

TSE[®] and SPEAK[®]

2001-2002
EDITION



SCORE USER GUIDE

TEST OF
SPOKEN
ENGLISH
&
SPEAKING
PROFICIENCY
ENGLISH
ASSESSMENT KIT



ETS[®]
Educational
Testing Service

www.toefl.org

The TSE program does not operate, license, endorse, or recommend any schools or study materials that claim to prepare people for the TSE or SPEAK test in a short time or that promise them high scores on the test.



Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.
Copyright © 2001 by Educational Testing Service. All rights reserved.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logos, SPEAK, the SPEAK logo, TOEFL, the TOEFL logo, TSE, the TSE logo, and TWE are registered trademarks of Educational Testing Service. The Test of English as a Foreign Language, Test of Spoken English, and Test of Written English are trademarks of Educational Testing Service.

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 online at www.toefl.org/copyrigh.html or sent to:

Proprietary Rights Office
Educational Testing Service
Rosedale Road
Princeton, NJ 08541-0001, USA
Phone: 1-609-734-5032

Preface

This 2001 edition of the *TSE Score User Guide* supersedes the *TSE Score User's Manual* published in 1995.

The *Guide* has been prepared for foreign student advisers, college deans and admissions officers, scholarship program administrators, department chairpersons and graduate advisers, teachers of English as a second language, licensing boards, and others responsible for interpreting TSE scores. In addition to describing the test, testing program, and rating scale, the *Guide* discusses score interpretation, TSE examinee performance, and TSE-related research.

Your suggestions for improving the usefulness of the *Guide* are most welcome. Please feel free to send any comments to us at the following address:

TSE Program Office
TOEFL Programs and Services
Educational Testing Service
PO Box 6157
Princeton, NJ 08541-6157, USA

Language specialists prepare TSE test questions. These specialists follow careful, standardized procedures developed to ensure that all test material is of consistently high quality. Each question is reviewed by several members of the ETS staff. The TSE Committee, an independent group of professionals in the fields of linguistics and language training that reports to the TOEFL Board, is responsible for the content of the test.

After test questions have been reviewed and revised as appropriate, they are selectively administered in trial situations and assembled into test forms. The test forms are then reviewed according to established ETS and TSE program procedures to ensure that the forms are free of cultural bias. Statistical analyses of individual questions, as well as of the complete tests, ensure that all items provide appropriate measurement information.

Table of Contents

Overview of the TSE Test	4
Purpose of the TSE test	4
Relationship of the TSE test to the TOEFL program	4
Development of the Original TSE Test	5
Revision of the TSE Test	6
The TSE Committee	6
Overview of the TSE test revision process	6
Purpose and format of the revised test	6
Test construct	7
Validity of the test	7
Reliability and SEM	9
Content and Program Format of the TSE Test .	10
Test content	10
Test registration	10
Administration of the test	11
Individuals with disabilities	11
Measures to protect test security	11
TSE score cancellation by ETS	12
Scores for the TSE Test	13
Scoring procedures	13
Scores and score reports	13
Confidentiality of TSE scores	13
Requests for TSE rescoring	15
TSE test score data retention	15
Use of TSE Scores	16
Setting score standards	16
TSE sample response tape	16
Guidelines for using TSE test scores	16
Statistical Characteristics of the TSE Test: Performance of Examinees on the Test of Spoken English	17
Speaking Proficiency English Assessment Kit (SPEAK)	21
Research	22
TOEFL research program	22
Research and related reports	22
References	27
Appendices	28
A. TSE Committee Members	28
B. TSE Rating Scale, TSE and SPEAK Band Descriptor Chart .	29
C. Glossary of Terms Used in TSE Rating Scale	31
D. Frequently Asked Questions and Guidelines for Using TSE or SPEAK Scores	32
E. Sample TSE Test	34
Where to Get TSE Bulletins	36

Overview of the TSE Test

Purpose of the TSE test

The primary purpose of the Test of Spoken English (TSE®) is to measure the ability of nonnative speakers of English to communicate orally in a North American English context. The TSE test is delivered in a semidirect format, which maintains reliability and validity while controlling for the subjective variables associated with direct interviewing. Because it is a test of general oral language ability, the TSE test is appropriate for examinees regardless of native language, type of educational training, or field of employment.

There are two separate registration categories within the TSE program: **TSE-A** and **TSE-P**.

TSE-A is for teaching and research assistant applicants who have been requested to take the TSE test by the admissions office or department chair of an academic institution. **TSE-A** is also for other undergraduate or graduate school applicants.

TSE-P is for all other individuals, such as those who are taking the TSE test to obtain licensure or certification in a professional or occupational field.

The TSE test has broad applicability because performance on the test indicates how oral language ability might affect the examinee's ability to communicate successfully in either academic or professional environments. TSE scores are used at many North American institu-

tions of higher education in the selection of international teaching assistants (ITAs). The scores are also used for selection and certification purposes in the health professions, such as medicine, nursing, pharmacy, and veterinary medicine, and for the certification of English teachers overseas and in North America.

TSE scores should not be interpreted as predictors of academic or professional success, but only as indicators of nonnative speakers' ability to communicate in English. The scores should be used in conjunction with other types of information about candidates when making decisions about their ability to perform in an academic or professional situation.

Relationship of the TSE test to the TOEFL program

The TSE program is administered by Educational Testing Service (ETS) through the Test of English as a Foreign Language (TOEFL) program.

Policies governing the TOEFL, TSE, and Test of Written English (TWE®) programs are formulated by the TOEFL Board, an external group of academic specialists in fields related to international admissions, student exchange and language education, and assessment. The Board was established by and is affiliated with the College Board and the Graduate Record Examinations Board.

Development of the Original TSE Test

The original Test of Spoken English was developed during the late 1970s in recognition of the fact that academic institutions often needed an accurate measure of speaking ability in order to make informed selection and employment decisions. At that time there was an emphasis in the fields of linguistics, language teaching, and language testing on accuracy in pronunciation, grammar, and fluency. The test was designed to measure these linguistic features and to evaluate a speaker's ability to convey information intelligibly to the listener. Test scores were derived for pronunciation, grammar, fluency, and overall comprehensibility.

In 1978 the TOEFL Research Committee and the TOEFL Board sponsored a study entitled "An Exploration of Speaking Proficiency Measures in the TOEFL Context" (Clark and Swinton, 1979). The report of this study details the measurement rationale and procedures used in developing the TSE test, as well as the basis for the selection of the particular formats and question types included in the original form of the test.

A major consideration in developing a measure of speaking ability was for it to be amenable to standardized administration at worldwide test centers. This factor immediately eliminated the subjective variables associated with direct, face-to-face interviewing. Providing the necessary training in interviewing techniques on a worldwide basis was considered impractical.

Another factor addressed during the development of the original TSE test was its linguistic content. Because the test would be administered in many countries, it had to be appropriate for all examinees regardless of native language or culture.

A third factor in test design considerations was the need to elicit evidence of general speaking ability rather than ability in a particular language-use situation. Because the test would be used to predict examinees' speaking ability in a wide variety of North American contexts, it could not use item formats or individual questions that would require extensive familiarity with a particular subject matter or employment context.

Two developmental forms of the TSE test were administered to 155 examinees, who also took the TOEFL test and participated in an oral proficiency interview modeled on that administered by the Foreign Service Institute (FSI). The specific items included on the prototype forms were selected with the goal of maintaining the highest possible correlation with the FSI rating and the lowest possible correlation with the TOEFL score to maximize the usefulness of the speaking test.

Validation of the TSE test was supported by research that indicated the relationship between the TSE comprehensibility scores and FSI oral proficiency levels, the intercorrelations among the four TSE scores, and the correlation of university instructors' TSE scores with student assessments of the instructors' language skills (Clark and Swinton, 1980).

Subsequent to the introduction of the test for use by academic institutions in 1981, additional research (Powers and Stansfield, 1983) validated TSE scores for selection and certification in health-related professions (e.g., medicine, nursing, pharmacy, and veterinary medicine).

Revision of the TSE Test

Since the introduction of the original TSE test in 1981, language teaching and language testing theory and practice have evolved to place a greater emphasis on overall communicative language ability. This contemporary approach includes linguistic accuracy as only one of several aspects of language competence related to the effectiveness of oral communication. For this reason, the TSE test was revised to better reflect current views of language proficiency and assessment. The revised test was first administered in July 1995.

The TSE Committee

In April 1992 the TOEFL Board approved the recommendation of the TOEFL Committee of Examiners to revise the TSE test and to establish a separate TSE Committee to oversee the revision effort.

TSE Committee members are appointed by the TOEFL Board Executive Committee. The TSE Committee includes specialists in applied linguistics and spoken English language teaching and testing, TSE chief raters, and representative score users. As the TSE test development advisory group, the TSE Committee approves the test specifications and score scale, reviews test questions and item performance, offers guidance for rater training and score use, and makes suggestions for further research, as needed.

Members of the TSE Committee are rotated on a regular basis to ensure the continued introduction of new ideas and perspectives related to the assessment of oral language proficiency. Appendix A lists current and former TSE Committee members.

Overview of the TSE test revision process

The TSE revision project begun in 1992 was a joint effort of the TSE Committee and ETS staff. This concentrated three-year project required

articulation of the underlying theoretical basis of the test and the test specifications as well as revision of the rating scale. Developmental research included extensive pilot testing of both test items and rating materials, a large-scale prototype research study, and a series of studies to validate the revised test and scoring system. Program publications underwent extensive revision, and the *TSE Standard-Setting Kit* was produced to assist users in establishing passing scores for the revised test. Extensive rater training and retraining were also conducted to set rating standards and assure appropriate implementation of the revised scoring system.

Purpose and format of the revised test

At the outset of the TSE revision project, it was agreed that the test purpose remained unchanged. That is, the test would continue to be one of general speaking ability designed to evaluate the oral language proficiency of nonnative speakers of English who were at or beyond the postsecondary level of education. It would continue to be of usefulness to the primary audience for the original TSE test (i.e., those evaluating prospective ITAs [international teaching assistants] and personnel in the health-related professions). In this light, it was designed as a measure of the examinee's ability to successfully communicate in North American English in an academic or professional environment.

It was also determined that the TSE test would continue to be a semidirect speaking test administered via audio-recording equipment using prerecorded prompts and printed test books, and that the examinee's recorded responses, or speech sample, would be scored independently by at least two trained raters. Pilot testing of each test form allows ETS to monitor the performance of all test questions.

Test construct

The TSE Committee commissioned a paper by Douglas and Smith (TOEFL MS-9, 1997) to provide a review of the research literature, outline theoretical assumptions about speaking ability, and serve as a guide for test revision. This paper, *Theoretical Underpinnings of the Test of Spoken English Revision Project*, described models of language use and language competence, emphasizing how they might inform test design and scoring. The paper also acknowledged the limitations of an audio-delivered test compared to a direct interview.

As derived from the theory paper, the construct underlying the revised test is communicative language ability. The TSE test was revised on the premise that language is a dynamic vehicle for communication, driven by underlying competencies that interact in various ways for effective communication to take place. For the purposes of the TSE, this communicative language ability has been defined to include strategic competence and language competence, the latter comprising discourse competence, functional competence, sociolinguistic competence, and linguistic competence.

Critical to the design of the test is the notion that these competencies are involved in the act of successful communication. Using language for an intended purpose or function (e.g., to apologize, to complain) is central to effective communication. Therefore, each test item consists of a language task that is designed to elicit a particular function in a specified context or situation.

Within this framework, a variety of language tasks and functions were defined to provide the structural basis of the revised test. The scoring system was also designed to provide a holistic summary of oral language ability across the communication competencies being assessed.

Validity of the test

A series of validation activities were conducted during the revision of the TSE test to evaluate the adequacy of the test design and to provide evidence for the usefulness of TSE scores. These efforts were undertaken with a process-oriented perspective. That is, the accumulation of validity data was used to inform test revision, make modifications as indicated, and confirm the

appropriateness of both the test design and scoring scale.

Validity refers to the extent to which a test actually measures what it purports to measure.* Although many procedures exist for determining validity, there is no single indicator or standard index of validity. The extent to which a test can be evaluated as a valid measure is determined by judging all available evidence. The test's strengths and limitations must be taken into account, as well as its suitability for particular uses and examinee populations.

