philosophy. Similarly, first-year undergraduates who are required to take courses in a wide range of subjects may require a level of language proficiency different from that of graduate students who are enrolled in a specialized field of study.

Section scores may also be taken into consideration in the setting and validating of score standards. For fields that require a substantial amount of reading, the Reading Comprehension score may be particularly important. In fields that require little writing, the Structure and Written Expression or TWE score may be less important. Assessment of the relationship of section scores to the criterion variables can further refine the process of interpreting TOEFL scores.

To be useful, data about subsequent performance must be collected for relatively large numbers of students over an extended period of time. Institutions that have only a small number of foreign applicants each year or that have only recently begun to require TOEFL scores may not find it feasible to conduct the recommended studies. Such institutions might find it helpful to seek information and advice from colleges and universities that have had more extensive experience with the TOEFL test. The TOEFL office suggests that institutions evaluate their TOEFL requirements regularly to ensure that they are consistent with the institutions' own academic requirements and the language training resources they can provide nonnative speakers of English.

# STATISTICAL CHARACTERISTICS OF THE TEST.....

## Level of Difficulty

It is generally agreed by measurement specialists that the TOEFL test will provide the best measurement in the critical score range of about 450 to 600 when the test is of moderate difficulty. One indicator of test difficulty is provided by the percentage of correct items. The mean percent correct for the sections for the 13 different forms administered between July 1995 and June 1996 falls within 58.3 percent and 81.6 percent of the maximum possible score. For Listening Comprehension, the average percent correct ranges from 58.3 to 75.8 percent, with a mean percent correct of 67.3. For Structure and Written Expression, the values range from 63.7 to 81.1 percent, with a mean percent correct of 69.7. For Reading Comprehension, the values range from 59.1 to 78.7 percent, with a mean percent correct of 69.1.

Percent correct, as a measure of difficulty, depends both on the inherent difficulty of the test and on the ability level of the group of examinees that took the test. Both factors are of concern in determining whether the test is properly matched to the ability level of the examinees. However, for the scaled scores that are reported to examinees and institutions, the effect of the differences in difficulty level among the various forms of the test is removed, or adjusted for, by a statistical process called score equating. (See "Calculation of TOEFL Scores," page 22.)

## Test Equating

TOEFL test equating has two major purposes: (1) to adjust minor differences in difficulty among different TOEFL forms to ensure that examinees having equal levels of English proficiency will receive equivalent scaled scores and (2) to ensure that scores from different TOEFL forms are on a common scale so that they are comparable. To equate scores, the TOEFL program employs a "true score" equating method based on item response theory (Cook and Eignor, 1991; Hambleton and Swaminathan, 1985; Lord, 1980). All new TOEFL forms are equated to the TOEFL base form administered in July 1995. The equating procedure consists of establishing what scores on the new TOEFL form and on the TOEFL base form correspond to the same level of English proficiency. Scores for the new TOEFL form and the base form corresponding to the same level of English proficiency are considered to be equivalent. An

examinee's equated score, then, is the score on the July 1995 (or base) form for each section corresponding to the examinee's score for each section on the current form. The examinee's converted, or reported, scores are obtained by applying the nonlinear conversion table originally obtained for each section on the base form to the examinee's equated section scores.

## Adequacy of Time Allowed

Although no single statistic has been widely accepted as a measure of the adequacy of time allowed for a separately timed section, two rules of thumb are used at ETS: (1) 80 percent of the group ought to be able to finish almost every question in each section, and (2) 75 percent of the questions in a section ought to be completed by almost all of the group. The Listening Comprehension section of the TOEFL test is paced by a recording; thus, every question is presented to every examinee and the criteria for speededness do not apply.

For Sections 2 and 3 of the 13 forms administered between July 1995 and June 1996, at least 94 percent of each group of examinees were able to complete **all** the questions in each section, and the three-quarter point in the sections was reached by 99.1 to 100.0 percent. Thus, one may reasonably conclude that, by these criteria, speed is not an important factor in TOEFL scores.

## Reliabilities and the Standard Error of Measurement

The TOEFL test is an accurate and dependable measure of proficiency in English as a foreign language. However, no test score is entirely without measurement error. This does not mean that someone has made a mistake in constructing or scoring the test. It means only that examinees' scores are not perfectly consistent, due to a number of factors. The extent to which test scores are free from errors in the measurement process is known as reliability. Reliability describes the tendency of individual examinees' scores to have the same relative positions in the group, no matter which form of the test the examinees take. Test reliability can be estimated by a variety of different statistical procedures. The two most commonly used statistical indices are the reliability coefficient and the standard error of measurement.

The term “reliability coefficient” is generic, reflecting the fact that a variety of coefficients exist because errors in the measurement process can arise from a variety of sources. For example, sources of error can be found from variations in the sample of tasks required by the testing instrument, or in the way that examinees respond during the course of a single test administration. Reliability coefficients that quantify these sources are known as measures of internal consistency, and they refer to the reliability of a measurement instrument at a single point in time. It is also possible to obtain reliability coefficients that take into account additional sources of error, such as changes in the performance of examinees from day to day and/or variations due to different test forms. Typically, these latter measures of reliability are difficult to obtain because they require that a group of examinees be retested with the same or another test form on another occasion.

In numerical value, reliability coefficients are always between .00 and .99, and generally between .60 and .95. The closer the value of the reliability coefficient to the upper limit, the greater the freedom of the test from error in measurement. Table 2 gives average internal consistency reliabilities of the scaled scores for each of the three multiple-choice sections and for the total test based on TOEFL test forms administered between July 1995 and June 1996. For a somewhat different view of reliability that looks at local dependence in TOEFL reading comprehension items and some listening comprehension items, see Wainer and Lukhele (in press).

**Table 2. Reliabilities and Standard Errors of Measurement (SEM)\***

Section	Reliability	SEM
1. Listening Comprehension	.90	2.0
2. Structure and Written Expression	.86	2.7
3. Reading Comprehension	.89	2.4
Total Score	.95	13.9

\* The medians of forms administered between July 1995 and June 1996. Based only on examinees tested in the United States and Canada.

The standard error of measurement (SEM) is an estimate of the probable extent of the error inherent in a test score due to the imprecision of the measurement process. As an example, suppose that a number of persons, all possessing the same degree of English language proficiency, were to take the same TOEFL test form. Despite their equal proficiency, these persons would not all get the same TOEFL score. A few would get much higher scores than the rest, a few much lower; however, most would obtain TOEFL scores that were close to the scores that represented their actual proficiency. The variation in scores could be attributable to differences in motivation, attentiveness, the particular items on the TOEFL test, and other factors such as those mentioned above. The standard error of measurement is an index of how much the scores of examinees having the same actual proficiency can be expected to vary.

Interpretation of the standard error of measurement is based on concepts in statistical theory and is applied with the understanding that errors of measurement can be expected to follow a particular sampling distribution. In the above example, the score that each of the persons with the same proficiency would achieve on the test if there were no errors of measurement is called the “true score.” The observed scores that these persons could be expected to actually receive are assumed to be normally distributed about this true score. That is, the true score is assumed to be the expected value (i.e., the mean) of the distribution of observed scores. The standard deviation of this distribution is the standard error of measurement.

Note that the standard error of measurement defined this way is actually the conditional standard error of measurement (CSEM) given a particular true score. That is, the standard deviation of the distribution for the observed scores corresponding to a particular true score is the CSEM given that true score. Typically the CSEMs for particular true scores peak in the middle of the score range and decrease as the true scores increase. This is because for higher true scores the corresponding observed scores have a smaller range of possible variation. As evidenced by TOEFL data from July 1995 to June 1996, for Section 2 the CSEM for a scaled score of 45 is 3.16, much bigger than 1.94, the CSEM for a scaled score of 60.

Once the CSEMs are defined and calculated, the SEM for a section scaled score can be computed as the weighted average of the CSEMs, with the weights based on the scaled score distribution.

When computing the CSEMs and then the SEM, because the true item and ability parameters are unknown, estimated item and ability parameters are used. The resulting CSEMs and SEM will likely differ somewhat from their actual true values (they are not necessarily just underestimates of the true values). However, the effect of estimation error on the reported values of the CSEMs and SEM is likely to be small for two reasons: (1) the effect of estimation error of item and ability parameters on the CSEMs (and SEM) is through its effect on the item characteristic curves, and in general the item characteristic curves are robust to modest changes in item and ability parameters; and (2) the CSEMs (and the SEM) are related to the item characteristic curves, through a summation process, and in the summation process, each item contributes only a small amount to the CSEMs. Unless estimation error causes the item contributions to all be inaccurate in the same direction (which is very unlikely), the effect will be canceled out through the summation process.

In most instances the SEM is treated as an average value and applied to all scores in the same way. It can be expressed in the same units as the reported score, which makes it quite useful in interpreting the scores of individuals. Table 2 shows that the SEM for Section 1 is 2.0 points; for Section 2, 2.7 points; for Section 3, 2.4 points; and for the total score, 13.9 points. There is, of course, no way of knowing just how much a particular person's actual proficiency may have been under- or overestimated from a single administration. However, the SEM can be used to provide score bands or confidence bands around observed scores to arrive at estimates of true scores of persons in a particular reference group.

Because the section and total score reliabilities (given in Table 2) are quite high for TOEFL, if the observed scores of examinees are not extreme it is fairly likely that their true scores lie within one SEM of their observed scores. For example, from the data in Table 2, we can be fairly confident that for Section 1, the examinees' true scores lie within 2 points of their observed scores. For the total score, it is fairly likely

that the examinees have true scores within 13.9 points of their reported scores. Alternatively, suppose a given examinee had a reported score of 50 on Section 3 of the test. We could then say that it is likely this person's true score was between 48 and 52. More precise methods for calculating score bands around observed scores to estimate true scores are available (see, for example, Harvill, 1991).

