ETS® TOEIC®

SCORE USER GUIDE

TOEIC®

SPEAKING
&
WRITING

TESTS

COMPUTER-DELIVERED TESTING

Foreword

The *Score User Guide for the TOEIC®* Speaking and Writing tests was prepared for testing centers, companies, schools, *TOEIC®* ETS Preferred Network (EPN) members, and others who have an interest in the TOEIC Speaking and Writing tests. This User Guide describes the uses of the TOEIC Speaking and Writing tests and explains how they are administered. *TOEIC®* program staff located around the world are available to explain how to use the TOEIC Speaking and Writing tests most effectively, whether it is to help develop language assessment, training and recruiting policies, or simply to answer questions that arise after reading this guide.

Companies around the world have come to recognize that English-language proficiency is a key to global competitiveness. In today's diverse and fast-paced global economy, the ability to communicate effectively in English—the international language of business—equips your organization with the skills needed to expand your business and achieve sustainable growth.

For further information or to make testing arrangements, please contact your local EPN office. A list of EPN offices can be found on the web at **www.ets.org/toeic**.

Table of Contents

Overview

About ETS

ETS is a not-for-profit organization with a mission to advance quality and equity in education for all people worldwide.

We help teachers teach, students learn, and parents measure the educational and intellectual progress of their children. We do this by:

- Listening to educators, parents and critics
- Learning what students and their institutions need
- Leading in the development of new and innovative products and services

Our Mission: **To advance quality and equity in education** by providing fair and valid assessments, research, and related services. Our products and services measure knowledge and skills, promote learning and educational performance, and support education and professional development for all people worldwide.

Our Vision: **To be recognized as the global leader** in providing fair and valid assessments, research, and related products and services to help individuals, parents, teachers, educational institutions, businesses, governments, countries, states, and school districts, as well as measurement specialists and researchers.

Our Values: **Social responsibility, equity, opportunity, and quality**

We practice these values by listening to educators, parents and critics. We learn what students and the institutions they attend need. We lead in the development of products and services to help teachers teach, students learn, and parents measure the intellectual progress of their children.

TOEIC® Tests

The *TOEIC®* program offers a series of assessments designed to measure and support English-language proficiency across all four communication skills—listening, reading, speaking and writing—and a range of proficiency levels. The *TOEIC®* tests are the most widely used around the world and set the standard for assessing English-language communication skills needed in the workplace. Organizations of all kinds trust *TOEIC®* scores to inform the decisions that matter.

Overview of the *TOEIC*® Speaking and Writing Tests

For over 40 years, companies around the world have used the *TOEIC*® Listening and Reading test for recruitment, placement, hiring, and promotion purposes. The *TOEIC*® Speaking and Writing tests are designed in response to a need in the global workplace to be able to directly measure speaking and writing skills in English. Many global corporations and institutions need information about their employees' speaking and writing skills, as well as their listening and reading skills. The TOEIC Speaking and Writing tests complement the TOEIC Listening and Reading test by providing a direct assessment of a person's spoken and written English skills. The TOEIC Speaking and Writing tests combined with the TOEIC Listening and Reading test provide both test takers and score users with a comprehensive view of a test taker's English language ability. Together, these tests provide an assessment tool for measuring all four English-language communication skills.

Target Population

- Personnel who use English in real-life work settings, such as businesses, hotels, hospitals, restaurants, international meetings, conventions, and sporting events
- Managerial, sales, and technical employees in international business, commerce, and industries that require English for their work
- Government agencies looking to make informed decisions
- Academic institutions preparing for student admissions, placement and exiting, as well as measuring the effectiveness of their English-language training programs
- Individuals who are preparing to enter the workplace
- Candidates for training to be conducted in English
- Individuals in English-language training programs

Purposes

- To verify current level of English proficiency
- To qualify for a new position and/or promotion in a company
- To enhance professional credentials
- To monitor progress in English
- To set learning goals
- To involve employers in advancing English ability

Test Design

In August 2005, ETS test developers met with key partners to discuss the test design requirements for the TOEIC Speaking and Writing tests. These key partners had important information about what kinds of information global corporations and institutions needed about test takers' language skills. As a result of these meetings, the following requirements were established:

- The tests should be linear (i.e., each test form is created to be administered in a certain sequence) and delivered by computer with the possibility of a paper option for writing.

- Each test should discriminate across a wide range of abilities including test takers of relatively low ability (as low as test takers with *TOEIC®* Listening and Reading combined scores of 400).