Construct validity research was initiated in the theory paper commissioned by the TSE Committee (Douglas and Smith, TOEFL MS-9, 1997). This document discusses the dynamic nature of the construct of oral language ability in the field of language assessment and points the way to a conceptual basis for the revised test. As a result of the paper and discussion among experts in the field, the basic construct underlying the test was defined as communicative language ability. This theoretical concept was operationalized in the preliminary test specifications.

To evaluate the validity of the test design, Hudson (1994) reviewed the degree of congruence between the test's theoretical basis and the test specifications. This analysis suggested a generally high degree of concordance. The test specifications were further revised in light of this review.

In a similar vein, the prototype test was examined by ETS staff for its degree of congruence with the test specifications. This review also led to modest revisions in the test specifications and item writing guidelines in order to provide a high degree of congruence between the theory, specifications, and test forms.

As a means of validating the test content, a discourse analysis of both native and nonnative speaker speech as elicited by the prototype test was conducted (Lazaraton and Wagner, TOEFL MS-7, 1996). The analysis indicated that the language functions intended were reliably and consistently elicited from both native and nonnative speakers, all of whom performed the same types of speech activities.

* The reader is referred to the American Psychological Association's *Standards for Educational and Psychological Testing* (1999), as well as Wainer and Braun's *Test Validity* (1988), for a thorough treatment of the concept of validity.

The test rating scale and score bands were validated through another process. ETS rating staff wrote descriptions of the language elicited in speech samples which were compared to the rating scale and score bands assigned to the samples. This was to determine the degree of agreement between elicited speech and the scoring system. The results confirmed the validity of the rating system.

The concurrent validity of the revised TSE test was investigated in a large-scale research study by Henning, Schedl, and Suomi (TOEFL RR-48, 1995). The sample for this study consisted of subjects representing the primary TSE examinee populations: prospective university teaching assistants (N = 184) and prospective licensed medical professionals (N = 158).

Prospective teaching assistants represented the fields of science, engineering, computer science, and economics. Prospective licensed medical professionals included foreign medical graduates who were seeking licenses to practice as physicians, nurses, veterinarians, or pharmacists in the United States. The subjects in both groups represented more than 20 native languages.

The instruments used in the study included an original version of the TSE test, a 15-item prototype version of the revised test, and an oral language proficiency interview (LPI). The original version and revised prototype were administered under standard TSE conditions.

The study utilized two types of raters: 16 linguistically "naive" raters who were untrained and 40 expert, trained raters. The naive raters, eight from a student population and eight from a potential medical patient population, were selected because they represented groups most likely to be affected by the English-speaking proficiency of the nonnative candidates for whom passing TSE scores are required. These raters were purposely chosen because they had little experience interacting with nonnative English

speakers, and scored only the responses to the prototype. The naive raters were asked to judge the communicative effectiveness of the revised TSE prototype responses of 39 of the subjects as part of validating the revised scoring method. The trained raters scored the examinees' performance on the original TSE test according to the original rating scale and performance on the prototype revised test according to the new rating scale. (The rating scale used in this study to score the revised TSE test was similar though not identical to the final rating scale approved by the TSE Committee in December 1995, which can be found in Appendix B.)

The use of naive raters in this study served to offer additional construct validity evidence for inferences to be made from test scores. That is, untrained, naive raters were able to determine and differentiate varying levels of communicative language ability from the speech performance samples elicited by the prototype test. These results also provided content validity for the rating scale bands and subsequent score interpretation.

Means and standard deviations were computed for the scores given by the trained raters. In this preliminary study, the mean of the scores on the prototype of the revised test was 50.27 and the standard deviation was 8.66. Comparisons made of the subjects' performance on the original TSE test and the prototype of the revised test showed a correlation between scores for the two versions was .83.

As part of the research study, a subsample of 39 examinees was administered a formal oral language proficiency interview recognized by the American Council on the Teaching of Foreign Languages, the Foreign Service Institute, and the Interagency Language Roundtable. The correlation between the scores on the LPI and the prototype TSE test was found to be .82, providing further evidence of concurrent validity for the revised test.

Reliability and SEM

Reliability can be defined as the extent to which test scores are free from errors in the measurement process. A variety of reliability coefficients can exist because errors of measurement can arise from a number of sources. *Interrater reliability* is an index of the consistency of TSE scores assigned by the first and second raters before adjudication. *Test form reliability* is an index of internal consistency among TSE items and provides information about the extent to which the items are assessing the same construct. *Test score reliability* is the degree to which TSE test scores are free from errors when the two sources of error variation are accounted for simultaneously, that is, the variations of examinee-and-rating interaction and of examinee-and-item interaction. Reliability coefficients can range from .00 to .99.* The closer the value of the coefficient to the upper limit, the less error of measurement. Table 1 provides means of interrater, test form, and test score reliabilities for the total examinee group

* This reliability estimate was reached by the use of the Spearman-Brown adjustment, which provides an estimate of the relationship that would be obtained if the average of the two ratings were used as the final score.

and the academic/professional subgroups over the 54 monthly administrations of the TSE test between July 1995 and January 2000.

The *standard error of measurement (SEM)* is an index of how much an examinee's actual proficiency (or true score) can vary due to errors of measurement. SEM is a function of the test score standard deviation and *test score reliability*. An examinee's TSE observed score is expected to be within the range of his or her TSE true score plus or minus the two SEMs (i.e., plus or minus approximately 4 points on the TSE reporting scale) about 95 percent of the time. The average SEM is also shown in Table 1.

Table 1. Average TSE Reliabilities and Standard Errors of Measurement (SEM) — Total Group and Subgroups

(Based on 64,701 examinees who took primary TSE and SPEAK forms between July 1995 and January 2000.)

	Total (N = 64,701)	Academic (N = 29,254)	Professional (N = 35,447)
Interrater Reliability	0.92	0.91	0.92
Test Form Reliability	0.98	0.97	0.98
Test Score Reliability	0.89	0.89	0.90
SEM	2.24	2.26	2.22

Content and Program Format of the TSE Test

Test content

The TSE test consists of 12 questions, 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 test is delivered via audio-recording equipment and a test book. An interviewer on the test tape leads the examinee through the test; the examinee responds into a microphone, and responses are recorded on a separate answer tape.

The time allotted for each response ranges from 30 to 90 seconds, the timing is based on pilot testing results. All the questions asked by the interviewer, as well as the response time, are printed in the test book. The questions on the test are of a general nature and are designed to inform the raters about the candidate's oral communicative language ability.

At the beginning of the test, the interviewer on the test tape asks some general questions that serve as a “warm up” to help examinees become accustomed to speaking on tape and to allow for adjustment of the audio equipment as needed. These initial, unnumbered questions are not scored. Next, the examinees are given 30 seconds to study a map and then are asked some questions about it. Subsequently, the examinees are asked to look at a sequence of pictures and tell the story that the pictures show. Then the examinees are asked to discuss topics of general interest and to describe information presented in a simple graph. Finally, the examinees are asked to present information from a revised schedule and indicate the revisions.

A short video, *Test of Spoken English: An Overview*, provides general information about the background, purpose, and format of the test. The video is approximately 20 minutes long and is available upon request. It is also included in the *TSE Standard-Setting Kit*.

Test registration

The TSE test is administered 12 times a year at test centers throughout the world. TSE administration dates are published in the *Information Bulletin for TSE*.^{*} The *Bulletin* includes a registration form, a general description of the test, the test directions, and a sample test. TSE candidates must complete the registration form and return it to TOEFL/TSE Services with the appropriate test fee. Copies of the *Bulletin* are distributed to TSE test centers, to American embassies, binational centers, language academies, and additional agencies and individuals who express interest in TSE. Often institutions or departments and employers that require TSE scores of applicants include copies of the *Bulletin* when responding to inquiries from nonnative speakers. A supply of *Bulletins* can also be obtained from TOEFL/TSE Services, PO Box 6151, Princeton, NJ 08541-6151, USA.

^{*} Individuals who plan to take the TSE test in India, Korea, or Taiwan should refer to the *Information Bulletin for TSE — India, Korea, Taiwan Edition*. In the People's Republic of China (PRC), where the Test of English as a Foreign Language is administered in the paper-based format, examinees must obtain the *PRC Edition of Bulletin of Information for TOEFL, TWE, and TSE*.

Administration of the test

The TSE test is administered under strictly controlled testing procedures. The actual testing time is approximately 20 minutes. The test can be administered to individuals with cassette tape recorders or to a group using a multiple-recording facility such as a language laboratory.

Because the scores of examinees are comparable only if the same procedures are followed at all test administrations, the TSE Program Office provides detailed guidelines for test center supervisors to ensure uniform administrations. The *TSE Supervisor's Manual* is mailed with the test materials to test supervisors well in advance of the test date. This publication describes the arrangements necessary to prepare for the test administration, discusses the kind of equipment needed, and gives detailed instructions for the actual administration of the test.

TSE regulations, as listed in the *Information Bulletin*, are enforced to prevent cheating and attempts at impersonation.

At the beginning of the administration, before the start of the actual test, examinees are given sealed test books. Once the test begins, examinees listen to a tape recording containing the general directions and test questions. The tape recorders on which examinees' responses are recorded are not stopped at any time during the test unless an unusual circumstance related to the test administration is identified by the administrator.

IMPORTANT: The TSE test is **NOT** administered as part of the TOEFL test. It is administered separately, at the present time.

Individuals with disabilities

The TSE Program Office, in response to requests from individuals with disabilities, will make special arrangements with test center supervisors, where local conditions permit, to administer the TSE test with accommodations. Among the

accommodations that can be provided are extended testing time, breaks, test reader, sign language interpreter, other aids customarily used by the test taker, large print, nonaudio (without oral stimulus), and braille. All requests for accommodations must be approved in accordance with TSE policies and procedures.

Nonstandard scores

The TSE Program Office recommends that alternative methods of evaluating English proficiency be used for individuals who cannot take the TSE under standard conditions. Criteria such as past academic record, recommendations from language teachers or others familiar with the applicant's English proficiency, and/or a personal interview are suggested in lieu of TSE scores.

However, as noted earlier, the TSE Program Office will make special arrangements to administer the test under nonstandard conditions for individuals with disabilities. Because the individual circumstances of nonstandard administrations vary so widely, the TSE Program Office is not able to compare scores obtained at such administrations with those obtained at standard administrations.

Measures to protect test security

To protect the validity of the test scores, the TSE Program Office continually reviews and refines procedures designed to increase the security of the test before, during, and after its administration. Because of the importance of TSE scores to applicants and to institutions, there are inevitably some individuals who engage in practices designed to increase their reported scores. The careful selection of supervisors, a low examinee-to-proctor ratio, and the detailed administration procedures given in the *Supervisor's Manual* are all designed to prevent attempts at impersonation, 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 TSE administration. To be admitted to a test center, every examinee must present an official identification document with a recognizable photograph, such as a valid passport.

Although the passport is the basic document accepted at all test centers, other specific photobearing documents are acceptable for individuals who may not be expected to have passports or who are taking the test in their own countries. Through foreign embassies in the United States and TSE supervisors in foreign countries, TOEFL/TSE Services verifies the types of official photobearing identification documents used in each country, such as national identity cards, registration certificates, and work permits. Detailed information about identification requirements is included in the *Information Bulletin*.

Photo file records

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. The form is collected by the test center supervisor from each examinee before he or she is admitted to the testing room. In addition to verifying the photo identity of the examinee, the supervisor verifies that the name on the official identification document is exactly the same as the name on the photo file record.

Supervision of examinees

Supervisors and room proctors are instructed to exercise extreme vigilance during a test administration to prevent examinees from giving or receiving assistance in any way. While taking the test, examinees may not have anything on their desks but their test books, tape recorders, and admission tickets. They are not permitted to make notes or marks of any kind in their test books.

If a supervisor is certain that someone has given or received assistance on the test, the examinee is dismissed from the testing room and his or her score is not reported. If a supervisor

suspects someone of cheating, a description of the incident is written on the Supervisor's Irregularity Report (included in the *Supervisor's Manual*), which is returned to ETS with the examinee's tape. Suspected and/or confirmed cases of cheating are investigated by the Test Security Office at ETS.

Preventing access to test materials

To ensure that examinees have not seen the test material in advance, new forms of the test are developed regularly.

To help prevent the theft of test materials, procedures have been devised for the secure distribution and handling of these materials. Test tapes and test books (individually sealed and packed in sealed plastic bags) are sent to test centers in sealed boxes that supervisors are required to place in locked storage that is inaccessible to unauthorized persons. Supervisors count the test books upon receipt, after the examinees have begun the test, and at the end of the administration. No one is permitted to leave the testing room until all test books and examinee answer tapes have been accounted for.