In comparing total scores for two examinees, the standard errors of measurement need to be taken into account. The standard error of the difference between TOEFL scores for two examinees is  $\sqrt{2}$  (or 1.414) times the standard error of measurement presented in Table 2 and takes into account the contribution of two error sources in the different scores. One should not conclude that one score represents a significantly higher level of proficiency in English than another score unless there is a difference of at least 39 points between them. In comparing section scores for two persons, the difference should be at least 6 points for Section 1, at least 8 points for Section 2, and at least 7 points for Section 3. (For additional information on the standard errors of score differences, see Anastasi, 1968, and Magnusson, 1967.)

Consideration of the standard error of measurement underscores the fact that no test score is entirely without measurement error, and that cut-off scores should not be used in a completely rigid fashion in evaluating an applicant's performance on the TOEFL test. Some justification for this position follows.

TOEFL scores are used by many different undergraduate and graduate programs in conjunction with candidates' other profiles to make admissions decisions. Each program has its own requirement as to candidates' English proficiency levels. Some may require higher spoken communication skills and others may require higher writing skills, demanding differential consideration of the section scores. At times TOEFL scores are used to prescreen candidates, and factors such as applicant pool as well as projected classroom size come into play. All these circumstances make setting a universal cut-off score impossible as well as unnecessary. However, many programs do have their own cut-off scores set to reflect perhaps the basic level of candidate English proficiency to survive their programs, as well as simply to prescreen and



reduce the prospective applicant pool. Keep in mind, however, that it is extremely difficult to defend any particular cut-off score. The process of setting cut-off scores has been identified by researchers as an example of a judgment or decision-making task (JDM), and as Jaeger (1994) noted, “responses to JDM tasks, including standard-setting tasks (cut-off scores being the outcome) are...responses to problem statements that are replete with uncertainty and less than complete information.” Also as clearly articulated by Brennan (1994), “standard setting is a difficult activity, involving many *a priori* decisions and many assumptions.”

Another problem with cut-off scores is that they are often perceived as arbitrary. As noted by van de Linden (1994, page 100):

The feelings of arbitrariness...stem from the fact that although cut scores have an “all or none” character, their exact location can never be defended sufficiently. Examinees with achievement just below a cut score differ only slightly from those with achievements immediately above the score. However, the personal consequences of this small difference may be tremendous, and it should be no surprise that these examinees can be seen as the victims of arbitrariness in the standard-setting procedure.

Still another problem with the setting of cut-off scores is that the particular method used to set the standard will clearly affect the results, i.e., different procedures will provide different cut-off scores. Standards are constructed rather than discovered, and there are no “true” standards. As Jaeger (1994) pointed out, “a right answer does not exist, except perhaps in the minds of those providing judgments.” All these factors support not using a cut-off score in a completely rigid fashion in evaluating an applicant’s performance on TOEFL. (For additional guidelines for using TOEFL test scores, see pages 25-28.)

## Reliability of Gain Scores

Some users of the TOEFL test are interested in the relationship between TOEFL scores that are obtained over time by the same examinees. For example, an English language instructor may be interested in the gains in TOEFL scores obtained by students in an intensive English language program. Typically, the available data will consist of differences calculated by subtracting TOEFL scores obtained at the completion of the program from those obtained at the beginning of the program. In interpreting these gain scores, we must inquire how reliable our estimates of these differences are, taking into account the characteristics of each of the two tests administered.

Unfortunately, it is a fact that the assessment of the difference between two test scores usually has substantially lower reliability than the reliabilities of the two tests taken separately. This is due to two factors. First, the errors of measurement that occur in each of the tests are accumulated in the difference score. Second, the common aspects of language proficiency that are measured on the two occasions are canceled out in the difference score. This latter factor means that, other things being equal, the reliability of the difference scores decreases as the correlation between pretest and posttest increases. This is because more of what is common between the two tests is canceled out of the difference score, and more of what is left over is made up of the accumulated errors of measurement in each of the two tests. As a numerical example, if the reliability of both the pretest and the posttest is about .90 and if the standard deviations of the scores are assumed to be equal, the reliability of the gain scores decreases from .80 to .50 as the correlation between pretest and posttest increases from .50 to .80. If the correlation between pretest and posttest is as high as the reliabilities of the two tests, the reliability of the gain scores is zero. For further discussion on the limitations in interpreting difference scores, see Linn and Slinde (1977), and Thorndike and Hagan (1977, pages 98-100).



The attribution of gain scores in a local setting requires caution, because gains may reflect increased language proficiency, a practice effect, and/or a statistical phenomenon called “regression toward the mean” (which essentially means that, upon repeated testing, high scorers tend to score lower and low scorers tend to score higher). Swinton (1983) analyzed data from a group of students at San Francisco State University that indicated that TOEFL score gains decrease as a function of proficiency level at the time of initial testing. For this group, student scores were obtained at the start of an intensive English language program and at its completion 13 weeks later. Students whose initial scores were in the 353-400 range showed an average gain of 61 points; students whose initial scores were in the 453-500 range showed an average gain of 42 points.

As part of this study, an attempt was made to remove the effects of practice and regression toward the mean by administering another form of the TOEFL test one week after the pretest. Initial scores in the 353-400 range increased about 20 points on the retest, and initial scores in the 453-500 range improved about 17 points on the retest. The greater part of these gains can be attributed to practice and regression toward the mean, although a small part may reflect the effect of one week of instruction.

Subtracting the retest gain (20 points) from the posttest gain (61 points), it was possible to determine that, within this sample, students with initial scores in the 353-400 range showed a real gain on the TOEFL test of 41 points during 13 weeks of instruction. Similarly, students in the 453-500 initial score range showed a 25-point gain in real language proficiency after adjusting for the effects of practice and regression. Thus, the lower the initial score, the greater will be the probable gain over a fixed period of instruction. Other factors, such as the nature of the instructional program, will affect gain scores also.

The TOEFL program has published a manual (Swinton, 1983) that describes a methodology suitable for conducting local studies of gain scores. University-affiliated and private English language institutes may wish to conduct gain score studies with their own students to determine the amount of time that is ordinarily required to improve from one score level to another.

## Intercorrelations Among Scores

The three multiple-choice sections of the TOEFL test are designed to measure different skills within the general domain of English proficiency. It is commonly recognized that these skills are interrelated; persons who are highly proficient in one area tend to be proficient in the other areas as well. If this relationship were perfect, there would be no need to report scores for each section. The scores would represent the same information repeated several times, rather than different aspects of language proficiency.

Table 3 gives the correlation coefficients measuring the extent of the relationships among the three sections and with the total test score. A correlation coefficient of 1.0 would indicate a perfect relationship between the two scores, and 0.0 would indicate a total lack of relationship. The table shows average correlations over the forms administered between July 1995 and June 1996. Correlations between the section scores and the total score are spuriously high because the section scores are included in the total. The observed correlations, ranging from .68 to .79, indicate that there is a fairly strong relationship among the skills tested by the three multiple-choice sections of the test, but that the section scores provide some unique information.

**Table 3. Intercorrelations Among the Scores\***

Section	1	2	3	Total
1. Listening Comprehension	—	.68	.69	.86
2. Structure and Written Expression	.68	—	.79	.92
3. Reading Comprehension	.69	.79	—	.92
Total Score	.86	.92	.92	—

\* The medians of correlation coefficients for forms administered between July 1995 and June 1996. Based only on examinees tested in the United States and Canada.

## Validity

In addition to evidence of reliability, there should be an indication that a test is valid — that it actually measures what it is intended to measure. For example, a test of basic mathematical skills that yielded very consistent scores would be considered reliable. But if those scores showed little relationship to students' performance in basic mathematics courses, the validity of the test would be questionable. This would be particularly true if the scores showed a stronger relationship to the students' performance in less relevant areas, such as language or social studies. The question of validity of the TOEFL test relates to how well it measures a person's proficiency in English as a second or foreign language.

Establishing the validity of a test is admittedly one of the most difficult tasks facing those who design the test. For this reason, validity is usually confirmed by analyzing the test from a number of perspectives.

Although researchers have stated definitions for many different types of validity, it is generally recognized that validity refers to the usefulness of inferences made from test scores (APA, 1985; Messick, 1987). To support inferences, validation should include several types of evidence, e.g., content-related, criterion-related, and construct-related. The nature of the evidence should depend on the specific inference or use of the test.

To establish content-related evidence, one must demonstrate that the content exhibited and behavior elicited on a test constitute an adequate sample of the content and behaviors of the subject or field tested. Criterion-related evidence of validity applies when one wishes to draw a relationship between a score on the test under consideration and a score on some other variable, called a criterion. Construct-related validity evidence should support the integrity of the intended constructs or behavioral domains as measured on the test. For a test that reports a total score and three section scores, such as TOEFL, research should provide evidence of the integrity of constructs and the validity of inferences associated with every score reported. Of the three kinds of validity evidence, content-related evidence is established by examining the content of the test, whereas criterion-related and construct-related evidence frequently involve judgments based on statistical relationships.

## Content Validity

Content-related evidence for the TOEFL test is a major concern of the TOEFL Committee of Examiners (see page 8), which has developed a comprehensive list of specifications for items appearing in the different sections of the test. The specifications identify the aspects of English communication, ability, and proficiency that are to be tested and describe appropriate techniques for testing them. The specifications are continually reviewed and revised as appropriate to ensure that the test reflects both current English usage and current theory as to the nature of second language proficiency.