- Each test should separate test takers into as *meaningful* levels as possible.

- Combined testing time for Speaking and Writing should be approximately 90 minutes.

The test designers used an Evidence-Centered Design (ECD) approach to create the specifications for the test. The methodology can be summarized as follows:

1. Articulate as clearly as possible the claims we wish to make about test takers based on their performance on the assessment.

2. Identify behaviors that would convince us that making the claims described under step 1 above would be warranted.

3. Specify tasks that the test taker will attempt to accomplish in order to show these behaviors and thus to provide evidence to support the claims. (This is different from writing test questions.)

4. Define in as much detail as possible the characteristics of the task that will always be the same and those that can be varied.

5. Enumerate the allowable variations for each task type.

6. Use the output from steps 4 and 5 to create sample tasks.

7. Evaluate the extent to which the sample tasks support the claim that was articulated in step 1. If there are discrepancies between the claim and the sample task, begin again with a different task (step 2) or go back to step 1 and modify the claim.

After going through this process several times, the test designers decided that the most effective strategy for defining meaningful levels of performance was to base the design of the test on three hierarchical claims. For the *TOEIC®* Speaking test, the claims are:

- The test taker can produce some language that is intelligible to native and proficient to nonnative English speakers.

- The test taker can carry out routine social and occupational interactions such as giving and receiving directions, asking for information, asking for clarification, and so forth.

- The test taker can create connected, sustained discourse appropriate to the typical workplace.

For the *TOEIC®* Writing test, the claims are:

- The test taker can produce well-formed sentences (including subordination).

- The test taker can produce multi-sentence-length text to convey straightforward information, questions, instructions, narratives, etc.

- The test taker can produce multi-paragraph-length text to express complex ideas, using, as appropriate, reasons, evidence, and extended explanations.

One of the strengths of the ECD process is its ability to demonstrate a strong and explainable relationship between the claims, the tasks the test requires test takers to do, the evaluation (scoring) of tasks, and the descriptions of overall performance on the test. However, the process also implies that if no external validity evidence is present, the claims made cannot extend beyond evidence that can be deduced from the test itself.

In the case of the *TOEIC®* Speaking and Writing assessments, the test designers felt tasks could be created that supported each of the claims discussed above. It is important to note that each task focuses on one of the three claim levels. It is also important to note that the claims are hierarchical. The assumption for these tests is that, in general, test takers who can successfully do the tasks that support Claim 3 can also successfully do

the tasks that support Claims 1 and 2. Likewise, the test takers who can successfully complete the tasks that support Claim 2 can do the tasks that support Claim 1.

The creation of the levels and the level descriptors was an integral part of the test design process from the beginning. The level descriptors say something about performance on Claim 1 tasks, performance on Claim 2 tasks, and performance on Claim 3 tasks. (The exception to this is Level 1. Because Level 1 test takers did not answer at all, we cannot describe their performance in other than negative terms.)

Of course, adopting a set of test specifications was not the end of the development or design process. Once the draft test specifications had been decided, four Speaking tests and four Writing tests were created. These forms were then administered in a December 2005–January 2006 field study to approximately 2,500 test takers. The tests were then scored using the ETS Online Scoring Network and raters trained for the purpose. The results of the field study were analyzed. The score patterns for all 2,500 test takers were reviewed and the test designers' expectations were met. The test designers worked with statistical analysis to convert weighted raw score totals to scale scores, and then determined how to divide the score scale into performance levels. The levels were set at those places on the scale where changes in score patterns were observed in terms of actual test taker performance and claims about performance.

Timing of the Test

The *TOEIC®* Speaking and Writing tests are timed tests. On the *TOEIC®* Speaking test, test takers have a given amount of time for each test question. Test takers cannot move to the next test question if they finish a question early, skip test questions, or return to previously viewed test questions. On the *TOEIC®* Writing test, test takers have a given amount of time for all of the first five questions on the test. Among those five questions, test takers are able to skip questions, return to previously viewed questions, or spend more time on one question than on another. For the last three questions on the test, test takers have a given amount of time for each test question. On those three questions, test takers cannot move to the next test question if they finish a question early, skip a question, or return to a previously viewed question.

Data from the field study confirmed that the timing of each question is both consistent with the claims of the test and adequate for test takers to be able to respond.