TSE supervisors return the test materials to ETS, where they are counted upon receipt. The ETS Test Security Office investigates all cases of missing test materials.

TSE score cancellation by ETS

TSE Services, on behalf of Educational Testing Service, seeks to report scores that accurately reflect the performance of the test taker. ETS has developed test administration and test security standards and procedures with the goals of assuring that all test takers have equivalent opportunities to demonstrate their abilities, and preventing some test takers from gaining unfair advantage over others. **ETS reserves the right to cancel any test score if, in ETS's judgment, there is an apparent discrepancy in photo identification, the test taker has engaged in misconduct in connection with the test, there is a testing irregularity, or there is substantial evidence that the test score is invalid for another reason.**

Scores for the TSE Test

Scoring procedures

TSE answer tapes are scored by trained TSE raters who are experienced teachers and specialists in the field of English or English as a second language. Raters are trained at qualifying workshops conducted by ETS staff. Prior to each test scoring session, raters review answer tapes at various points on the TSE rating scale to maintain accurate scoring. Raters undergo retraining if score discrepancies indicate that it is warranted.

Each TSE tape is rated independently by two raters; neither knows the scores assigned by the other. Each rater evaluates each item response and assigns a score level using descriptors of communicative effectiveness that are delineated in the TSE rating scale (see Appendix B). Examinee scores are produced from the combined average of these independent item ratings. If the two ratings do not show adequate agreement, the tape is rated by a third independent rater. Final scores for tapes requiring third ratings are based on averaging the two closest averages and disregarding the discrepant average. The TSE and SPEAK Band Descriptor Chart (Appendix B) is used by raters.

Scores and score reports

The TSE test yields a single holistic score of communicative language ability reported on a scale of 20 to 60. Assigned score levels are averaged across items and raters, and the scores are reported in increments of five (i.e., 20, 25, 30, 35, 40, 45, 50, 55, and 60). Score level performance is described below.

Scale	Description
60	Communication almost always effective: task performed very competently
55	
50	Communication generally effective: task performed competently
45	
40	Communication somewhat effective: task performed somewhat competently
35	
30	Communication generally not effective: task performed poorly
25	
20	No effective communication: no evidence of ability to perform task

If responses to more than one of the items are missing, no test score is reported and the examinee is offered a retest at no charge.

Two types of score records are issued for the TSE: the examinee's score record, which is sent directly to the examinee, and official score reports, which are sent directly by ETS to institutions or agencies specified by the examinee on the TSE admission ticket. Payment of the test fee entitles the examinee to designate two recipients of the official score report. The official score report includes the examinee's name, registration number, native country, native language, date of birth, test date, and TSE score. (See sample report on page 14.)

Additional score reports

TSE examinees may request that official score reports be sent to additional institutions at any time up to two years after they take the test.

Additional score reports, for which there is a fee, are mailed within two weeks after receipt of the Score Report Request Form found in the TSE *Bulletin*.

Confidentiality of TSE scores

Information retained in the TSE files is the same as the information printed on the examinee's score record and on the official score report. An official score report will be sent only to those institutions or agencies designated on the admission ticket 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.

The scores are not to be released by institutional recipients without the explicit permission of the examinees.

The TSE program recognizes the right of examinees to privacy with regard to information that is stored in data or research files held by Educational Testing Service and the program's responsibility to protect information in its files from unauthorized disclosure. Therefore, ETS does not fax or give TSE results by telephone to examinees or institutions. The TOEFL/TSE office will not release TSE scores or other information without the examinee's written consent.

Score reports are valid only if received directly from Educational Testing Service. TSE 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.



**Test of Spoken English
OFFICIAL SCORE REPORT**

NOTE: If you have any reason to believe that someone has tampered with this score report, please call toll free, 800-257-9547 to have the scores verified. Remember, scores more than two years old cannot be verified. Photostat copies should not be accepted.

TSE SCORE

EXAMINEE'S ADDRESS:

SAMPLE

REGISTRATION NUMBER		NAME (Family or Surname, Given, Middle)	
Month	Year	CENTER NUMBER	Month/Day/Year
TEST DATE			DATE OF BIRTH
INSTITUTION CODE		DEPARTMENT	NATIVE COUNTRY
DEPARTMENT		NATIVE LANGUAGE	

SEE OTHER SIDE FOR EXPLANATION OF SCORES.

Test of Spoken English, P.O. Box 6157, Princeton, NJ 08541-6157, USA

DOs and DON'Ts

DO verify the information on an examinee's score record by calling TOEFL/TSE Services at

1-800-257-9547

(8:30 am – 4:30 pm New York time)

DON'T accept scores that are more than two years old.

DON'T accept score reports from other institutions that were obtained under the SPEAK program. SPEAK scores are only valid for the institution that administered the test.

DON'T accept photocopies of score reports.

Examinee identification service

The examinee identification service provides photo identification of examinees taking the TSE. If there is reason to suspect an inconsistency between a high test score and relatively weak spoken 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 for 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 report.

Requests for photo file records should be sent to:

TOEFL/TSE Program Office
Educational Testing Service
PO Box 6157
Princeton, NJ 08541-6157
USA

Requests for TSE rescoring

An examinee who questions the accuracy of the reported score may request to have the response tape rated again by a rater who did not score the tape previously. If the TSE score increases or decreases, a revised examinee's score record is issued, and revised official score reports are sent to the institutions that received original scores. This revised score becomes the official TSE score. If rescoring confirms the original TSE score, the examinee is so notified by letter from TOEFL/TSE Services-Princeton.

Requests must be received within six months of the test date, and there is a fee for this service. The results of the rescoring are available about three weeks after the receipt at TOEFL/TSE

Services-Princeton of the TSE Rescoring Request Form and fee. The form is available in the TSE *Bulletin*. Experience has shown that very few score changes result from this procedure.

TSE test score data retention

Because language proficiency can change considerably in a relatively short period, TOEFL/TSE Services-Princeton will not report or verify scores that are more than two years old. Individually identifiable test scores are retained for only two years.

TSE test score data that may be used at **any time** for informational, research, statistical, or training purposes are not individually identifiable.

Use of TSE Scores

Setting score standards

Educational Testing Service does not set passing or failing scores on the TSE. Each institution or agency that uses TSE scores must determine what score is acceptable, depending on the level of oral communicative language ability it deems appropriate for a particular purpose. It should be noted that scores on the revised TSE and the original test are different in meaning. Because the tests are different, there cannot be a score-by-score correspondence on the two measures. The TSE program has prepared the *TSE Standard-Setting Kit* to assist institutions and agencies in arriving at score standards for the revised test.

TSE sample response tape

The TSE program has developed a *TSE Sample Response Tape* as a supplement to this guide. The 30-minute audio tape contains selected sample responses from the revised TSE and is intended to provide score users with a better understanding of the levels of communicative effectiveness represented by particular TSE scores. The tape includes several speech samples elicited from nonnative English speakers of different native language backgrounds. The speech samples represent various levels of spoken English proficiency derived from the TSE rating scale and are arranged from high score to low score.

Guidelines for using TSE test scores

The following guidelines are presented to assist institutions in the interpretation and use of TSE scores.

1. Use the TSE score only as a measure of ability to communicate orally in English. Do not use it to predict academic or work performance.
2. Base the evaluation of an applicant's potential for successful academic work or job performance on all available relevant information and recognize that the TSE score is only one indicator of ability to perform effectively in a given academic or professional context.
3. Consider the kinds and levels of English oral language required at different levels of study in different academic disciplines or in varied professional assignments. Also consider the resources available at the institution for improving the English speaking proficiency of nonnative speakers.
4. Consider that examinee scores are based on a 20-minute tape that represents spontaneous speech samples.
5. Review the TSE rating scale and *TSE Sample Response Tape*. The scale appears in Appendix B and the tape can be ordered from ETS.
6. Conduct a local validity study to assure that the TSE scores required by the institution are appropriate.

It is important to base the evaluation of international candidates' potential performance on all available relevant information, not solely on TSE scores. The TSE measures an individual's oral communicative language ability in English in a North American context, but does not measure listening, reading, or writing skills in English. The TOEFL and TWE tests may be used to measure those skills.

General oral communicative effectiveness is only one of many qualities necessary for successful academic or job performance. Other qualities may include command of subject matter, interpersonal skills, and interest in the field or profession. The TSE test does not provide information about aptitude, motivation, command of subject matter or content areas, teaching ability, or cultural adaptability, all of which may have significant bearing on the ability to perform effectively in a given situation.

As part of its general responsibility for the tests it produces, the TSE program is concerned about the interpretation and use of TSE scores by recipient institutions. The TSE Program Office encourages individual institutions to request its assistance with any questions related to the proper use of TSE scores.

Statistical Characteristics of the TSE Test: Performance of Examinees on the Test of Spoken English

This section contains information about the performance of examinees who took the Test of Spoken English between July 1995 and January 2000. The psychometric data were collected during the first five years of the administration of the revised TSE.

Contents

Reliability and SEM	9
Table 1: Average TSE Score Reliabilities and SEMs — Total Group and Subgroups	9
Performance of Examinees on the TSE Test	17
Table 2: Percentile Ranks for TSE Scores — Total Group	18
Table 3: Percentile Ranks for TSE Scores — Academic Examinees	18
Table 4: Percentile Ranks for TSE Scores — Applicants for Professional License	18
Table 5: TSE Total Score Means and Standard Deviations — All Examinees Classified by Geographic Region and Native Language.....	19
Table 6: TSE Total Score Means and Standard Deviations — All Examinees Classified by Geographic Region and Native Country	20

The data presented here are based on TSE test scores obtained by 82,868 examinees between July 1995 and January 2000. It should be noted that this test record database includes both first-time test takers and repeating examinees.

These tables summarize the performance of self-selected groups of examinees who took the TSE test during the period specified; the data are not necessarily representative of the general TSE population.

Table 2 gives the percentile ranks for the total scale scores for the total group between July 1995 and January 2000.

Tables 3 and 4 show the percentile ranks for the total scale scores for the total groups of academic and professional license examinees, as well as for the four largest language groups in each of these categories, between July 1995 and January 2000.

Table 2. Percentile Ranks for TSE Scores — Total Group

(Based on 82,868 examinees who took TSE between July 1995 and January 2000.)

TSE Score	Percentile Rank
60	97
55	90
50	75
45	51
40	24
35	6
30	1
25	<1
20	<1
Score Mean	45.27
S.D.	6.77

Table 3. Percentile Ranks for TSE Scores — Academic Examinees*

TSE Score	Academic Total (36,747)	Chinese (12,093)	Korean (3,608)	Tagalog (2,778)	Hindi (1,530)
60	98	>99	99	98	97
55	91	98	97	91	84
50	76	91	93	73	53
45	53	72	81	41	22
40	25	36	49	12	5
35	6	7	15	1	<1
30	1	1	3	0	<1
25	<1	<1	<1	0	0
20	0	0	<1	0	0
Score Mean	45.04	42.27	40.60	46.64	49.46
S.D.	6.65	5.17	5.67	5.37	5.28

*Based on examinees who, on their TSE answer sheets, indicated that they were teaching or research assistant applicants, or undergraduate or graduate school applicants, to an academic institution between July 1995 and January 2000.

Table 4. Percentile Ranks for TSE Scores — Applicants for Professional License**

TSE Score	Professional Total (46,121)	Tagalog (9,490)	Korean (5,564)	Chinese (2,973)	Arabic (2,440)
60	97	99	>99	99	98
55	89	93	99	96	92
50	75	75	97	87	76
45	50	44	89	65	46
40	23	14	58	31	16
35	6	2	21	6	2
30	1	<1	5	1	<1
25	<1	0	<1	<1	0
20	0	0	0	0	0
Score Mean	45.45	46.18	39.07	43.24	45.99
S.D.	6.85	5.35	4.99	5.65	5.59

**Based on examinees who, on their TSE answer sheets, indicated that they were taking the TSE test to obtain licensure or certification in a professional or occupational field between July 1995 and January 2000.

Tables 5 and 6 may be useful in comparing the performance on the TSE test of a particular examinee with that of other examinees from the same country and with that of examinees who speak the same language. It is important to point out that the data do not permit the generalization that there are fundamental differences in the ability of the various national and language groups to learn English or in the level of English proficiency they can attain. The tables are based simply on the performance of those examinees native to particular countries and languages who happened to take the TSE test.