A TOEFL research study by Duran, Canale, Penfield, Stansfield, and Liskin-Gasparro (1985) analyzed one form of the TOEFL test from several different frameworks related to contemporary ideas about aspects of communicative competence. These frameworks take into account the grammatical, sociolinguistic, and discourse competencies required to answer TOEFL items correctly. Although the competencies and the degree to which the TOEFL test measures them vary considerably across sections, the results indicate that successful performance on the test requires a wide range of competencies.

Information regarding the perceptions of college faculty of the validity of the Listening Comprehension section is available in *A Survey of Academic Demands Related to Listening Skills* (Powers, 1985). Powers found that the kinds of listening comprehension questions used in the TOEFL test were rated (by faculty) as being among the most appropriate of those considered.

Bachman, Kunnan, Vanniarajan, and Lynch (1988) suggest that the reading passages in Section 3 tend to be entirely academic in focus. This is consistent with the intended use of the test as a measure of proficiency in English for academic purposes.

Although American cultural content is present in the test, care has been taken to ensure that knowledge of such content is not required to succeed in responding to any of the items. Angoff (1989), in a study using one form of the TOEFL test with more than 20,000 examinees tested abroad and more than 5,000 examinees tested in the United States, established that there was no detected cultural advantage for examinees who had resided more than one year in the United States.

In 1984, the TOEFL program held an invitational conference to discuss the content validity of the test. The conference brought together some two dozen specialists in the testing of English as a second language. The papers presented at the conference are available in *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference* (Stansfield, 1986). These papers provide additional information about the language tasks that appear on the TOEFL test and are an important reference for an understanding of the content validity of the test. Subsequent changes in the test, designed to make it more reflective of communicative competence, are enumerated on pages 92 and 93 of the proceedings.

### Criterion-Related Validity

Some of the earliest and most basic TOEFL research attempted to match performance on the test with other indicators of English language proficiency, thus providing criterion-related evidence of TOEFL's validity. In some cases these indicators were tests themselves.

A study conducted by Maxwell (1965) at the Berkeley campus of the University of California found an .87 correlation between total scores on the TOEFL test and the English proficiency test used for the placement of foreign students at that campus. This correlation was based on a total sample of 238 students (202 men and 36 women, 191 graduates and 47 undergraduates) enrolled at the university during the fall of 1964. Upshur (1966) conducted a study to determine the correlation between TOEFL and the Michigan Test of English Language Proficiency. This was based on a total group of 100 students enrolled at San Francisco State College (N = 50), Indiana University (N = 38), and Park College (N = 12) and yielded a correlation of .89. Other studies comparing TOEFL and Michigan Test scores have been done by Pack (1972) and Gershman (1977). In 1966 a study was carried out at the American Language Institute (ALI) at Georgetown University comparing scores on TOEFL with scores on the ALI test developed at Georgetown. The correlation of the two tests for 104 students was .79.

In addition to comparing TOEFL with other tests, some of these studies included investigations of how performance on TOEFL related to teacher ratings. In the ALI Georgetown study the correlation between TOEFL and these ratings for 115 students was .73. Four other institutions reported similar correlations. Table 4 gives the data from these studies. At each of the institutions (designated by code letters in the table) the students were ranked in four, five, or six categories based on their proficiency in English as determined by university tests or other judgments of their ability to pursue regular academic courses (American Language Institute, 1966).

**Table 4. Correlations of Total TOEFL Scores with University Ratings**

University	Number of Students	Correlations with Ratings
A	215	.78
B	91	.87
C	45	.76
D	279	.79

In a study conducted on the five-section version of the test used prior to 1976, Pike (1979) investigated the relationship of the TOEFL test and its subsections to a number of alternate criterion measures, including writing samples, cloze tests, oral interviews, and sentence-combining exercises. In general, the results confirmed a close relationship between the five sections of the TOEFL test and the English skills they were intended to measure. Among the most significant findings of this study were the correlations between TOEFL subscores and two nonobjective measures: oral interviews and writing samples (essays).

Table 5 gives the correlation coefficients for the three language groups participating in the study. Moreover, the figures are shown for both the total interview ratings and the grammar and vocabulary subscores; the essay ratings are listed according to two different scoring schemes — one focusing on essay content and one on essay form. The strong correlations and common variances found in Pike’s study between some of the sections of the TOEFL test led to the combining and revising of those sections to form the current three-part version of the test.

**Table 5. Correlations of TOEFL Subscores with Interview and Essay Ratings**

		N	Interview			Essay	
			Gram.	Vocab.	Total	Content	Form
Listening Comprehension	Peru	95	.84	.84	.84	.83	.91
	Chile	143	.76	.75	.78	.76	.83
	Japan	192	.84	.83	.82	.59	.72
English Structure	Peru	95	.86	.87	.87	.86	.92
	Chile	143	.88	.87	.87	.88	.98
	Japan	192	.70	.69	.71	.55	.81
Vocabulary	Peru	95	.82	.83	.82	.80	.84
	Chile	143	.77	.77	.75	.74	.83
	Japan	192	.55	.62	.59	.45	.66
Reading Comprehension	Peru	95	.88	.87	.87	.84	.85
	Chile	143	.74	.76	.75	.67	.82
	Japan	192	.62	.62	.62	.61	.73
Writing Ability	Peru	95	.86	.85	.86	.85	.93
	Chile	143	.79	.78	.75	.77	.88
	Japan	192	.59	.62	.60	.64	.73

Further evidence for the criterion-related validity of the TOEFL, TSE, and TWE tests was provided by Henning and Cascallar (1992) in a study relating performance on these examinations to independent ratings of oral and written communicative language ability over a variety of controlled academic communicative functions.

### Construct Validity

In early attempts to obtain construct-related evidence of validity for the TOEFL test, two studies were conducted comparing the performance of native and nonnative speakers of English on the test. Angoff and Sharon (1970) found that the mean TOEFL scores of native speakers in the United States were much higher than those of foreign students who had taken the same test. Evidence that the test was quite easy for the American students is found in the observations that their mean scores were not only high but homogeneously high relative to those of the foreign students; that their score distributions were highly negatively skewed; and that a high proportion of them earned maximum or near-maximum scores on the test.

A more detailed study of native speaker performance on the TOEFL test was conducted by Clark (1977). Once again, performance on the test as a whole proved similar to that of the native speakers included in the Angoff and Sharon study. The mean raw score for the native speakers, who took two different forms of the TOEFL test, was 134 (out of 150). This compared to mean scores of 88 and 89 for the nonnative speakers who had originally taken the same forms. However, additional analysis showed that the native speakers did not perform equally well on all three sections of the test.

Such information is useful for test development because it provides guidelines on which to base evaluations of questions at the review stage. The information from these comparisons of native and nonnative speakers of English also provides evidence of the construct validity of the TOEFL test as a measure of English language proficiency.

More recent evidence for the construct validity of the TOEFL test is available in a series of studies investigating the factor structure and dimensionality of the test (Boldt, 1988; Hale, Rock, and Jirele, 1989; Oltman, Stricker, and Barrows, 1988). Evidence for the validity of constructs measured by current and prospective listening and vocabulary item types is presented in Henning (1991a, 1991b). A number of other construct validity studies are available in the TOEFL Research Report Series (see pages 43-45), the most recent of which bear on some construct validity evidence for the reading and listening portions of TOEFL (Freedle and Kostin, 1993, 1996; Nissan, DeVincenzi, and Tang, 1996; and Schedl, Thomas, and Way, 1995).

Other evidence of TOEFL's validity is presented in studies that have focused on the relationship of the TOEFL test to some widely used aptitude tests. The findings of these studies contribute to the construct-related validity evidence by showing the extent to which the test has integrity as a measure of proficiency in English as a foreign language. One of these studies (Angelis, Swinton, and Cowell, 1979) compared the performance of nonnative speakers of English on the TOEFL test with their performance on the verbal portions of the GRE Aptitude (now General) Test (graduate-level students) or both the SAT and the Test of Standard Written English (undergraduates). As indicated in Table 6, the GRE verbal performance of the nonnative speakers was much lower and less reliable than the performance of the native speakers. Similar results were reported for undergraduates on the SAT verbal and the TSWE (Table 7).

**Table 6. TOEFL/GRE Verbal Score Comparisons**

	Mean	S.D.	Rel.	S.E.M.
TOEFL	523	69	.95	15
(Nonnatives) (N = 186) GRE-V	274	67	.78	30
Native Speakers (N = 1,495) GRE-V	514	128	.94	32

**Table 7. TOEFL/SAT and TSWE Score Comparisons**

	Mean	S.D.	Rel.	S.E.M.
TOEFL	502	63	.94	16
(Nonnatives) (N = 210) SAT-V	269	67	.77	33
Native Speakers (N = 1,765) SAT-V	425	106	.91	32
(Nonnatives) (N = 210) TSWE	28	8.8	.84	4
Native Speakers (N = 1,765) TSWE	42.35	11.09	.89	3.7

Wilson (1982) conducted a similar study of all GRE, TOEFL, and GMAT examinees during a two-year period extending from 1977 to 1979. These results, depicted in Table 8, combined with those obtained in the earlier study by Angelis, Swinton, and Cowell (1979), warrant an important conclusion for admissions officers: verbal aptitude test scores of nonnative examinees are significantly lower on average than the scores earned by native English speakers. On the other hand, quantitative aptitude scores are not greatly affected by a lack of language proficiency. Further, analyses of each study show that only when TOEFL scores reach approximately the 625 level do verbal aptitude test scores of foreign candidates reach the level normally obtained by native English speakers.