**Table 5. TSE Total Score Means and Standard Deviations(1) —
All Examinees Classified by Geographic Region and Native Language**
(Based on 82,868 examinees who took TSE between July 1995 and January 2000)⁽²⁾

Native Language		Number of Examinees	Mean	Standard Deviation	Native Language		Number of Examinees	Mean	Standard Deviation	
AFRICAN	Afrikaans	374	56	5	ASIAN (continued)	Sinhalese	101	46	6	
	Amharic	120	44	6		Sundanese	*	*	*	
	Bemba	*	*	*		Tagalog	12,268	46	5	
	Berber	*	*	*		Tamil	1,759	50	6	
	Chichewa	*	*	*		Tatar	*	*	*	
	Efik-Ibibio	*	*	*		Telugu	1,121	48	5	
	Ewe	34	46	7		Thai	464	41	6	
	Fula (Peulh)	*	*	*		Tibetan	*	*	*	
	Ga	*	*	*		Tulu	27	51	6	
	Ganda (Luganda)	*	*	*		Urdu	825	48	6	
	Hausa	*	*	*		Uzbek	*	*	*	
	Ibo (Igbo)	420	46	5		Vietnamese	1,032	41	6	
	Kanuri	*	*	*		EUROPEAN	Albanian	39	46	6
	Kikuyu	72	45	5			Armenian	43	49	6
	Kirundi	*	*	*			Basque (Euskara)	*	*	*
	Lingala	*	*	*			Belarusian	*	*	*
	Luba-Lulua	*	*	*			Bulgarian	173	48	6
	Luo	29	47	5			Catalan (Provençal)	33	45	6
	Malagasy	*	*	*			Czech	105	49	5
	Malinke-Bambara-Dyula	*	*	*			Danish	62	55	5
	Mende	*	*	*			Dutch	444	53	6
	Nyanja	*	*	*			English	2,065	56	6
	Oromo (Galla)	*	*	*			Estonian	*	*	*
	Ruanda	*	*	*			Finnish	143	49	7
	Sesotho	*	*	*			French	1,511	48	7
	Setswana	*	*	*			Galician	*	*	*
	Shona	56	49	5			German	1,412	53	6
	Siswati	*	*	*			Greek	478	49	6
	Somali	34	43	8			Hungarian (Magyar)	223	48	6
	Swahili	76	46	7			Icelandic	28	53	5
	Tigrinya	40	43	5			Italian	395	48	6
	Twi-Fante (Akan)	98	46	6			Latvian	*	*	*
	Wolof	*	*	*			Lithuanian	34	46	8
Xhosa	*	*	*	Macedonian	29		47	6		
Yoruba	454	46	6	Maltese	*		*	*		
Zulu	*	*	*	Norwegian	127		51	7		
ASIAN	Assamese	31	49	7	Polish		1,044	46	5	
	Azeri	*	*	*	Portuguese		878	47	6	
	Bengali	782	49	6	Romanian		526	48	6	
	Bhili	*	*	*	Russian	1,099	46	6		
	Bikol	182	44	5	Serbo-Croatian	477	47	6		
	Burmese	36	47	7	Slovak	77	47	6		
	Cebuano (Visayan)	2,491	46	5	Slovene	*	*	*		
	Chinese	15,066	42	5	Spanish	3,598	46	7		
	Georgian	*	*	*	Swedish	289	53	6		
	Gujarati	1,501	46	6	Turkish	689	46	6		
	Hindi	2,547	49	6	Turkmen	*	*	*		
	Ilocano	960	44	5	Ukrainian	141	47	6		
	Indonesian	240	43	6	Yiddish	*	*	*		
	Japanese	3,080	41	6	Yupiks	*	*	*		
	Javanese	27	41	7	MIDDLE EASTERN	Arabic	3,218	46	6	
	Kannada (Kanarese)	384	50	5		Farsi (Persian)	754	45	6	
	Kashmiri	27	51	7		Hebrew	435	51	6	
	Kazakh	*	*	*		OTHER/NOT REPORTED	Not Reported	573	45	7
	Khmer (Kampuchean)	*	*	*			Other	979	45	6
	Konkani	171	51	5			PACIFIC REGION	Fijian	*	*
	Korean	9,192	40	5	Madurese	31		40	5	
	Kurdish	*	*	*	Marshallese	*		*	*	
	Lao	*	*	*	Minankabau	*		*	*	
	Malay	114	47	7	Pidgin	*		*	*	
	Malayalam	914	46	6	Samoan	*		*	*	
	Marathi	873	49	5	Tongan	*	*	*		
	Mongolian	*	*	*	SOUTH AMERICAN	Guarani	*	*	*	
	Nepali	58	44	7		Quechua	*	*	*	
	Oriya	76	47	6						
	Panay-Hiligaynon	1,139	45	5						
	Pashto	27	49	6						
	Punjabi	772	46	6						
	Samar-Leyte	138	45	5						
Sindhi	118	49	6							

* (1) Because of the unreliability of statistics based on small samples, means are not reported for subgroups of fewer than 25 examinees.
(2) Includes 573 examinees who did not report their native languages and 979 examinees who reported "other" languages.

**Table 6. TSE Total Score Means and Standard Deviations(1) —
All Examinees Classified by Geographic Region and Native Country**

(Based on 80,218 examinees who took TSE from July 1995 and January 2000)⁽²⁾

Geographic Region and Native Country		Number of Examinees	Mean	Standard Deviation	Geographic Region and Native Country		Number of Examinees	Mean	Standard Deviation	
AFRICA	Algeria	36	44	7	ASIA (continued)	Japan	3,133	41	6	
	Angola	*	*	*		Kiribati	*	*	*	
	Benin	*	*	*		Korea (DPR)	46	40	7	
	Botswana	*	*	*		Korea (ROK)	9,150	40	5	
	Burkina Faso	*	*	*		Kyrgyzstan	*	*	*	
	Burundi	*	*	*		Laos	*	*	*	
	Cameroon	52	44	6		Macau	27	44	8	
	Comoros	*	*	*		Malaysia	182	47	7	
	Congo Republic	*	*	*		Mauritius	*	*	*	
	Cote d'Ivoire	*	*	*		Mongolia	*	*	*	
	Egypt	1,712	45	5		Myanmar (Burma)	37	46	8	
	Eritrea	29	43	6		Nepal	52	44	7	
	Ethiopia	143	44	6		Pakistan	783	48	6	
	Gabon	*	*	*		Philippines	17,540	46	5	
	Gambia	*	*	*		Singapore	196	49	7	
	Ghana	175	46	6		Sri Lanka	301	45	6	
	Guinea	*	*	*		Taiwan	2,503	42	5	
	Kenya	199	46	46		Tajikistan	*	*	*	
	Lesotho	*	*	*		Thailand	463	41	6	
	Liberia	*	*	*		Uzbekistan	36	45	5	
	Lybia	34	46	5		Vietnam	1,036	40	6	
	Madagascar	*	*	*		EUROPE	Albania	31	46	6
	Malawi	*	*	*			Andorra	*	*	*
	Mali	*	*	*			Armenia	*	*	*
	Mauritania	*	*	*			Austria	114	52	5
	Morocco	66	44	5			Azores	*	*	*
	Mozambique	*	*	*			Belarus	53	47	6
	Namibia	*	*	*			Belgium	228	50	6
	Nigeria	1,071	46	6			Bosnia/Herzegovina	140	45	6
	Reunion	*	*	*			Bulgaria	173	48	6
	Rwanda	*	*	*			Croatia	90	48	6
	Sao Tome and Principe	*	*	*			Cyprus	126	49	6
	Senegal	*	*	*			Czech Republic	105	49	6
	Seychelles	*	*	*			Denmark	62	55	6
	Sierra Leone	*	*	*			England	136	56	5
	Somalia	33	42	8			Estonia	*	*	*
	South Africa	774	56	5			Finland	151	49	7
	Sudan	97	46	5			Former Yugoslav Rep. of Macedonia	30	47	6
	Swaziland	*	*	*			France	697	47	6
	Tanzania	32	49	8			Georgia	*	*	*
	Togo	*	*	*			Germany	1,133	53	6
	Tunisia	*	*	*			Greece	367	48	6
	Uganda	*	*	*			Hungary	179	48	6
	Zaire (Congo-DRC)	27	51	7			Iceland	29	53	5
	Zambia	*	*	*			Ireland	26	59	3
	Zimbabwe	71	51	6			Italy	386	48	6
	AMERICAS	Anguilla	*	*			*	Kazakhstan	37	46
Argentina		536	46	6	Latvia		64	44	6	
Aruba		*	*	*	Lithuania		38	46	8	
Bahamas		*	*	*	Luxembourg		*	*	*	
Barbados		*	*	*	Malta		*	*	*	
Belize		*	*	*	Moldova		30	46	7	
Bolivia		29	47	8	Monaco		*	*	*	
Brazil		754	47	6	Netherlands	*	*	*		
Canada		1,467	55	7	Northern Ireland	*	*	*		
Chile		154	46	7	Norway	128	51	7		
Colombia		643	46	6	Poland	1,038	46	5		
Costa Rica		75	50	7	Portugal	130	49	6		
Cuba		125	41	6	Romania	554	48	6		
Dominica		*	*	*	Russia	653	47	7		
Dominican Republic (Commonwealth of)		*	*	*	Scotland	*	*	*		
Ecuador		55	46	7	Slovak Republic	69	47	6		
El Salvador		61	45	7	Slovenia	*	*	*		
Grenada		*	*	*	Spain	516	47	6		
Guadeloupe		*	*	*	Sweden	280	53	6		
Guatemala		41	47	6	Switzerland	219	51	6		
Guyana		*	*	*	Turkey	676	46	6		
Haiti		100	42	7	Ukraine	345	46	6		
Honduras		30	47	7	United Kingdom	27	53	6		
Jamaica		31	53	5	Wales	*	*	*		
Maldives		*	*	*	Yugoslavia	301	47	6		
Mexico		480	47	7	MIDDLE EAST	Iran	722	46	6	
Netherlands Antilles		*	*	*		Iraq	320	46	5	
Nicaragua		156	38	6		Israel	489	50	6	
Northern Mariana Islands		*	*	*		Jordan	281	46	6	
Panama		61	45	7		Kuwait	38	47	7	
Paraguay		*	*	*		Lebanon	194	50	7	
Peru		233	44	6		Oman	*	*	*	
Puerto Rico		145	47	7		Saudi Arabia	128	47	6	
St. Vincent and the Grenadines		*	*	*		Syria	364	48	6	
Suriname		*	*	*		United Arab Emirates	*	*	*	
Trinidad and Tobago		60	52	5		Yemen	*	*	*	
United States of America		257	51	8		OTHER/NOT REPORTED	Not Reported	370	46	7
Uruguay		33	47	7	Other		85	46	6	
Venezuela		210	46	7	PACIFIC REGION	American Samoa	*	*	*	
ASIA		Afghanistan	66	45		5	Australia	63	57	6
		Azerbaijan	*	*		*	Fiji	*	*	*
		Bangladesh	247	47		6	Marshall Islands	*	*	*
		Brunei Darussalam	*	*		*	New Caledonia	*	*	*
		Cambodia (Kampuchea)	*	*		*	New Zealand	*	*	*
		China (People's Republic of)	10,493	42		5	Papua New Guinea	*	*	*
		Hong Kong	2,010	44		6	Western Samoa	*	*	*
		India	10,802	48		6				
	Indonesia	256	43	6						

* (1) Because of the unreliability of statistics based on small samples, means are not reported for subgroups of fewer than 25 examinees.

(2) Includes 370 examinees who did not report their country of birth or who reported English as their native language.

Speaking Proficiency English Assessment Kit (SPEAK)

SPEAK

The TSE program offers the Speaking Proficiency English Assessment Kit (SPEAK), which enables institutions to administer at their own convenience retired forms of the TSE test for local evaluation purposes.

SPEAK was developed by the TOEFL program to provide a valid and reliable instrument for assessing the English speaking proficiency of people who are not native speakers of the language. It can be used for selection of those who are employed as teaching assistants or in other capacities. It can also be used by intensive English language programs to place their students at appropriate levels.

SPEAK is available for direct purchase for on-site testing by university-affiliated English language institutes, institutional or agency testing offices, intensive English language programs, government departments, and other organizations serving public or private educational programs. It is important to remember that SPEAK is designed for internal use only.