**Table 8. TOEFL, GRE, and GMAT Score Comparisons, 1977-79**

	GRE Sample			GMAT Sample		
	All	Foreign ESL	TOEFL	All	Foreign ESL	TOEFL
N	831,650	2,442	2,442	563,849	3,918	3,918
Verbal Mean	479	345	NA	26	15.7	NA
SD	129	95	NA	9	7.7	NA
Quantitative Mean	518	606	NA	27	29	NA
SD	135	136	NA	8	9.2	NA
Analytical Mean	496	400	NA	NA	NA	NA
SD	120	114	NA	NA	NA	NA
Total Mean	NA	NA	552	462	389.8	541.8
SD	NA	NA	61	105	97.5	71.7

To provide guidelines for those who may be evaluating applicants presenting scores from more than one of the above tests, Angelis, Swinton, and Cowell (1979) conducted special analyses. Results indicated that, for graduate-level applicants, 475 on the TOEFL test is a critical decision point for interpretations of GRE verbal scores. Applicants above that level tend to have GRE verbal scores that, although lower than scores for native speakers, fall within an interpretable range of verbal ability for students with homogeneous TOEFL scores. Those below the 475 TOEFL level tend to have such low GRE verbal scores that such interpretations cannot easily be made. At the undergraduate level, 435 on TOEFL is a key decision point. SAT verbal scores for applicants below that level are not likely to be informative. Similarly, Powers (1980) found that a TOEFL score of 450 is required before GMAT verbal scores begin to discriminate effectively among examinees.

These results suggest that, when TOEFL scores enter the range normally considered for admissions decisions, it is also possible to draw valid inferences from scores on aptitude tests.

As noted earlier, interpreting the relationship between language proficiency and aptitude and achievement test scores in verbal areas can be complex. Few of even the most qualified international applicants approach native proficiency in English. Thus, verbal aptitude scores of nonnative English speakers are likely to be depressed somewhat even when TOEFL test scores are high. Only when TOEFL scores are at an average native speaker level (approximately 625 or above) does the distribution of scores on a verbal aptitude test become similar to the distribution obtained by native English speakers. Cultural factors and cross-national differences in educational programs may also affect performance on tests of verbal ability.

As noted above, the TOEFL program has published three research reports that can assist in evaluating the effect of language proficiency on an applicant's performance on specific standardized tests. *The Performance of Nonnative Speakers of English on TOEFL and Verbal Aptitude Tests* (Angelis, Swinton, and Cowell, 1979) gives comparative data about the performance of a group of foreign students on the TOEFL test and either the GRE verbal or the SAT verbal and the TSWE. *The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language* (Powers, 1980) compares performance on TOEFL and GMAT. Additional information and comparisons are available in *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL* (Wilson, 1982).

TOEFL is currently a three-section test. Support for the three-section format is provided by the pattern of correlations between each of the TOEFL sections and other tests (Angelis, Swinton, and Cowell, 1979). The GRE verbal score correlates highest with the Reading Comprehension section of TOEFL (.623). The same section correlates highest (.681) with the SAT verbal score. This is to be expected since both verbal aptitude tests rely heavily on reading and vocabulary. For the College Board's TSWE, the highest correlation (.708) is with Section 2 of TOEFL, Structure and Written Expression. Again, this is to be expected because the TSWE uses knowledge of grammar and related linguistic elements as indicators of writing ability. In all three cases, the lowest correlations are those with TOEFL Section 1, Listening Comprehension. Because none of the other tests includes items that attempt to measure ability to understand spoken English, this again is to be expected.

In another study cited earlier, comparing performance of nonnative speakers of English on TOEFL and the Graduate Management Admission Test, Powers (1980) reported the same pattern of correlations. As indicated in Table 9, the highest GMAT verbal-TOEFL correlation is that for the Vocabulary and Reading Comprehension section. Correlations for Section 2 are slightly lower and those for Section 1 (listening) are the lowest. The fact that the correlations for the quantitative section of the GMAT are the lowest of all (ranging from .29 to .39) provides support for the discriminating power of the TOEFL test as a measure of verbal skills in contrast to quantitative skills.

**Table 9. Correlations Between GMAT and TOEFL Scores\***

GMAT Score	TOEFL Scores			
	Listening Comprehension	Structure and Written Expression	Reading Comprehension	Total
GMAT Verbal	.58	.66	.69	.71
GMAT Quantitative	.29	.37	.39	.39
GMAT Total	.52	.61	.64	.66

\*Based on 5,781 examinees with TOEFL and GMAT scores.



# OTHER TOEFL PROGRAMS AND SERVICES

## TWE Test (Test of Written English)

This 30-minute essay test provides the examinee with an opportunity to perform writing tasks similar to those required of students in North American universities. This includes the ability to generate and organize ideas on paper, to support those ideas with examples or evidence, and to use the conventions of standard written English.

The examinee is given one topic on which to write. As with other TOEFL test items, the TWE essay questions are developed by specialists in English or ESL, and each essay question is field-tested and reviewed by a committee of composition specialists, the TWE Committee. A pretested topic will be approved for use in the TWE test only if it elicits a range of responses at a variety of proficiency levels, does not appear to unfairly advantage or disadvantage any examinee or group of examinees, and does not require special subject matter knowledge. The essay questions are also reviewed for racial and cultural bias and content appropriateness according to established ETS sensitivity review procedures.

After a test administration, each TWE essay is read by two trained and qualified raters, who assign scores based on a six-point, criterion-referenced scoring guide. Neither reader knows the score assigned by the other. In the case of a discrepancy of more than one point, a third reader scores the essay.

The Test of Written English score is not incorporated into the total TOEFL score. Instead, a separate TWE score is reported on the TOEFL score report. Score recipients receive a copy of the *TWE Scoring Guide*, which describes the proficiency levels associated with the six holistic score points. Sample essays at the six score levels are published in the *TOEFL Test of Written English Guide*.

TWE test results can assist institutions in evaluating the academic writing proficiency of their ESL and EFL students and in placing these students in appropriate writing courses.

## TSE Test (Test of Spoken English)

The Test of Spoken English was developed by ETS under the direction of the TOEFL Policy Council and TSE Committee to provide a reliable measure of proficiency in spoken English. Because the TSE test is a test of general oral language ability, it is appropriate for examinees regardless of native language, type of educational training, or field of employment.

The TSE test has broad applicability in that performance on the test indicates how oral communicative language ability might affect the examinee's ability to communicate successfully in an academic or professional environment. TSE scores are used at many North American institutions of higher education in the selection of international teaching assistants, sometimes called ITAs. The scores are also used for selection and certification purposes in the health professions, such as medicine, nursing, pharmacy, and veterinary medicine.

The Test of Spoken English is administered 12 times a year on the same dates as the TOEFL test. The test takes approximately 20 minutes and can be administered to individuals with cassette tape recorders or to a group using a language laboratory.

The TSE test requires examinees to demonstrate their ability to communicate orally in English by responding orally under timed conditions to a variety of printed and aural stimuli that are designed to elicit a variety of responses. All examinee responses are recorded on tape.

The test consists of 12 items, each of which requires examinees to perform a particular speech act. Examples of these speech activities, also called "language functions," include narrating, recommending, persuading, and giving and supporting an opinion. The time allotted for each answer ranges from 30 to 90 seconds and is written in parentheses after each question.

TSE answer tapes are rated by trained specialists in the field of English or English as a second language. (The rating scale is different from the rating scale used for the TSE test prior to July 1995.) Raters assign a score level for the response to each item by using a set of descriptors that describe performance at various levels of proficiency based on communicative language features. Examinee scores are produced from the average of these two ratings and are reported on a score scale of 20 to 60. Official score reports are sent to institutions designated by the examinees.

The TSE rating scale and a sample score report are printed in the *TSE Manual for Score Users*. A TSE sample response tape is also available to provide score users with sample examinee responses at the levels of communicative effectiveness represented by particular TSE scores.



## **SPEAK® Kit (Speaking Proficiency English Assessment Kit)**

The Speaking Proficiency English Assessment Kit (SPEAK) was developed by the TOEFL program to provide institutions with a valid and reliable instrument for assessing the spoken English of nonnative speakers.

SPEAK consists of several components including a Rater Training Kit, test materials, and an Examinee Practice Set, each of which is purchased separately. The Rater Training Kit includes the materials necessary for training individuals to rate examinees' recorded responses to the test. The training materials consist of a *Rater Training Guide*, sample response cassette, training cassettes, testing cassettes, practice rating sheet, and scoring key. Raters determine whether they have mastered the necessary rating skills by comparing the ratings they assign to the rater-testing cassettes with the correct ratings provided in the *Guide*.

SPEAK test results can be used to evaluate the speaking proficiency of applicants for teaching assistantships who are not native speakers of English, to measure improvement in speaking proficiency over a period of time, or to identify teaching assistants and others who may need additional instruction in English.

Two SPEAK test forms (Test Forms A and B) are available to purchasers of the SPEAK kit. The SPEAK testing materials, which allow repeated test administration at any convenient location, consist of 30 reusable examinee test books, a cassette tape for actual administration of the test, a scoring key, and a pad containing 100 rating sheets.

It is important to be aware that SPEAK is designed for internal or local use only. SPEAK tests are available for direct purchase by university-affiliated language institutes, institutional or agency testing offices, and other organizations or offices serving educational programs.

## **SLEP® Test (Secondary Level English Proficiency Test)**

The Secondary Level English Proficiency test is designed for students entering grades 7 through 12 who are nonnative speakers of English. The test is a measure of proficiency in two primary areas: understanding spoken English and understanding written

English. The SLEP test is based on the assumption that language proficiency is a critical factor in determining the degree to which students can benefit from instruction. It is not an aptitude test or a measure of academic achievement; it is a measure of English language ability. The results of the test can be very helpful in making placement decisions related to assignment to ESL classes, placement in a mainstream English-medium program, or exit from an ESL program. Because the SLEP scale is sensitive to small gains in language skills, the test can be useful for program evaluation purposes.