Although the test design of the TSE and SPEAK is the same, the scores on these two tests are not equivalent because the TSE is administered and scored under standardized conditions. The SPEAK test is administered and scored following standards set by each institution using the test. Consequently, a SPEAK score is valid only in the institution where SPEAK was administered. Additional information about SPEAK is available upon request.

The *TSE Standard-Setting Kit* is available to assist institutions in arriving at score standards for the revised TSE/SPEAK test.

Launched in the early 1980s, SPEAK was revised in 1996. It includes:

- *SPEAK Rater Training Kit* — the kit includes materials for training staff to rate examinees' oral responses and general test administration information.
- *Test Forms* — six SPEAK test forms (A, B, C, D, E, and F) are available in exercise sets. Each form contains 30 test books, one cassette test tape, the rating scale, and a pad of score sheets.
- *Examinee Practice Set* — the set contains 15 identical practice test books and 15 practice test cassettes. The test provided is the disclosed sample TSE test found in *TSE Bulletins* and on the TOEFL Web site, with the audio component delivered via audio cassettes. The materials enable examinees to become familiar with the format of the SPEAK test.

Research

TOEFL 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 broad areas of inquiry, such as test validation, information, reliability, use, construction, implementation, examinee performance, and applied technology. The areas of inquiry for completed research projects are highlighted in the schema on page 26.

Since the studies are usually specific to the TOEFL tests and associated testing programs, most of the actual research work is conducted by Educational Testing Service staff members rather than by outside researchers. Many projects, however, include outside consultants and the cooperation of other institutions, particularly those with programs in the teaching of English as a foreign or second language.

The TOEFL Board supports this ongoing program. The TOEFL Committee of Examiners, an external committee of specialists in linguistics, language testing, or the teaching of English as a

foreign or second language and language research specialists from the academic community, sets guidelines for the scope of the TOEFL research program and reviews and approves TOEFL funded research projects.

Research and related reports

An ongoing series of research studies and activities related to the revised TSE test continues to address issues of importance to the TSE and SPEAK programs, examinees, and score users. As needed, the TSE Committee suggests further TSE or SPEAK research. The results of research studies conducted under the direction of the TOEFL programs are available to the public in published reports.

To date, there are several TSE or SPEAK-related listings in the TOEFL Research Report Series, the TOEFL Technical Report Series, and the TOEFL Monograph Series. Additional projects are in progress and under consideration. When a new research, technical, or monograph report is published, an abstract and ordering information are posted on the TOEFL Web site. The complete list of available research studies can be found at:

<http://www.toefl.org/research/rrpts.html>, and
<http://www.toefl.org/research/rschindx.html>.

Research Reports

RR-4. An Exploration of Speaking Proficiency Measures in the TOEFL Context. Clark and Swinton. October 1979. Describes a three-year study involving the development and experimental administration of test formats and item types aimed at measuring the English-speaking proficiency of non-native speakers; results grouped into a prototype Test of Spoken English.

RR-7. The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings. Clark and Swinton. December 1980. Examines the performance of teaching assistants on the Test of Spoken English in relation to their classroom performance as judged by students; reports that the TSE® test is a valid predictor of oral language proficiency for nonnative English-speaking graduate teaching assistants.

RR-13. The Test of Spoken English as a Measure of Communicative Ability in the Health Professions. Powers and Stansfield. January 1983. Provides results of using a set of procedures for determining standards of language proficiency in testing pharmacists, physicians, veterinarians, and nurses and for validating the use of the TSE test in health-related professions.

RR-18. A Preliminary Study of Raters for the Test of Spoken English. Bejar. February 1985. Examines the scoring patterns of different TSE raters in an effort to develop a method for predicting disagreements; reports that the raters varied in the severity of their ratings but agreed substantially on the ordering of examinees.

RR-36. A Preliminary Study of the Nature of Communicative Competence. Henning and Cascallar. February 1992. Provides information on the comparative contributions of some theory-based communicative competence variables to domains of linguistic, discourse, sociolinguistic, and strategic competencies and investigates these competency domains for their relation to components of language proficiency as assessed by the TOEFL, TWE, and TSE tests.

RR-40. Reliability of the Test of Spoken English Revisited. Boldt. November 1992. Examines effects of scale, section, examinee, and rater as well as the interactions of these factors on the TSE test; offers suggestions for improving reliability.

RR-46. Multimethod Construct Validation of the Test of Spoken English. Boldt and Oltman. December 1993. Uses factor analysis and multidimensional scaling to explore the relationships among TSE subsections and rating dimensions; results show the roles of test section and proficiency scales in determining TSE score variation.

RR-48.* Analysis of Proposed Revisions of the Test of Spoken English. Henning, Schedl, and Suomi. March 1995. Compares a prototype revised TSE with the original version of the test with respect to interrater reliability, frequency of rater discrepancy, component task adequacy, scoring efficacy, and other aspects of validity; results underscore the psychometric quality of the revised TSE.

RR-49. A Study of the Characteristics of the SPEAK Test. Sarwark, Smith, MacCallum, and Cascallar. March 1995. Investigates issues of reliability and validity associated with the original locally administered and scored SPEAK test, the “off-the-shelf” version of the original TSE; results indicate that this version of the SPEAK test is reasonably reliable for local screening and is an appropriate measure of English-speaking proficiency in U.S. instructional settings.

RR-58.* Using Just Noticeable Differences to Interpret Test of Spoken English Scores. Stricker. August 1997. This study explored the value of obtaining a Just Noticeable Difference (JND) — the difference in scores needed before observers discern a difference in examinees’ English proficiency — for the current Test of Spoken English as a means of interpreting scores in practical terms, using college students’ ratings of their international teaching assistants’ English proficiency and adapting classical psychophysical methods. The test’s concurrent validity against these ratings was also appraised. Three estimates of the JND were obtained. They varied considerably in size, but all were substantial when compared with the standard deviation of the TSE scores, the test’s standard error of measurement, and guidelines for the effect size for mean differences. The TSE test correlated moderately with the rating criterion. The JND estimates appear to be meaningful and useful in interpreting the practical significance of TSE scores, and the test has some concurrent validity.

* Studies related to current versions of the TSE and SPEAK tests launched in July 1995 and July 1996, respectively.

RR-63. * **Validating the Revised Test of Spoken English Against a Criterion of Communicative Success.** Powers, Schedl, Wilson-Leung, and Butler. March 1999. A communicative competence orientation was taken to study the validity of test score inferences derived from the current Test of Spoken English. To implement the approach, a sample of undergraduate students, primarily native speakers of English, provided a variety of reactions to, and judgments of, the test responses of a sample of TSE examinees. The TSE scores of these examinees, previously determined by official TSE raters, spanned the full range of TSE score levels. Undergraduate students were selected as “evaluators” because they, more than most other groups, are likely to interact with TSE examinees, many of whom become teaching assistants.

The objective was to determine the degree to which official TSE scores are predictive of listeners’ ability to understand the messages conveyed by TSE examinees. Analyses revealed a strong association between TSE score levels and the judgments, reactions, and understanding of listeners. This finding applied to all TSE tasks and to nearly all of the several different kinds of evaluations made by listeners.

RR-65. * **Monitoring Sources of Variability Within the Test of Spoken English Assessment System.** Myford and Wolfe. June 2000. An analysis of TSE data showed that, for each of two TSE administrations, the examinee proficiency measures were found to be trustworthy in terms of their precision and stability. The standard error of measurement varied across the score distribution, particularly in the tails of the distribution.

The items on the TSE appear to work together; ratings on one item correspond well to ratings on the other items. Consequently, it is appropriate to generate a single summary measure to capture the essence of examinee performance across the 12 items. However, the items differed little in terms of difficulty, thus limiting the instrument’s ability to discriminate among levels of proficiency.

The TSE rating scale functions as a five-point scale, and the scale categories are clearly distinguishable. Raters differed somewhat in the levels of severity they exercised when they rated examinee performances. The vast majority used the scale in a consistent fashion.

* Studies related to current versions of the TSE and SPEAK tests launched in July 1995 and July 1996, respectively.

Technical Reports

TR-15. * **Strengthening the Ties That Bind: Improving the Linking Network in Sparsely Connected Rating Designs.** Myford and Wolfe. August 2000. The purpose of this study was to evaluate the effectiveness of a strategy for linking raters when there are large numbers of raters involved in a scoring session and the overlap among raters is minimal. In sparsely connected rating designs, the number of examinees any given pair of raters has scored in common is very limited. Connections between raters may be weak and tentative at best. The linking strategy employed involved having all raters in a Test of Spoken English scoring session rate a small set of six benchmark audiotapes, in addition to those examinee tapes that each rater scored as part of his or her normal workload. Using output from *Facets* analyses of the rating data, the researchers looked at the effects of embedding blocks of ratings from various smaller sets of these benchmark tapes on key indicators of rating quality. The researchers found that all benchmark sets were effective for establishing at least the minimal connectivity needed in the rating design in order to allow placement of all raters and all examinees on a single scale. When benchmark sets were used, the highest scoring benchmark (i.e., those examinees that scored 50s and 60s across the items) produced the highest quality linking (i.e., the most stable linking). The least consistent benchmark sets (i.e., those that were somewhat harder to rate because an examinee's performance varied across items) tended to provide fairly stable links. The most consistent benchmarks (i.e., those that were somewhat easier to rate because an examinee's performance was similar across items) and middle scoring benchmarks (i.e., those from examinees who scored 30s and 40s across the items) tended to provide less stable linking. Low scoring benchmark sets provided the least stable linking. When a single benchmark tape was used, the highest scoring single tape provided higher quality linking than either the least consistent or most consistent benchmark tape.

* Studies related to current versions of the TSE and SPEAK tests launched in July 1995 and July 1996, respectively.

Monograph Series

MS-7. * **The Revised Test of Spoken English: Discourse Analysis of Native Speaker and Nonnative Speaker Data.** Lazaraton and Wagner. December 1996. Describes a qualitative discourse analysis of native speaker and nonnative speaker responses to the current TSE test; results indicated that the match between intended task functions (as per the content specifications) and the actual functions employed by native speakers was quite close.

MS-9. * **Theoretical Underpinnings of the Test of Spoken English Revision Project.** Douglas and Smith. May 1997. The purpose of this paper is to lay a theoretical foundation for the revisions leading to the current Test of Spoken English. The revision project was undertaken in response to concerns expressed by researchers and score users about the validity of the TSE test and to a request by the TOEFL Committee of Examiners to make the Test of Spoken English more reflective of current thinking on the assessment of oral language skills. The paper first discusses communicative competence as a basis for understanding the nature of language knowledge, and then describes sociolinguistic and discourse factors that influence spoken language performance. Test method characteristics that influence test performance are also discussed, as are types of evidence necessary for establishing reliability and validity of the current TSE test. The paper concludes with a discussion of the implications of the theory for the interpretation of examinee performance with regard to academic and professional contexts of language use.

TOEFL Research Reports, Technical Reports, and Monographs Related to TSE and SPEAK Tests*

AREA	TSE/SPEAK
TEST VALIDATION	
Construct Validity	RR-4, 7, 13, 36, 46, 48,** MS-7,** MS-9**
Face/Content Validity	RR-49
Predictive Validity	RR-7, 13, 49, 63**
Concurrent Validity	RR-4, 7, 48,** 49, 58**
Response Validity	
TEST INFORMATION	
Score Interpretation	RR-36
Underlying Processes	RR-36
Diagnostic Value	
Performance Descriptors	
Reporting/Scaling	RR-48,** 58**
EXAMINEE PERFORMANCE	
Difference Variables	
Language Acquisition/Loss	
Sample Dimensionality	
Person Fit	
TEST USE	
Decisions/Cut Scores	RR-13
Test/Item Bias	
Socio/Pedagogical Impact	
Satisfying Assumptions	
Examinee/User Populations	
TEST CONSTRUCTION	
Format Rationale/Selection	RR-48**
Equating	RR-58**
Item Pretesting/Selection	
Component Length/Weight	RR-48**
TEST IMPLEMENTATION	
Testing Time	
Scoring/Rating	RR-4, 18, 48,** 49, 65,** 66,** TR15**
Practice/Sequence Effects	
TEST RELIABILITY	
Internal Consistency	RR-40
Alternate Forms	
Test-Retest	
Inter-/Intrater	RR-4, 7, 18, 40, 49
APPLIED TECHNOLOGY	
Innovative Formats	
Machine Test Construction	
Computer-Adaptive Testing	
Item Banking	

* Research Reports are identified by their series number preceded by "RR"; Technical Reports are listed by their series number preceded by "TR"; Monographs are preceded by "MS."

**Studies related to current versions of the TSE and SPEAK tests launched in July 1995 and July 1996, respectively.

References

- American Psychological Association. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association, 1999.
- Bejar, I. *A Preliminary Study of Raters for the Test of Spoken English* (TOEFL Research Report 18). Princeton, NJ: Educational Testing Service, 1985.
- Boldt, R. F. *Reliability of the Test of Spoken English Revisited*. (TOEFL Research Report 40). Princeton, NJ: Educational Testing Service, 1992.
- Boldt, R. F., and Oltman, P. *Multimethod Construct Validation of the Test of Spoken English* (TOEFL Research Report 46). Princeton, NJ: Educational Testing Service, 1993.
- Clark, J. L. D., and Swinton, S. S. *An Exploration of Speaking Proficiency Measures in the TOEFL Context* (TOEFL Research Report 4). Princeton, NJ: Educational Testing Service, 1979.
- Clark, J. L. D., and Swinton, S. S. *The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings* (TOEFL Research Report 7). Princeton, NJ: Educational Testing Service, 1980.
- Douglas, D., Murphy, J., and Turner, C. *The St. Petersburg Protocol: An Agenda for a TSE Validity Mosaic* (ETS internal document). Princeton, NJ: Educational Testing Service, 1996.
- Douglas, D., and Smith, J. *Theoretical Underpinnings of the Test of Spoken English Revision Project* (TOEFL Monograph Series 9). Princeton, NJ: Educational Testing Service, 1997.
- Henning G., and Cascallar, E. C. *A Preliminary Study of the Nature of Communicative Competence* (TOEFL Research Report 36). Princeton, NJ: Educational Testing Service, 1992.
- Henning, G., Schedl, M., and Suomi, B. K. *Analysis of Proposed Revisions of the Test of Spoken English*. (TOEFL Research Report 48). Princeton, NJ: Educational Testing Service, 1995.
- Hudson, T. *A Conceptual Validation of the Theory to Test Specification Congruence of the Revised Test of Spoken English* (ETS internal document). Princeton, NJ: Educational Testing Service, 1994.
- Lazaraton, A., and Wagner, S. *The Revised TSE: Discourse Analysis of Native Speaker and Nonnative Speaker Data* (TOEFL Monograph Series 7). Princeton, NJ: Educational Testing Service, 1996.
- Myford, C. M., and Wolfe, E. W. *Monitoring Sources of Variability Within the Test of Spoken English Assessment System*. (TOEFL Research Report 65). Princeton, NJ: Educational Testing Service, 2000.
- Myford, C. M., and Wolfe, E. W. *Strengthening the Ties That Bind: Improving the Linking Network in Sparsely Connected Rating Designs* (TOEFL Technical Report 15). Princeton, NJ: Educational Testing Service, 2000.
- Pike, L. W. *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., and Stansfield, C. W. *The Test of Spoken English as a Measure of Communicative Ability in the Health Professions: Validation and Standard Setting* (TOEFL Research Report 13). Princeton, NJ: Educational Testing Service, 1983.
- Powers, D. E., Schedl, M. A., Wilson-Leung, S., and Butler, F.A. *Validating the Revised Test of Spoken English Against a Criterion of Communicative Success* (TOEFL Research Report 63). Princeton, NJ: Educational Testing Service, 1999.
- Sarwark, S. M., Smith, J., MacCallum, R., and Cascallar, E. C. *A Study of the Characteristics of the SPEAK Test* (TOEFL Research Report 49). Princeton, NJ: Educational Testing Service, 1995.
- Stricker, L. J. *Using Just Noticeable Differences to Interpret Test of Spoken English Scores* (TOEFL Research Report 58). Princeton, NJ: Educational Testing Service, 1997.
- Wainer, H., and Braun, H. I. (Eds.) *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

Appendices

Appendix A

TSE Committee Members (2001-2002)

Richard F. Young, Chair Member	(2000-2003) (1997-2000)	University of Wisconsin-Madison
Tim McNamara	(2001-2004)	University of Melbourne, Australia
James E. Purpura	(1997-2003)	Teachers College at Columbia University
Emma Castillo	(2000-2002)	Philippine Normal University
Barbara Hoekje	(1999-2002)	Drexel University
Marysia Johnson	(2000-2003)	Arizona State University
Julia Delahunty (ex officio)		Middlesex County College
Mark C. Miller (ex officio)		University of Delaware

Former Members (1992-2001)

Frances Butler, Chair	(1992-1994)	University of California-Los Angeles
Dan Douglas, Chair Member	(1994-1997) (1992-1994)	Iowa State University
Miriam Friedman Ben-David	(1992-1996)	Educational Commission for Foreign Medical Graduates (ECFMG)
Richard Cameron	(1999-2000)	University of Illinois-Chicago
Richard Gaughran	(1997-2001)	Comenius University, Slovakia
Frederick L. Jenks	(1994-1997)	Florida State University
Mark Miller	(1992-1994)	University of Delaware
Joseph A. Murphy	(1994-1997)	Nagasaki Junshin Catholic University, Japan
Cynthia L. Myers	(1996-1999)	Iowa State University
Barbara S. Plakans	(1996-1999)	The Ohio State University
Jennifer St. John	(1992-1995)	University of Ottawa, Canada
Jan Smith	(1992-1996)	University of Minnesota
Carolyn E. Turner, Chair Member	(1997-2000) (1995-1997)	McGill University, Canada

Appendix B

TEST OF SPOKEN ENGLISH (TSE) RATING SCALE

Approved by TSE Committee, December 1995

- 60 Communication almost always effective: task performed very competently.**
Functions performed clearly and effectively
Appropriate response to audience/situation
Coherent, with effective use of cohesive devices
Use of linguistic features almost always effective; communication not affected by minor errors
- 50 Communication generally effective: task performed competently.**
Functions generally performed clearly and effectively
Generally appropriate response to audience/situation
Coherent, with some effective use of cohesive devices
Use of linguistic features generally effective; communication generally not affected by errors
- 40 Communication somewhat effective: task performed somewhat competently.**
Functions performed somewhat clearly and effectively
Somewhat appropriate response to audience/situation
Somewhat coherent, with some use of cohesive devices
Use of linguistic features somewhat effective; communication sometimes affected by errors
- 30 Communication generally not effective: task generally performed poorly.**
Functions generally performed unclearly and ineffectively
Generally inappropriate response to audience/situation
Generally incoherent, with little use of cohesive devices
Use of linguistic features generally poor; communication often impeded by major errors
- 20 No effective communication: no evidence of ability to perform task.**
No evidence that functions were performed
No evidence of ability to respond appropriately to audience/situation
Incoherent, with no use of cohesive devices
Use of linguistic features poor; communication ineffective due to major errors



APPENDIX B TSE AND SPEAK BAND DESCRIPTOR CHART

50

30

20

	60	50	40	30	20
Overall features to consider:	<p>Communication almost always effective: task performed very competently</p> <p>Speaker volunteers information freely, with little or no effort, and may go beyond the task by using additional appropriate functions.</p> <ul style="list-style-type: none"> Native-like repair strategies Sophisticated expressions Very strong content Almost no listener effort required 	<p>Communication generally effective: task performed competently</p> <p>Speaker volunteers information, sometimes with effort; usually does not run out of time.</p> <ul style="list-style-type: none"> Linguistic weaknesses may necessitate some repair strategies that may be slightly distracting Expressions sometimes awkward Generally strong content Little listener effort required 	<p>Communication somewhat effective: task performed somewhat competently</p> <p>Speaker responds with effort; sometimes provides limited speech sample and sometimes runs out of time.</p> <ul style="list-style-type: none"> Sometimes excessive, distracting, and ineffective repair strategies used to compensate for linguistic weaknesses (e.g., vocabulary and/or grammar) Adequate content Some listener effort required 	<p>Communication generally not effective: task generally performed poorly</p> <p>Speaker responds with much effort; provides limited speech sample and often runs out of time.</p> <ul style="list-style-type: none"> Repair strategies excessive, very distracting, and ineffective Much listener effort required Difficult to tell if task is fully performed because of linguistic weaknesses, but function can be identified 	<p>No effective communication; no evidence of ability to perform task</p> <p>Extreme speaker effort is evident; speaker may repeat prompt, give up on task, or be silent.</p> <ul style="list-style-type: none"> Attempts to perform task end in failure Only isolated words or phrases intelligible, even with much listener effort Function cannot be identified
	<p>Functions performed clearly and effectively</p>	<p>Functions generally performed clearly and effectively</p>	<p>Functions performed somewhat clearly and effectively</p>	<p>Functions generally performed unclearly and ineffectively</p>	<p>No evidence that functions were performed</p>
Functional competence is the speaker's ability to select functions to reasonably address the task and to select the language needed to carry out the function.	<p>Speaker is highly skillful in selecting language to carry out intended functions that reasonably address the task.</p>	<p>Speaker is able to select language to carry out functions that reasonably address the task.</p>	<p>Speaker may lack skills in selecting language to carry out functions that reasonably address the task.</p>	<p>Speaker often lacks skills in selecting language to carry out functions that reasonably address the task.</p>	<p>Speaker is unable to select language to carry out the functions.</p>
	<p>Appropriate response to audience/situation</p>	<p>Generally appropriate response to audience/situation</p>	<p>Somewhat appropriate task response to audience/situation</p>	<p>Generally inappropriate response to audience/situation</p>	<p>No evidence of ability to respond appropriately to audience/situation</p>
Sociolinguistic competence is the speaker's ability to demonstrate an awareness of audience and situation by selecting language, register (level of formality) and tone, that is appropriate.	<p>Speaker almost always considers register and demonstrates audience awareness.</p> <ul style="list-style-type: none"> Understanding of context, and strength in discourse and linguistic competence, demonstrate sophistication 	<p>Speaker generally considers register and demonstrates sense of audience awareness.</p> <ul style="list-style-type: none"> Occasionally lacks extensive range, variety, and sophistication; response may be slightly unpolished 	<p>Speaker demonstrates some audience awareness, but register is not always considered.</p> <ul style="list-style-type: none"> Lack of linguistic skills that would demonstrate sociolinguistic sophistication 	<p>Speaker usually does not demonstrate audience awareness since register is often not considered.</p> <ul style="list-style-type: none"> Lack of linguistic skills generally masks sociolinguistic skills 	<p>Speaker is unable to demonstrate sociolinguistic skills and fails to acknowledge audience or consider register.</p>
	<p>Coherent, with effective use of cohesive devices</p>	<p>Coherent, with some effective use of cohesive devices</p>	<p>Somewhat coherent, with some use of cohesive devices</p>	<p>Generally incoherent, with little use of cohesive devices</p>	<p>Incoherent, with no use of cohesive devices</p>
Discourse competence is the speaker's ability to develop and organize information in a coherent manner and to make effective use of cohesive devices to help the listener follow the organization of the response.	<p>Response is coherent, with logical organization and clear development.</p> <ul style="list-style-type: none"> Contains enough details to almost always be effective Sophisticated cohesive devices result in smooth connection of ideas 	<p>Response is generally coherent, with generally clear, logical organization, and adequate development.</p> <ul style="list-style-type: none"> Contains enough details to be generally effective Some lack of sophistication in use of cohesive devices may detract from smooth connection of ideas 	<p>Coherence of the response is sometimes affected by lack of development and/or somewhat illogical or unclear organization, sometimes leaving listener confused.</p> <ul style="list-style-type: none"> May lack details Mostly simple cohesive devices are used Somewhat abrupt openings and closures 	<p>Response is often incoherent; loosely organized, and inadequately developed or disjointed, discourse, often leave listener confused.</p> <ul style="list-style-type: none"> Often lacks detail Simple conjunctions used as cohesive devices, if at all Abrupt openings and closures 	<p>Response is incoherent.</p> <ul style="list-style-type: none"> Lack of linguistic competence interferes with listener's ability to assess discourse competence
	<p>Use of linguistic features almost always effective; communication not affected by minor errors</p>	<p>Use of linguistic features generally effective; communication generally not affected by errors</p>	<p>Use of linguistic features somewhat effective; communications sometimes affected by errors</p>	<p>Use of linguistic features generally poor; communication often impeded by major errors</p>	<p>Use of linguistic features poor; communication ineffective due to major errors</p>
Linguistic competence is the effective selection of vocabulary, control of grammatical structures, and accurate pronunciation along with smooth delivery in order to produce intelligible speech.	<ul style="list-style-type: none"> Errors not noticeable Accent not distracting Range in grammatical structures and vocabulary Delivery often has native-like smoothness 	<ul style="list-style-type: none"> Errors not unusual, but rarely major Accent may be slightly distracting Some range in vocabulary and grammatical structures, which may be slightly awkward or inaccurate Delivery generally smooth with some hesitancy and pauses 	<ul style="list-style-type: none"> Minor and major errors present Accent usually distracting Simple structures sometimes accurate, but errors in more complex structures common Limited ranges in vocabulary; some inaccurate word choices Delivery often slow or choppy; hesitancy and pauses common 	<ul style="list-style-type: none"> Limited linguistic control; major errors present Accent very distracting Speech contains numerous sentence fragments and errors in simple structures Frequent inaccurate word choices; generally lack of vocabulary for task completion Delivery almost always plodding, choppy and repetitive; hesitancy and pauses very common 	<ul style="list-style-type: none"> Lack of linguistic control Accent so distracting that few words are intelligible Speech contains mostly sentence fragments, repetition of vocabulary, and simple phrases Delivery so plodding that only few words are produced

Appendix C

GLOSSARY OF TERMS USED IN TSE RATING SCALE

Communication: Recognition by the listener of a speaker's intended meaning.