There are three different forms of the SLEP test, all developed to the same test specifications, equated, and norm referenced. Each test form contains 150 multiple-choice questions of eight different item types and is divided into two sections, Listening Comprehension and Reading Comprehension. The questions in the first section of the test use taped samples of spoken English to test listening comprehension and do not rely heavily on written materials. The questions in the second section measure vocabulary, grammar, and overall reading comprehension and are based on written and visual materials. Answer sheets are easily scored, and technical data for interpreting test results are provided in the *SLEP Test Manual*.

SLEP testing materials are available for direct purchase. The basic package of testing materials for each form contains 20 SLEP test books, 100 two-ply answer sheets, a copy of the *SLEP Test Manual*, and a cassette recording of the listening comprehension questions. Each item in the basic package may also be purchased separately.

## **Fee Voucher Service for TOEFL and TSE Score Users**

The TOEFL program offers a fee voucher service for the convenience of organizations and agencies that pay TOEFL and/or TSE test fees for some or all of their students or applicants. Each fee voucher card shows the name and code number of the participating institution and is valid only for the specified testing year and the specific program (TOEFL or TSE) indicated thereon. To participate in the service, institutions must sponsor a minimum of 10 candidates per year.

Fee voucher cards are distributed by the participating institution or agency directly to the applicants for whom it will pay the test fees. The applicants, in turn, submit the completed cards in lieu of personal payment with their completed registration forms. Following each TOEFL test administration, the sponsor receives the test scores of each sponsored examinee who submitted a fee voucher card and an invoice for the number of cards accepted and processed at ETS.

It is important that applicants register before the registration closing date for the administration at which they wish to test. Test centers will not accept fee voucher cards as admission documents.

### **TOEFL Fee Certificate Service**

The TOEFL Fee Certificate Service allows family or friends in the United States, Canada, and other countries where US dollars are available to purchase certificates from the TOEFL program office. A purchaser can then send the certificate to an individual living in a country with currency exchange restrictions to use as proof of payment for the test fee when the prospective test taker registers for a TOEFL administration.

Although the fee certificates are especially useful to individuals living in countries or areas in which US dollars are difficult or impossible to obtain, the certificates will be accepted as a valid form of TOEFL registration fee payment anywhere in the world (except Japan, Taiwan, and the People's Republic of China) up to 14 months from the date of issue.

### **TOEFL Magnetic Score-Reporting Service**

A magnetic score-reporting service for TOEFL official score-report recipients is available by subscription. The service provides TOEFL score reports twice a month to participating institutions and agencies for a nominal annual charge. Although individual paper score reports continue to be sent to institutions and agencies that are designated TOEFL score recipients, the scores can be sent only to the central address or admissions office listed in the TOEFL files. This service can be ordered for specific offices or departments.

The score records are in single record format on 9 track/1600/6250 bpi magnetic tapes, 3<sup>1</sup>/<sub>2</sub> inch floppy disks, formatted for an IBM or IBM-compatible personal computer, or cartridges. Each tape or disk is accompanied by a roster containing all examinee data included on the tape or disk.

The tapes or disks are prepared for each institution or agency with only the score records of TOEFL examinees who correctly marked the code number of the institution or agency on their answer sheets when they took the test or who submitted a written request that their scores be reported to that institution or agency. The magnetic score-reporting service provides a convenient way to merge students' TOEFL score data with other student data.

Subscription to this service is for one year (July to June) and may begin at any time during the year.

### **Examinee Identification Service for TOEFL and TSE Score Users**

This service provides photo identification of examinees taking the TOEFL and TSE tests. The photo file record is collected by the test center supervisor from each examinee before he or she is admitted to the testing room.

The official score reports routinely sent to institutions and other score recipients designated by the test taker, and the examinee's own copy of the score report, bear an electronically reproduced photo image of the examinee and a copy of the test taker's signature. In a small number of cases, it may not be possible to reproduce an examinee's photo image on the score report. Instead, the words "Photo Available Upon Request" will be printed on the reports. Copies of photographs for these examinees may be obtained by using the Examinee Identification Service.

If there is reason to suspect an inconsistency between a high test score and relatively weak English proficiency, an institution or agency that has received either an official score report from ETS or an examinee's score record from an examinee may request a copy of that examinee's photo file record up to 18 months following the test date shown on the score report. The written request for examinee identification must be accompanied by a photocopy of the examinee's score record or official score report.

## Support for External Research Studies

The TOEFL program will make available certain types of test data or perform analyses of pertinent data requested by external researchers for studies related to assessing English language proficiency. The researchers must agree to (1) protect the confidentiality of the data, (2) assume responsibility for the analyses and conclusions of the studies, and (3) reimburse the TOEFL program for the costs associated with the compilation and formatting of the data.

TOEFL program funding of independent research, if requested and granted, is usually limited to providing test materials and related services without charge and/or the cost of the data access and data analysis.

Individuals interested in utilizing TOEFL test data or materials for research studies should write to the TOEFL program office.

For more information about the programs and services described on pages 39-42, visit TOEFL OnLine at <http://www.toefl.org> or write to:

TOEFL Program Office  
Educational Testing Service  
P.O. Box 6155  
Princeton, NJ 08541-6155



# RESEARCH PROGRAM

The purpose of the TOEFL research program is to further knowledge in the field of language assessment and second language acquisition about issues related to psychometrics, language learning and pedagogy, and the proper use and interpretation of language assessment tools. In light of these diverse goals, the TOEFL research agenda calls for continuing research in several broad areas of inquiry, including test validation, reliability, use, construction, and examinee performance.

The TOEFL Research Committee reviews and approves all research projects and sets guidelines for the scope of the TOEFL research program.

## TOEFL Research Report Series

The results of research studies conducted under the direction of the TOEFL Research Committee are available to the public through the TOEFL Research Report and Technical Report Series. In addition to those listed below, a number of new projects are in progress or under consideration.

Research report titles available as of July 1997:

1. *The Performance of Native Speakers of English on the Test of English as a Foreign Language*. Clark. November 1977.
2. *An Evaluation of Alternative Item Formats for Testing English as a Foreign Language*. Pike. June 1979.
3. *The Performance of Nonnative Speakers of English on TOEFL and Verbal Aptitude Tests*. Angelis, Swinton, and Cowell. October 1979.
4. *An Exploration of Speaking Proficiency Measures in the TOEFL Context*. Clark and Swinton. October 1979.
5. *The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language*. Powers. December 1980.
6. *Factor Analysis of the Test of English as a Foreign Language for Several Language Groups*. Powers and Swinton. December 1980.
7. *The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings*. Clark and Swinton. December 1980.
8. *Effects of Item Disclosure on TOEFL Performance*. Angelis, Hale, and Thibodeau. December 1980.
9. *Item Performance Across Native Language Groups on the Test of English as a Foreign Language*. Alderman and Holland. August 1981.
10. *Language Proficiency as a Moderator Variable in Testing Academic Aptitude*. Alderman. November 1981.
11. *A Comparative Analysis of TOEFL Examinee Characteristics, 1977-1979*. Wilson. September 1982.
12. *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL*. Wilson. October 1982.
13. *The Test of Spoken English as a Measure of Communicative Proficiency in the Health Professions*. Powers and Stansfield. January 1983.
14. *A Manual for Assessing Language Growth in Instructional Settings*. Swinton. February 1983.
15. *A Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*. Bridgeman and Carlson. September 1983.
16. *Summaries of Studies Involving the Test of English as a Foreign Language, 1963-1982*. Hale, Stansfield, and Duran. February 1984.
17. *TOEFL from a Communicative Viewpoint on Language Proficiency: A Working Paper*. Duran, Canale, Penfield, Stansfield, and Liskin-Gasparro. February 1985.
18. *A Preliminary Study of Raters for the Test of Spoken English*. Bejar. February 1985.
19. *Relationship of Admission Test Scores to Writing Performance of Native and Nonnative Speakers of English*. Carlson, Bridgeman, Camp, and Waanders. August 1985.
20. *A Survey of Academic Demands Related to Listening Skills*. Powers. December 1985.
21. *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference*. Stansfield. May 1986.
22. *Patterns of Test Taking and Score Change for Examinees Who Repeat the Test of English as a Foreign Language*. Wilson. January 1987.