Effectiveness of communication: The degree to which an intended message is successfully and efficiently conveyed to a listener.

Task: The performance of an appropriate language function in a specified context.

Function: The use of language for an intended purpose (e.g., to apologize, to complain).

Perform competently: To provide a reasonable response to an intended task.

Compensatory strategies: Communication techniques such as paraphrase, examples, synonyms, redundancy, and demonstration to make one's communication more effective or to compensate for language deficiencies.

Coherence: The clear and logical organization of the speaker's utterances.

Cohesive devices: Cohesive components, such as conjunctions and transitional expressions, which tie utterances together and help the listener understand the organization of the response.

Response to audience/situation: The sensitivity of the speaker to the listener and the social situation. Such sensitivity is demonstrated by the speaker's choice of vocabulary, use of idiomatic expression, degree of formality, degree of politeness, speed, volume, and tone of voice.

Accuracy: The degree to which pronunciation, grammar, fluency, and vocabulary approach that of a native speaker who has or is receiving a postsecondary education.

Pronunciation: The production of speech sounds.

Grammar: The linguistic rules for producing phrases and sentences.

Fluency: Smoothly flowing speech.

Vocabulary: Words and expressions that are appropriate for the intended message.

Appendix D

FREQUENTLY ASKED QUESTIONS AND GUIDELINES FOR USING TSE OR SPEAK SCORES^{1, 2}

FAQs:

1. What does the TSE test assess?

TSE scores are a reflection of an examinee's oral communicative language ability on a scale from 20 to 60 (from "No effective communication" to "Communication almost always effective"). Raters evaluate the speech samples and assign score levels using descriptors of communicative effectiveness related to task/function, coherence and use of cohesive devices, appropriateness of response to audience/situation, and linguistic accuracy.

2. How are scores on the TSE test computed?

There are 12 items on the test, and each item receives two independent holistic ratings from trained TSE raters. The 12 scores are averaged across raters and reported in five-point increments (i.e., 20, 25, 30, 35, 40, 45, 50, 55, 60). If the two ratings do not show adequate agreement, the tape is rated by a third independent rater. Final scores for tapes requiring third ratings are based on averaging the two closest averages and disregarding the discrepant average.

3. What are the similarities and differences between the TSE and SPEAK tests?

The TSE and SPEAK tests are similar in content and are both used to evaluate the speaking ability in English of persons whose native language is not English.¹ Both tests are delivered in a semidirect format, which maintains reliability and validity while controlling for some of the subjective variables associated with direct interviewing. However, the two tests differ in that the TSE is a secure test that is administered and scored by ETS; the SPEAK is administered and scored by individual institutions. The SPEAK tests are former (retired) TSE test forms.

4. Can the original TSE/SPEAK and the revised TSE/SPEAK scores be compared or converted?

No, the scores on the two measures are different in meaning because the original and the revised tests are different in content, format, and score design. Since the tests are different, there cannot be a score-by-score correspondence.

5. How can institutions set their cut scores (passing scores)?

The TSE program has prepared the *TSE Standard-Setting Kit* to assist institutions in choosing their cut scores on the TSE. The kit consists of a video that gives basic information about the test, an audiotape with sample responses, and a manual that provides instructions on how to set up and conduct a standard-setting meeting. If you are interested in purchasing this kit, contact the TSE/SPEAK Director for information.

6. What score requirement (passing scores) are most institutions choosing?

It is not advisable for an institution to choose a cut score based on those chosen by other institutions. It is important that each institution determine what cut score is acceptable in its particular context by having a standard-setting meeting as explained in the *TSE Standard-Setting Kit*.

7. Are TSE and SPEAK scores equivalent?

Although the test design of the TSE and SPEAK is the same, the scores on these two tests are not equivalent because the TSE is administered and scored under standardized conditions. The SPEAK test is administered and scored following standards set by each institution using the test. Consequently, a SPEAK score is valid only in the institution where it was administered; it is not valid in any other institution.

¹ It is not valid to use the TSE or SPEAK tests to assess the oral communicative ability of native speakers of English. Because the highest score on the TSE/SPEAK rating scale is 60, it might mistakenly be assumed that only native speakers of English or perfect responses can receive that score. Theoretically, an educated native speaker of English would be capable of scoring well beyond 60, if such a score existed.

² Tests were revised in 1995 and 1996, respectively.

Appendix D (continued)

GUIDELINES FOR USING TSE OR SPEAK TEST SCORES

The following guidelines are presented to assist institutions in the interpretation and use of TSE and/or SPEAK scores:

- 1. Consider that examinee scores are based on a 20-minute test that represents spontaneous speech samples.**
Each set of responses is a snapshot of an examinee's performance under particular conditions; an examinee's performance might vary from day to day, depending on the communicative situation.
- 2. Use TSE or SPEAK scores only as a measure of ability to communicate orally in English. The scores should not be used to predict academic, teaching, or professional performance.**
The evaluation of an examinee's potential for successful academic work, teaching, or professional performance should be based on all available relevant information, including command of subject matter, interpersonal skills, and interest in his or her field or profession. For example, it is recommended that, for ITA (international teaching assistant) assessment, other tests of classroom communication, such as teaching performance tests, be used in addition to the TSE or SPEAK test.
- 3. Set score standards for your institution.**
Each institution that uses TSE or SPEAK must determine what cut score is acceptable in its particular context by conducting a standard-setting meeting.
A TSE Standard-Setting Kit is available to assist institutions in arriving at score standards for the revised TSE/SPEAK test. This kit includes a videotape about the revised TSE, a benchmark tape of sample responses at each score level, and materials that can be duplicated and used at standard-setting meetings. The kit may be ordered by filling out the order form in the *TOEFL Products and Services Catalog* where it is described.
- 4. Consider setting more than one passing score.**
An institution might find it appropriate to choose one passing score for those who are ready to enter a teaching or professional environment immediately, and might choose another score for those who would be accepted into positions on a provisional basis.
- 5. Consider that the levels of English oral communicative ability required in different academic disciplines, levels of study, or professional assignments vary.**
This fact may suggest the need for different standards in different departments or for different purposes.

Appendix E

SAMPLE TSE TEST*

The TSE test is designed to measure proficiency in spoken English. Because spoken language proficiency can be achieved only after a relatively long period of study and much practice, an attempt to study English for the first time shortly before taking the test will not be very helpful.

To help you become familiar with the TSE test, several practice questions are provided below.

ON THE DAY OF THE TEST

On the day of the test, you will be given a test book and asked to listen to and read the general directions before you begin. It is a good idea to become familiar with the directions before the day of the test. The practice questions below are similar but not identical to questions you will find in the actual test. Therefore, responses to these practice questions may not be acceptable on an actual test. During the TSE test your responses will be recorded on tape. It might be helpful to record your practice responses on tape, then listen to hear how your speech actually sounds.

GENERAL DIRECTIONS

In the Test of Spoken English, you will be able to demonstrate how well you speak English. The test will last approximately 20 minutes. You will be asked questions by an interviewer on tape. The questions are printed in the test book and the time you will have to answer each one is printed in parentheses after each question. You are encouraged to answer the questions as completely as possible in the time allowed. While most of the questions on the test may not appear to be directly related to your academic or professional field, each question is designed to tell the raters about your oral language ability. The raters will evaluate how well you communicate in English.

As you speak, your voice will be recorded. Your score for the test will be based on your speech sample. Be sure to speak loudly enough for the machine to record clearly what you say. Do not stop your tape recorder at any time during the test unless you are told to do so by the test supervisor. If you have a problem with your tape recorder, notify the test supervisor immediately.

TSE PRACTICE QUESTIONS**

First, the interviewer will ask you three questions. These questions are for practice and will not be scored, but it is important that you answer them.

Sample questions:

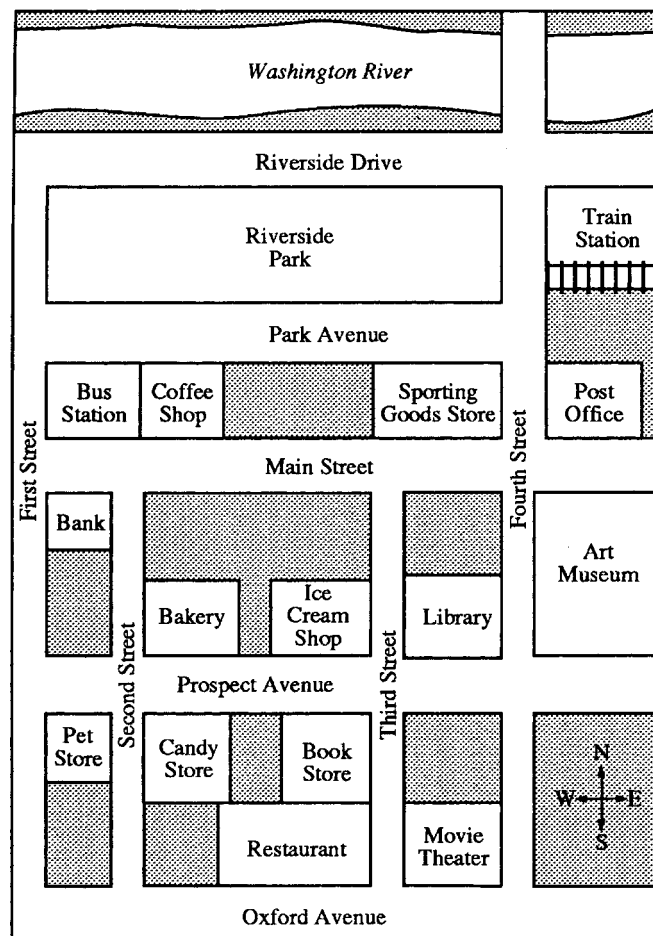
What is the ID number on the cover of your test book? (10 seconds)

What is the weather like today? (10 seconds)

What are your plans for the rest of the day? (10 seconds)

Then the test will begin. Be sure to speak clearly and say as much as you can in responding to each question.

Imagine that we are colleagues. The map below is of a neighboring town that you have suggested I visit. You will have 30 seconds to study the map. Then I'll ask you some questions about it.

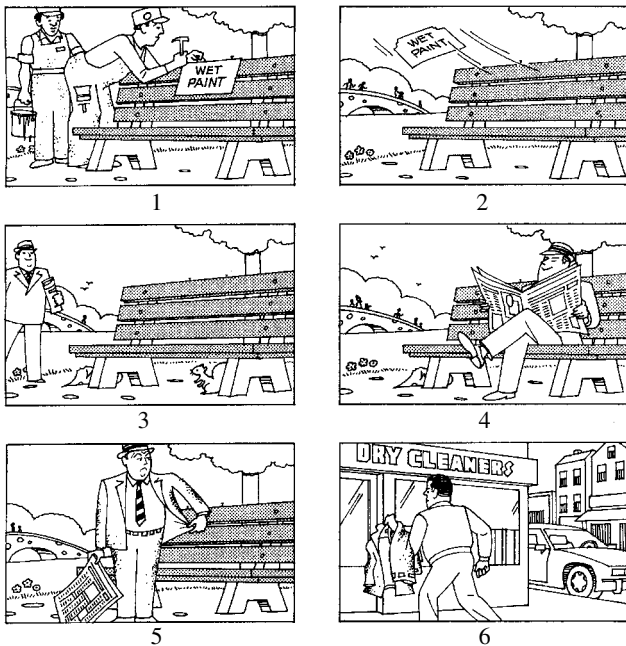


1. Choose one place on the map that you think I should visit and give me some reasons why you recommend this place. (30 seconds)
2. I'd like to see a movie. Please give me directions from the bus station to the movie theater. (30 seconds)
3. One of your favorite movies is playing at the theater. Please tell me about the movie and why you like it. (60 seconds)

* Copies of this sample test are available at <http://www.toefl.org>, or by contacting the TSE program.