23. *Development of Cloze-Elide Tests of English as a Second Language*. Manning. April 1987.
24. *A Study of the Effects of Item Option Rearrangement on the Listening Comprehension Section of the Test of English as a Foreign Language*. Golub-Smith. August 1987.
25. *The Interaction of Student Major-Field Group and Test Content in TOEFL Reading Comprehension*. Hale. January 1988.
26. *Multiple-Choice Cloze Items and the Test of English as a Foreign Language*. Hale, Stansfield, Rock, Hicks, Butler, and Oller. March 1988.
27. *Native Language, English Proficiency, and the Structure of the Test of English as a Foreign Language*. Oltman, Stricker, and Barrows. July 1988.
28. *Latent Structure Analysis of the Test of English as a Foreign Language*. Boldt. November 1988.
29. *Context Bias in the Test of English as a Foreign Language*. Angoff. January 1989.
30. *Accounting for Random Responding at the End of the Test in Assessing Speededness on the Test of English as a Foreign Language*. Secolsky. January 1989.
31. *The TOEFL Computerized Placement Test: Adaptive Conventional Measurement*. Hicks. January 1989.
32. *Confirmatory Factor Analysis of the Test of English as a Foreign Language*. Hale, Rock, and Jirele. December 1989.
33. *A Study of the Effects of Variations of Short-term Memory Load, Reading Response Length, and Processing Hierarchy on TOEFL Listening Comprehension Item Performance*. Henning. February 1991.
34. *Note Taking and Listening Comprehension on the Test of English as a Foreign Language*. Hale. February 1991.
35. *A Study of the Effects of Contextualization and Familiarization on Responses to TOEFL Vocabulary Test Items*. Henning. February 1991.
36. *A Preliminary Study of the Nature of Communicative Competence*. Henning and Cascallar. February 1992.
37. *An Investigation of the Appropriateness of the TOEFL Test as a Matching Variable to Equate TWE Topics*. DeMauro. May 1992.
38. *Scalar Analysis of the Test of Written English*. Henning. August 1992.
39. *Effects of the Amount of Time Allowed on the Test of Written English*. Hale. June 1992.
40. *Reliability of the Test of Spoken English Revisited*. Boldt. November 1992.
41. *Distributions of ACTFL Ratings by TOEFL Score Ranges*. Boldt, Larsen-Freeman, Reed, and Courtney. November 1992.
42. *Topic and Topic Type Comparability on the Test of Written English*. Golub-Smith, Reese, and Steinhaus. March 1993.
43. *Uses of the Secondary Level English Proficiency (SLEP) Test: A Survey of Current Practice*. Wilson. March 1993.
44. *The Prediction of TOEFL Reading Comprehension Item Difficulty for Expository Prose Passages for Three Item Types: Main Idea, Inference, and Supporting Idea Items*. Freedle and Kostin. May 1993.
45. *Test-Retest Analyses of the Test of English as a Foreign Language*. Henning. June 1993.
46. *Multimethod Construct Validation of the Test of Spoken English*. Boldt and Oltman. December 1993.
47. *An Investigation of Proposed Revisions to Section 3 of the TOEFL Test*. Schedl, Thomas, and Way. March 1995.
48. *Analysis of Proposed Revisions of the Test of Spoken English*. Henning, Schedl, and Suomi. March 1995.
49. *A Study of Characteristics of the SPEAK Test*. Sarwark, Smith, MacCallum, and Cascallar. March 1995.

50. *A Comparison of the Performance of Graduate and Undergraduate School Applicants on the Test of Written English.* Zwick and Thayer. May 1995.
51. *An Analysis of Factors Affecting the Difficulty of Dialogue Items in TOEFL Listening Comprehension.* Nissan, DeVincenzi, and Tang. February 1996.
52. *Reader Calibration and Its Potential Role in Equating for the Test of Written English.* Myford, Marr, and Linacre. May 1996.
53. *An Analysis of the Dimensionality of TOEFL Reading Comprehension Items.* Schedl, Gordon, Carey, and Tang. March 1996.
54. *A Study of Writing Tasks Assigned in Academic Degree Programs.* Hale, Taylor, Bridgeman, Carson, Kroll, and Kantor. June 1996.
55. *Adjustment for Reader Rating Behavior in the Test of Written English.* Longford. August 1996.
56. *The Prediction of TOEFL Listening Comprehension Item Difficulty for Minitalk Passages: Implications for Construct Validity.* Freedle and Kostin. August 1996.
57. *Survey of Standards for Foreign Student Applicants.* Boldt and Courtney. August 1997.
58. *Using Just Noticeable Differences to Interpret Test of Spoken English Scores.* Stricker. August 1997.
3. *Development of Procedures for Resolving Irregularities in the Administration of the Listening Comprehension Section of the TOEFL Test.* Way and McKinley. February 1991.
4. *Cross-Validation of the Proportional Item Response Curve Model.* Boldt. April 1991.
5. *The Feasibility of Modeling Secondary TOEFL Ability Dimensions Using Multidimensional IRT Models.* McKinley and Way. February 1992.
6. *An Exploratory Study of Characteristics Related to IRT Item Parameter Invariance with the Test of English as a Foreign Language.* Way, Carey, and Golub-Smith. September 1992.
7. *The Effect of Small Calibration Sample Sizes on TOEFL IRT-Based Equating.* Tang, Way, and Carey. December 1993.
8. *Simulated Equating Using Several Item Response Curves.* Boldt. January 1994.
9. *Investigation of IRT-Based Assembly of the TOEFL Test.* Chyn, Tang, and Way. March 1995.
10. *Estimating the Effects of Test Length and Test Time on Parameter Estimation Using the HYBRID Model.* Yamamoto. March 1995.
11. *Using a Neural Net to Predict Item Difficulty.* Boldt and Freedle. December 1996.
12. *How Reliable is the TOEFL test?* Wainer and Lukhele. August 1997.
13. *Concurrent Calibration of Dichotomously and Polytomously Scored TOEFL Items Using IRT Models.* Tang and Eignor. August 1997.

### **TOEFL Technical Report Series**

This series presents reports of a technical nature, such as those related to issues of multidimensional scaling or item response theory. As of July 1997 there are 13 reports in the series.

1. *Developing Homogeneous Scales by Multidimensional Scaling.* Oltman and Stricker. February 1991.
2. *An Investigation of the Use of Simplified IRT Models for Scaling and Equating the TOEFL Test.* Way and Reese. February 1991.

## TOEFL Monograph Series

As part of the foundation for the TOEFL 2000 project (see page 10), a number of papers were commissioned from experts within the fields of measurement and language teaching and testing. Critical reviews and expert opinions were invited to inform TOEFL program development efforts with respect to test construct, test user needs, and test delivery. These monographs are also of general scholarly interest. Thus, the TOEFL program is pleased to make these reports available to colleagues in the fields of language teaching and testing and international student admissions in higher education.

1. *A Review of the Academic Needs of Native English-Speaking College Students in the United States.* Ginther and Grant. September 1996.
2. *Polytomous Item Response Theory (IRT) Models and Their Applications in Large-Scale Testing Programs: Review of Literature.* Tang. September 1996.
3. *A Review of Psychometric and Consequential Issues Related to Performance Assessment.* Carey. September 1996.
4. *Assessing Second Language Academic Reading from a Communicative Competence Perspective: Relevance for TOEFL 2000.* Hudson. September 1996.
5. *TOEFL 2000 — Writing: Composition, Community, and Assessment.* Hamp-Lyons and Kroll. March 1997.
6. *A Review of Research into Needs in English for Academic Purposes of Relevance to the North American Higher Education Context.* Waters. November 1996.
7. *The Revised Test of Spoken English (TSE): Discourse Analysis of Native Speaker and Nonnative Speaker Data.* Lazaraton and Wagner. December 1996.
8. *Testing Speaking Ability in Academic Contexts: Theoretical Considerations.* Douglas. April 1997.
9. *Theoretical Underpinnings of the Test of Spoken English Revision Project.* Douglas and Smith. May 1997.
10. *Communicative Language Proficiency: Definition and Implications for TOEFL 2000.* Chapelle, Grabe, and Berns. May 1997.

See TOEFL OnLine at <http://www.toefl.org> for new reports as they are published.



# PUBLICATIONS

## **TOEFL Products and Services Catalog**

The catalog provides summaries with photographs of the priced training and student study materials developed by the TOEFL program staff. There are also brief descriptions of the testing programs and related services.

## **Bulletin of Information for TOEFL, TWE, and TSE**

This publication is the primary source of information for individuals who wish to take the TOEFL, TWE, and TSE tests at Friday or Saturday testing program administrations. The *Bulletin* tells examinees how to register, lists the test centers, provides a brief description of the tests, and explains score reporting and other procedures. It also contains the TOEFL, TWE, and TSE calendar, which includes the test dates, registration deadline dates, and mailing dates for official score reports. In addition, there are practice questions, detailed instructions for filling out the answer sheet on the day of the test, an explanation of procedures to be followed at the test center, and information about interpreting scores on the tests.

Copies of the *Bulletin* are available at many counseling or advising centers, United States embassies, and offices of the United States Information Service (USIS). In countries and regions where registration is handled by TOEFL representatives, the representatives distribute appropriate editions of the *Bulletin* to examinees and local institutions.

## **Test Center Reference List**

The *Test Center Reference List* provides TOEFL, TWE, and TSE test dates, registration deadline dates, score report mailing dates, and test center locations for the Friday and Saturday testing programs. It also tells how to obtain the appropriate edition of the *Bulletin*. The free list is distributed at the beginning of each testing year to institutions and organizations that use TOEFL, TWE, and TSE test scores.

## **Test Forms Available to TOEFL Examinees**

At some test center locations, examinees who actually take the test on dates announced in advance by the TOEFL office may obtain the test books used at these administrations free of charge. In addition, these examinees may order a list of the correct answers, a cassette recording of Section 1 (Listening Comprehension), and a copy of their answer sheet with the raw scores marked.

Information about when and how examinees may avail themselves of this service is given in the appropriate *Bulletin* editions for the areas where the service is available.

An order form with information about how to order and pay for the materials is printed on the inside back covers of the test books for these test administrations.

The availability of this material is subject to change without notice.

## **Guidelines for TOEFL Institutional Validity Studies**

This publication provides institutions currently using the TOEFL test with a set of general guidelines to consider when planning local predictive validity studies. It covers preliminary considerations, selecting criteria, specifying subgroups, determining size of group to be studied, selecting predictors, and determining decision standards, and provides reference sources.

## **TOEFL Test and Score Data Summary**

The performance of groups of examinees who took the TOEFL test during the most recently completed testing year (June-July) is summarized here. Percentile ranks for section and total scaled scores are given for graduate and undergraduate students, as well as for applicants applying for a professional license. Means and standard deviations are provided in table format for both males and females. Of particular interest to many admissions administrators are the data on section and mean scores for examinees classified by native language and geographic region and native country.



## Institutional Testing Program Brochure

The Institutional Testing Program (ITP) brochure contains a description of the TOEFL and Pre-TOEFL tests offered under this program. The brochure also provides sample test questions, details about ETS policy regarding testing, information about TOEFL and Pre-TOEFL score interpretation and the release of examinee score data, and an order form.