** Please note that the graphics used in the TSE practice questions are not the same size as those found in an actual test book.

Now please look at the six pictures below. I'd like you to tell me the story that the pictures show, starting with picture number 1 and going through picture number 6. Please take one minute to look at the pictures and think about the story. Do not begin the story until you are told to do so.

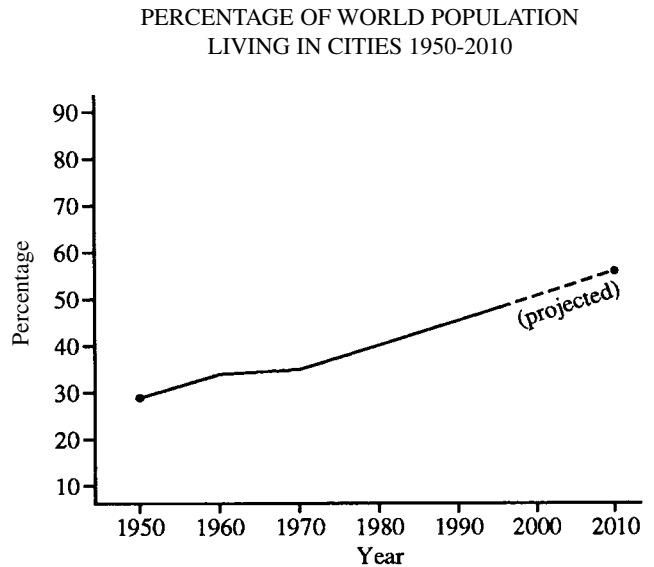


4. Tell me the story that the pictures show. (60 seconds)
5. What could the painters have done to prevent this? (30 seconds)
6. Imagine that this happens to you. After you have taken the suit to the dry cleaners, you find out that you need to wear the suit the next morning. The dry cleaning service usually takes two days. Call the dry cleaners and try to persuade them to have the suit ready later today. (45 seconds)
7. The man in the pictures is reading a newspaper. Both newspapers and television news programs can be good sources of information about current events. What do you think are the advantages and disadvantages of each of these sources? (60 seconds)

Now I'd like to hear your ideas about a variety of topics. Be sure to say as much as you can in responding to each question. After I ask each question, you may take a few seconds to prepare your answer, and then begin speaking when you're ready.

8. Many people enjoy visiting zoos and seeing the animals. Other people believe that animals should not be taken from their natural surroundings and put into zoos. I'd like to know what you think about this issue. (60 seconds)
9. I'm not familiar with your field of study. Select a term used frequently in your field and define it for me. (60 seconds)

10. The graph below presents the actual and projected percentage of the world population living in cities from 1950 to 2010. Tell me about the information given in the graph. (60 seconds)



11. What might this information mean for the future? (45 seconds)
12. Now imagine that you are the president of the Forest City Historical Society. A trip to Washington, D.C. has been organized for the members of the society. At the last meeting you gave out a schedule for the trip, but there have been some changes. You must remind the members about the details of the trip and tell them about the changes indicated on the schedule. In your presentation do not just read the information printed, but present it as if you were talking to a group of people. You will have one minute to plan your presentation. Do not begin speaking until you are told to do so.

**FOREST CITY HISTORICAL SOCIETY
TRIP TO WASHINGTON, D.C.**

Date: Saturday, April 12

Transportation: Chartered Bus

Depart: 8:00 ~~8:30~~ a.m. — Community Center parking lot

Itinerary: 10:30 a.m. — Guided Tour of White House
12:30 p.m. — Lunch* - Rock Creek Park
3:00 p.m. — National Museum of History and Technology (lecture - 4:00 p.m.)
6:30 p.m. — Dinner - ~~Embassy~~ ^{Capital Inn} Restaurant Georgetown

Return: 10:00 p.m. (approximately)

Cost: ~~\$20.00~~ (excluding admissions and dinner)
\$25.00

* Bring your own

(90 seconds)

Where to Get TSE Bulletins

Bulletins are usually available from local colleges and universities. In addition, *Bulletins* are available at many of the locations listed below; at United States educational commissions and foundations, United States Information Service (USIS) offices, binational centers, and private organizations; and directly from Educational Testing Service.

ALGERIA, OMAN, QATAR, SAUDI ARABIA, AND SUDAN

AMIDEAST
Testing Programs
1730 M Street, NW, Suite 1100
Washington, DC 20036-4505, USA
Telephone: 202-776-9649
www.amideast.org

EGYPT

AMIDEAST/CAIRO
23, Mossadak Street
Dokki, Cairo, Egypt
Telephone: 20-2-337-8265
www.amideast.org

or

AMIDEAST
American Cultural Center
3 Pharaana Street
Azarita, Alexandria, Egypt
Telephone: 20-3-482-9091
www.amideast.org

EUROPE, East/West

Citogroup-TOEFL
P.O. Box 1203
6801 BE Arnhem
Netherlands
Email: registration@citogroep.nl
Telephone: 31-26-352-1577
Fax: 31-26-352-1200
www.citogroep.nl

GAZA

AMIDEAST
Ahamad Abd al-Aziz Street
Behind Al-Karmel Secondary School
Remal Quarter
Gaza City
Telephone: 972-8-286-9338
www.amideast.org

HONG KONG

Hong Kong Examinations
Authority
San Po Kong Sub-Office
17 Tseuk Luk Street
San Po Kong
Kowloon, Hong Kong
Telephone: 852-2328-0061, ext. 365
www.hkea.edu.hk

INDIA/BHUTAN

Institute of Psychological and
Educational Measurement
119/25-A Mahatma Gandhi Marg
Allahabad, 211001, U.P. India
Telephone: 91-532-624881
or 624988
www.ipem.org

INDONESIA

International Education Foundation (IEF)
Menara Imperium, 28th Floor, Suite B
Metropolitan Kuningan
Superblok, Kav. 1
Jalan H.R. Rasuna Said
Jakarta 12980
Indonesia
Telephone: 62-21-8317330
www.iie.org/iie/ief/

JAPAN

Council on International Educa-
tional Exchange
TOEFL Division
Cosmos Aoyama B1
5-53-67 Jingumae, Shibuya-ku
Tokyo 150-8355, Japan
Telephone: (813) 5467-5520
www.cieej.or.jp

JERUSALEM

AMIDEAST/West Bank
Al-Watanieh Towers, 1st Floor
34 El-Bireh Municipality Street
El-Bireh, Palestinian National Authority
East Jerusalem 91193
Telephone: 972 or 970-2-240-8023
www.amideast.org

JORDAN

AMIDEAST
1 Akram Rashid, Um As-Summaq
P.O. Box 1249
Amman, 11118 Jordan
Telephone: 962-6-581-0930
www.amideast.org

KOREA

Korean-American Educational
Commission (KAEC)
M.P.O. Box 112
Seoul 121-600, Korea
Telephone: 82-2-3275-4000
www.fulbright.or.kr

KUWAIT

AMIDEAST
Yousef Al-Qenai Street
Bldg. 15, First Floor
Salmiya, Kuwait
Mail: P.O. Box 44818
Hawalli 32063, Kuwait
Telephone: 965-575-0670
www.amideast.org

LEBANON

AMIDEAST
Bazerkan Bldg., Nejme Square
1st Floor
Riad El Solh
Beirut, 2011 3302 Lebanon
Telephone: 961-1-989901
www.amideast.org

MALAYSIA/SINGAPORE

MACCEE
Testing Services
8th Floor Menara John Hancock
Jalan Gelenggang
Damansara Heights
50490 Kuala Lumpur, Malaysia
Telephone: 6-03-253-8107
www.macee.org.my/

MEXICO

Institute of International Education
Londres 16, 2nd Floor
Colonia Juarez, D.F., Mexico
Telephone: 525-209-9100,
ext. 3500, 3510, 4511
www.iie.org/latinamerica/

MOROCCO

AMIDEAST
15, rue Jabal El Ayachi, Agdal
Rabat, Morocco
Telephone: 212-3-767-5081
www.amideast.org

PEOPLE'S REPUBLIC OF CHINA

China International Examinations
Coordination Bureau
167 Haidian Road
Haidian District
Beijing 100080
People's Republic of China
Telephone: 86 (10) 6251-3994
www.neea.edu.cn

SYRIA

AMIDEAST
Ahmed Mrewed Street
Next to Nadi Al Sharq
Nahas Building No. 3
First Floor
Damascus, Syria
Telephone: 963-11-331-4420
www.amideast.org

TAIWAN

The Language Training & Testing
Center
P.O. Box 23-41
Taipei, Taiwan 106
Telephone: (8862) 2362-6045
www.lttc.ntu.edu.tw

TUNISIA

AMIDEAST
22, rue Al Amine Al Abassi
BP 351 Tunis-Cite Jardins 1002
Tunis-Belvedere, Tunisia
Telephone: 216-1-790-559
www.amideast.org

UNITED ARAB EMIRATES

AMIDEAST
c/o Higher Colleges of Technology
Muroor Road (4th Street)
P.O. Box 5464
Abu Dhabi, UAE
Telephone: 971-2-4-45-6720
www.amideast.org

VIETNAM

Institute of International Education
City Gate Building
104 Tran Hung Dao, 5th Floor
Hanoi, Vietnam
Telephone: (844) 822-4093
www.iie.org/iie/vietnam/

YEMEN

AMIDEAST
Algiers St., #66
P.O. Box 15508
Sana'a, Yemen
Telephone: 967-1-206-222
www.amideast.org

COMMONWEALTH OF INDEPENDENT STATES*

Web Site Address and
Telephone Numbers of
ASPRIAL/AKSELS/ACET
Offices
www.actr.org

RUSSIA

P.O. Box 1
Leninsky Prospect 2
Office 530
Russia 117049, Moscow
Moscow – (095) 237-91-16
(095) 247-23-21

Novosibirsk – (3832) 34-42-93
St. Petersburg – (812) 311-45-93
Vladivostok – (4232) 22-37-98
Volgograd – (8442) 36-42-85
Ekaterinburg – (3432) 61-60-34

ARMENIA, Yerevan
(IREX Office)
(8852) 56-14-10

AZERBAIJAN, Baku
(99412) 93-84-88

BELARUS, Minsk
(10-37517) 284-08-52, 284-11-70

GEORGIA, Tbilisi
(10-995-32) 93-28-99, 29-21-06

KAZAKSTAN, Almaty
(3272) 63-30-06, 63-20-56

KYRGYSTAN, Bishkek
(10-996-312) 22-18-82

MOLDOVA, Chisinau
(10-3732) 23-23-89, 24-80-12

TURKMENISTAN, Ashgabat
(993-12) [within NIS (3632)] 39-90-65
39-90-66

UKRAINE

Kharkiv – (38-0572)-45-62-46 (temporary)
(38-0572)-18-56-06

Kyiv – (044) 221-31-92, 224-73-56
Lviv – (0322) 97-11-25

Odesa – (0487) 32-15-16
UZBEKISTAN, Tashkent
(998-712) 56-42-44
(998-71) 152-12-81, 152-12-86

ALL OTHER COUNTRIES AND AREAS

TOEFL Publications
P.O. Box 6154
Princeton, NJ 08541-6154, USA
Telephone: 1-609-771-7100
www.toefl.org

TSE[®]



SPEAK[®]

TOEFL/TSE Services

**Educational Testing Service
P.O. Box 6151
Princeton, NJ 08541-6151
USA**

Phone: 1-609-771-7100

Fax: 1-609-771-7500

E-mail: toefl@ets.org

Web site: <http://www.toefl.org>

TSE[®] and SPEAK[®] are sponsored by the
Test of English as a Foreign Language Program



70606-008659 • S91M25 • Printed in U.S.A.
I.N. 407521