## TOEFL Test of Written English Guide

This publication provides a detailed description of the TWE test as well as the TWE scoring criteria and procedures. It also provides guidelines for the interpretation and use of TWE scores, statistical data related to examinee performance on the test, and sample TWE items and essays.

## TSE Score User's Manual

The *Manual* details the development, use, and scoring of the Test of Spoken English and its off-the-shelf version, SPEAK (Speaking Proficiency English Assessment Kit). Guidelines for score use and interpretation are also provided.

## Secondary Level English Proficiency Test Brochure

This publication describes the SLEP test, a convenient, off-the-shelf testing program for nonnative English speaking students entering grades 7 through 12. It includes sample test questions and ordering information.

## The Researcher

This publication contains brief descriptions of all the studies done by ETS researchers specific to the TOEFL tests and testing programs. Published annually, *The Researcher* is available to anyone interested in ongoing research in such areas as language assessment, examinee performance, reliability, and test validation. (See pages 43-46 for a list of titles in the series.)

To obtain additional copies of the *TOEFL Test and Score Manual* or any of the free publications described above, order on-line at <http://www.toefl.org> or write to:

TOEFL Program Office  
Educational Testing Service  
P.O. Box 6155  
Princeton, NJ 08541-6155



# TOEFL STUDY MATERIALS FOR THE PAPER-BASED TESTING PROGRAM.....

The study materials described here are official publications of the TOEFL program. They are produced by test specialists at ETS to help individuals planning to take TOEFL understand the specific linguistic skills the test measures and become familiar with the multiple-choice formats used.

## TOEFL Sample Test

This popular and very economical study product has been expanded and completely updated. It contains instructions for taking the TOEFL test and marking the answers, one practice test, answer sheets for “gridding” the answers to the multiple-choice questions, an answer key, recorded material for the Listening Comprehension section of the test, and scoring information. It also contains practice exercises for the Test of Written English.

## TOEFL Practice Tests, Volume 1 TOEFL Practice Tests, Volume 2

These products were created for those who want more than one test form for practice. Volume 1 contains two tests; Volume 2 contains four. Each volume provides instructions for taking the test, answer sheets, keys, recorded listening comprehension material with corresponding scripts, and scoring information. *TOEFL Practice Tests* provide hours of exercise material.

## TOEFL Test Preparation Kit

(new edition available Spring 1998)

The *Test Preparation Kit* is the most comprehensive TOEFL study product produced by ETS test specialists. This kit provides the user with extensive practice material in all three sections of the TOEFL test, as well as the Test of Written English. The kit contains

- four audio cassettes with more than 230 minutes of recorded answer sheet instructions and listening comprehension material

- workbook with practice and review materials, answer sheets, lists of the correct answers, and a unit devoted to the TWE test
- *sealed* Test Exercise Book containing the TOEFL and TWE tests — just like the material distributed at the test center

The *TOEFL Test Preparation Kit* gives the student an opportunity to hear and practice the kinds of questions that are contained in the paper-based TOEFL test.

## Econo Ten-packs

*Econo Ten-packs* help to reduce costs for those working in group settings, such as ESL study programs, language laboratories, and training classes and workshops. *Ten-packs* are available for the *TOEFL Test Preparation Kit* and the *TOEFL Practice Tests, Volume 2*. Each pack contains 10 sets of printed material from the corresponding study product.

**Note:** The instructor needs to purchase only one product package containing the recorded materials.

Information about ordering TOEFL study materials can be found in the *TOEFL Products and Services Catalog* (see page 47) or on our website at

<http://www.toefl.org>

# REFERENCES

Alderman, D. L. *TOEFL item performance across seven language groups* (TOEFL Research Report 9). Princeton, NJ: Educational Testing Service, 1981.

American Association of Collegiate Registrars and Academic Officers. *AACRAO-AID participant selections and placement study* (Report to the Office of International Training Agency for International Development, U.S. Department of State). Washington, DC: Author, 1971.

American Language Institute (Georgetown). *A report on the results of English testing during the 1966 Pre-University Workshop at the American Language Institute*. Unpublished manuscript. Georgetown University, 1966.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Psychological Association, 1985.

Anastasi, A. *Psychological testing* (3rd ed.). New York: Macmillan, 1968.

Angelis, P. J., Swinton, S. S., and Cowell, W. R. *The performance of nonnative speakers of English on TOEFL and verbal aptitude tests* (TOEFL Research Report 3). Princeton, NJ: Educational Testing Service, 1979.

Angoff, W. A. *Context bias in the test of English as a foreign language* (TOEFL Research Report 29). Princeton, NJ: Educational Testing Service, 1989.

Angoff, W. A., and Sharon, A. T. *A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to United States colleges* (ETS Research Bulletin No. 70-8). Princeton, NJ: Educational Testing Service, 1970.

Bachman, L. F., Kunnan, A., Vanniarajan, S., and Lynch, B. Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries. *Language Testing*, 1988, 5(2), 128-159.

Boldt, R. F. *Latent structure analysis of the Test of English as a Foreign Language* (TOEFL Research Report 28). Princeton, NJ: Educational Testing Service, 1988.

Brennan, R. L. (1994) *Standard Setting from the Perspective of Generalizability Theory*. Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments. Washington, DC.

Chase, C. I., and Stallings, W. M. *Tests of English language as predictors of success for foreign students*. Indiana Studies in Prediction No. 8. Monograph of the Bureau of Educational Studies and Testing. Bloomington, IN: Bureau of Educational Studies and Testing, Indiana University, 1966.

Clark, J. L. D. *The performance of native speakers of English on the Test of English as a Foreign Language* (TOEFL Research Report 1). Princeton, NJ: Educational Testing Service, 1977.

Cook, L. and Eignor, D. R.. *IRT equating methods. Educational Measurement: Issues and Practice*, 10(3), 37-45. 1991.

Cowell, W. R. Item-response theory pre-equating in the TOEFL testing program. In P. W. Holland and D. B. Rubin (Eds.), *Test equating* (pp. 149-161). New York: Academic Press, 1982.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W., and Liskin-Gasparro, J. *TOEFL from a communicative viewpoint on language proficiency: A working paper* (TOEFL Research Report 17). Princeton, NJ: Educational Testing Service, 1985.

Freedle, R., and Kostin, I. *The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main idea, inference, and supporting idea items*. (TOEFL Research Report 44). Princeton, NJ: Educational Testing Service, 1993.

Freedle, R., and Kostin, I. *The prediction of TOEFL listening comprehension item difficulty for minitalk passages: Implications for construct validity* (TOEFL Research Report 56). Princeton, NJ: Educational Testing Service, 1996.

Gershman, J. *Testing English as a foreign language: Michigan/TOEFL study*. Unpublished manuscript. Toronto Board of Education, 1977.

Hale, G. A., Rock, D. A., and Jirele, T. *Confirmatory factor analysis of the Test of English as a Foreign Language* (TOEFL Research Report 32). Princeton, NJ: Educational Testing Service, 1989.

Hambleton, R. K., and Swaminathan, H. *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff, 1985.

Harvill, L. M., (1991). An NCME Instructional Module on Standard Error of Measurement. *Educational Measurement: Issues and Practice*, 10(2), 33-41. Washington, DC.

Heil, D. K., and Aleamoni, L. M. *Assessment of the proficiency in the use and understanding of English by foreign students as measured by the Test of English as a Foreign Language* (Report No. RR-350). Urbana: University of Illinois. (ERIC Document Reproduction Service No. ED 093 948), 1974.

Henning, G. *A study of the effects of variations of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance* (TOEFL Research Report 33). Princeton, NJ: Educational Testing Service, 1991a.

Henning, G. *A study of the effects of contextualization and familiarization on responses to TOEFL vocabulary test items* (TOEFL Research Report 35). Princeton, NJ: Educational Testing Service, 1991b.

Henning, G., and Cascallar, E. *A preliminary study of the nature of communicative competence* (TOEFL Research Report 36). Princeton, NJ: Educational Testing Service, 1992.

Homburg, T. J. TOEFL and GPA: an analysis of correlations. In R. Silverstein (Ed.), *Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing*. Occasional Papers on Linguistics, No. 6. Carbondale: Southern Illinois University, 1979.

Hwang, K. Y., and Dizney, H. F. Predictive validity of the Test of English as a Foreign Language for Chinese graduate students at an American university. *Educational and Psychological Measurement*, 1970, 30, 475-477.

Jaeger, R. M. (1994). *On the Cognitive Construction of Standard-Setting Judgments: The Case of Configural Scoring*. Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments. Washington, DC.

Linn, R., and Slinde, J. The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 1977, 47(1), 121-150.

Livingston, S. A., and Zieky, M. J. *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service, 1982.

Lord, F. M. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum, 1980.

Lord, F. M., and Novick, M. R. *Statistical theories of mental test scores*. Reading, MA: Addison Wesley, 1968.

Magnusson, D. *Test theory*. Boston: Addison-Wesley, 1967.

Maxwell, A. *A comparison of two English as a foreign language tests*. Unpublished manuscript. University of California (Davis), 1965.

McKinley, R. L. An introduction to item response theory. *Measurement and Evaluation in Counseling and Development*, 1989, 22(1), 37-57.

Messick, S. *Validity* (ETS Research Bulletin No. 87-40). Princeton, NJ: Educational Testing Service, 1987. Also appears in R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: MacMillan, 1988.

Nissan, S., DeVincenzi, F., and Tang, K.L. *An analysis of factors affecting the difficulty of dialog items in TOEFL listening comprehension* (TOEFL Research Report 51). Princeton, NJ: Educational Testing Service, 1996.

Odunde, O. J. Test of English as a Foreign Language and first year GPA of Nigerian students (Doctoral dissertation, University of Missouri-Columbia, 1980). *Dissertation Abstracts International*, 42, 3419A-3420A. (University Microfilms No. 8202657), 1982.

Oltman, P. K., Stricker, L. J., and Barrows, T. *Native language, English proficiency, and the structure of the Test of English as a Foreign Language* (TOEFL Research Report 27). Princeton, NJ: Educational Testing Service, 1988.

Pack, A.C. A comparison between TOEFL and Michigan Test scores and student success in (1) freshman English and (2) completing a college program. *TESL Reporter*, 1972, 5, 1-7, 9.

Pike, L. *An evaluation of alternative item formats for testing English as a foreign language* (TOEFL Research Report 2). Princeton, NJ: Educational Testing Service, 1979.

Powers, D. E. *The relationship between scores on the Graduate Management Admission Test and the Test of English as a Foreign Language* (TOEFL Research Report 5). Princeton, NJ: Educational Testing Service, 1980.

Powers, D. E. *A survey of academic demands related to listening skills* (TOEFL Research Report 20). Princeton, NJ: Educational Testing Service, 1985.

Schrader, W. B., and Pitcher, B. *Interpreting performance of foreign law students on the Law School Admission Test and the Test of English as a Foreign Language* (Statistical Report 70-25). Princeton, NJ: Educational Testing Service, 1970.

Schedl, M., Thomas, N., and Way, W. *An investigation of proposed revisions to the TOEFL test* (TOEFL Research Report 47). Princeton, NJ: Educational Testing Service, 1995.

Sharon, A. T. *Test of English as a Foreign Language as a moderator of Graduate Record Examinations scores in the prediction of foreign students' grades in graduate school* (ETS Research Bulletin No. 71-50). Princeton, NJ: Educational Testing Service, 1971.

Stansfield, C. W., Ed. *Toward communicative competence testing: Proceedings of the Second TOEFL Invitational Conference* (TOEFL Research Report 21). Princeton, NJ: Educational Testing Service, 1986.

Swineford, F. *Test analysis: Test of English as a Foreign Language* (Statistical Report 71-112). Princeton, NJ: Educational Testing Service, 1971.

Swinton, S. S. *A manual for assessing language growth in instructional settings* (TOEFL Research Report 14). Princeton, NJ: Educational Testing Service, 1983.

*Test of English as a Foreign Language: Interpretive information*. Princeton, NJ: College Entrance Examination Board and Educational Testing Service, 1970.

Thorndike, R. I., and Hagen, E. P. *Measurement and evaluation in education* (4th ed.). New York: Wiley, 1977.

Upshur, J. A. *Comparison of performance on "Test of English as a Foreign Language" and "Michigan Test of English Language Proficiency."* Unpublished manuscript. University of Michigan, 1966.

Van der Linden, W. J. (1994). *A Conceptual Analysis of Standard Setting in Large-Scale Assessments*. Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments. Washington, DC.

Wainer, H., and Lukhele, R. *How reliable is the TOEFL test?* (TOEFL Technical Report 12). Princeton, NJ: Educational Testing Service, in press.

Wilson, K. M. *GMAT and GRE Aptitude Test performance in relation to primary language and scores on TOEFL* (TOEFL Research Report 12). Princeton, NJ: Educational Testing Service, 1982.

**TOEFL** .....

*Thirty-Four Years of Commitment to High Standards and Quality Service Around the World*



# ETS OFFICES SERVING TOEFL CANDIDATES AND SCORE USERS

---

## Main Office

TOEFL/TSE Services  
Educational Testing  
Service  
P.O. Box 6151  
Princeton, NJ 08541-6151  
  
Phone: 609-771-7100  
Fax: 609-771-7500  
TTY: 609-734-9362

## Washington, DC

Educational Testing  
Service  
Suite 620  
1776 Massachusetts  
Avenue, NW  
Washington, DC 20036  
  
Phone: 202-659-0616  
Fax: 202-659-8075  
TTY: 202-659-8067

## South

Educational Testing  
Service  
Suite 400  
Lakeside Centre  
1979 Lakeside Parkway  
Tucker, GA 30084  
  
Phone: 770-934-0133  
Fax: 770-723-7436  
TTY: 770-934-2624

## Midwest

Educational Testing  
Service  
Suite 300  
One Rotary Center  
1560 Sherman Avenue  
Evanston, IL 60201  
  
Phone: 847-869-7700  
Fax: 847-492-5141  
TTY: 847-869-7738

## Southwest

Educational Testing  
Service  
Suite 700  
2 Renaissance Square  
40 North Central Avenue  
Phoenix, AZ 85004  
  
Phone: 602-252-5400  
Fax: 602-252-7499  
TTY: 602-252-0276

## West

Educational Testing  
Service  
Suite 310  
Trans Pacific Centre  
1000 Broadway  
Oakland, CA 94607  
  
Phone: 510-873-8000  
Fax: 510-873-8118  
TTY: 510-465-5571

## Puerto Rico

Educational Testing  
Service  
Suite 315  
American International  
Plaza  
250 Munoz Rivera Avenue  
Hato Rey, PR 00918  
  
Phone: 787-753-6363  
Fax: 787-250-7426  
TTY: 787-758-4598



# TOEFL REPRESENTATIVES

**Algeria, Oman, Qatar, Saudi Arabia, Sudan, United Arab Emirates**  
AMIDEAST  
Testing Programs  
1730 M Street, NW, Suite 1100  
Washington, DC 20036-4505, USA

**Australia, New Zealand, Papua New Guinea, Solomon Islands, Vanuatu**  
Australian Council for  
Educational Research  
ACER-ETS Administration Office  
Private Bag 55  
Camberwell, Victoria 3124  
Australia

**Bahrain**  
AMIDEAST  
P.O. Box 10410  
Manama, Bahrain

**Brazil**  
Instituto Brasil-Estados Unidos  
Av. Nossa Senhora de Copacabana  
690-6° Andar  
22050-000 Rio de Janeiro, RJ  
Brasil

**Egypt**  
AMIDEAST  
6 Kamel El Shennawy Street  
Second Floor, Apartment 5  
Garden City, Cairo, Egypt  
or  
AMIDEAST  
American Cultural Center  
3 Pharaana Street  
Azarita, Alexandria  
Egypt

**Europe, (East/West)**  
CITO-TOEFL  
P.O. Box 1203  
6801 BE Arnhem  
Netherlands

**Hong Kong**  
Hong Kong Examinations  
Authority  
San Po Kong Sub-Office  
17 Tseuk Luk Street  
San Po Kong  
Kowloon, Hong Kong

**India/Bhutan**  
Institute of Psychological and  
Educational Measurement  
Post Box No. 19  
119/25-A Mahatma Gandhi Marg  
Allahabad, U.P. 211 001, India

**Indonesia**  
International Education  
Foundation (IEF)  
Menara Imperium  
Lantai 28, Suite B  
Jl. H.R. Rasuna Said, Kav. 1  
Kuningan, Jakarta Sealtan 12980  
Indonesia

**Japan**  
Council on International  
Educational Exchange (CIEE)  
TOEFL Division  
Cosmos Aoyama B1  
5-53-67 Jingumae, Shibuya-ku  
Tokyo 150, Japan

**Jordan**  
AMIDEAST  
P.O. Box 1249  
Amman, Jordan

**Korea**  
Korean-American Educational  
Commission (KAEC)  
K.P.O. Box 643  
Seoul 110-606, Korea

**Kuwait**  
AMIDEAST  
P.O. Box 44818  
Hawalli 32063, Kuwait

**Lebanon**  
AMIDEAST  
P.O. Box 135-155  
Ras Beirut, Lebanon  
or  
AMIDEAST  
P.O. Box 70-744  
Antelias, Beirut, Lebanon

**Malaysia/Singapore**  
MACEE  
Testing Services  
191 Jalan Tun Razak  
50400 Kuala Lumpur, Malaysia

**Mexico**  
Institute of International  
Education  
Londres 16, 2nd Floor  
Apartado Postal 61-115  
06600 Mexico D.F., Mexico

**Morocco**  
AMIDEAST  
25 bis, Patrice Lumumba  
Apt. No. 8  
Rabat, Morocco

**Pakistan**  
World Learning Inc.  
P.O. Box 13042  
Karachi 75350, Pakistan

**People's Republic of China**  
China International Examinations  
Coordination Bureau  
No. 167 Haidian Road  
Haidian District  
Beijing 100080  
People's Republic of China

**Syria**  
AMIDEAST  
P.O. Box 2313  
Damascus, Syria

**Taiwan**  
The Language Training & Testing  
Center  
P.O. Box 23-41  
Taipei, Taiwan

**Thailand, Cambodia, Laos, Vietnam**  
Institute of International Education  
G.P.O. Box 2050  
Bangkok 10501, Thailand

**Tunisia**  
AMIDEAST  
BP 351 Tunis-Belvedere 1002  
Tunis, Tunisia

**United Arab Emirates**  
AMIDEAST  
c/o Higher Colleges of Technology  
P.O. Box 5464  
Abu Dhabi, UAE

**All Other Countries and Areas**  
TOEFL/TSE Services  
P.O. Box 6161  
Princeton, NJ 08541-6161  
USA





**Test of English as a Foreign Language**  
**P.O. Box 6155**  
**Princeton, NJ 08541-6155**  
**USA**

---

To obtain more information  
about TOEFL products and services,  
use one of the following:

**Phone: 609-771-7100**

**E-mail: [toefl@ets.org](mailto:toefl@ets.org)**

**Website: <http://www.toefl.org>**