TOEIC® Speaking Test Format

The *TOEIC*® Speaking test is designed to measure a person's ability to communicate in spoken English in the context of daily life and the global workplace. The test is composed of eleven tasks and takes approximately twenty minutes to complete. The range of ability among English learners who will take the TOEIC Speaking test is expected to be broad; that is, both very capable speakers and speakers of limited ability can take the test. The test is designed to provide information about language ability to speakers across a range of language proficiency levels. The following table shows how the tasks in the TOEIC Speaking test are organized:

Question	Task	Evaluation Criteria
1–2	Read a text aloud	• pronunciation • intonation and stress
3–4	Describe a picture	all of the above, plus • grammar • vocabulary • cohesion
5–7	Respond to questions	all of the above, plus • relevance of content • completeness of content
8–10	Respond to questions using information provided	all of the above
11	Express an opinion	all of the above

TOEIC® Writing Test Format

The *TOEIC*® Writing test is designed to measure the ability to use written English to perform communication tasks that are typical to daily life and the international workplace. The test is composed of eight tasks and takes approximately one hour to complete. As with the *TOEIC*® Speaking test, the range of ability among English learners who will take the TOEIC Writing test is expected to be broad. The test is designed to provide useful information about test takers across a range of language proficiency levels. The following table shows how the tasks in the TOEIC Writing test are organized:

Question	Task	Evaluation Criteria
1–5	Write a sentence based on a picture	• grammar • relevance of the sentences to the pictures
6–7	Respond to a written request	• quality and variety of your sentences • vocabulary • organization
8	Write an opinion essay	• whether your opinion is supported with reasons and/or examples • grammar • vocabulary • organization

Administration of the *TOEIC®* Speaking and Writing tests

When and where are the *TOEIC®* Speaking and Writing tests offered?

The *TOEIC®* Speaking and Writing tests are available throughout the world. Testing can be arranged through corporations or other organizations that ask employees, job applicants, or students to take the TOEIC Speaking and Writing tests. In addition, many language-training programs and schools offer TOEIC testing. If testing has available through a particular organization, contact your local ETS Preferred Network (EPN) office to find out when and where the tests can be taken.

How often can the *TOEIC®* Speaking and Writing tests be taken?

Public testing for the *TOEIC®* Speaking and Writing tests is scheduled by the local EPN office. Contact your local EPN office for more information on test dates, locations, and how to register.

Institutional testing is scheduled by the institution or sponsoring organization and can be taken as offered. Test takers should contact the institution or sponsoring organization for more information. Institutions or sponsoring organizations should contact the local EPN office to arrange dates for testing.

Test Security

In administering a worldwide testing program, the *TOEIC®* program and its EPN office consider the maintenance of security at testing centers to be of utmost importance. To offer score users the most valid and reliable measurements of English-language proficiency available, the TOEIC program continuously reviews and refines procedures to increase the security of the test before, during, and after administrations.

Test Center Procedures

The following procedures and regulations apply during the entire test session, which begins when test takers are admitted to the test center and ends when they leave the test center.

- No test taker will be admitted after the test start time has passed.
- Books, dictionaries, papers, notes, rulers, calculators, watch alarms, mobile phones, smart phones, smart watches, listening, recording, scanning or photographic devices, highlighters or aids of any kind are not allowed in the testing room.
- Unauthorized paper of any kind is not permitted in the testing room.
- Test takers must have the test center administrator's permission to leave the testing room. Any lost time cannot be made up.
- There is no scheduled break during the *TOEIC®* test. If a test taker must leave the testing room, they are required to give the test center administrator identification document(s) before leaving the room. It will not be possible to make up the time lost.

Disabilities

The *TOEIC®* program and its local EPN office, in response to requests from individuals with disabilities, will make special arrangements with test center administrators, where local conditions permit, to administer the *TOEIC®* test with accommodations. Among the accommodations that can be provided are extended testing time, large print, and braille versions of the test, or other aids customarily used by the test taker. All requests for accommodations must be approved in accordance with TOEIC program policies and procedures. These procedures are located on the website at **https://www.ets.org/toeic/test-takers/speaking-writing/disability-accommodations/**.

Testing Irregularities

Testing Irregularities refer to problems with the administration of a test. When testing irregularities occur, they may affect an individual or group of test takers. Such problems include, without limitation, administrative errors such as improper timing, improper seating, defective materials, or defective equipment; improper access to test content; and other disruptions of test administrations such as natural disasters and other emergencies. When testing irregularities occur, ETS or the local EPN office may decline to score the test, or may cancel the scores. When in their judgment it is appropriate to do so, ETS or the local EPN office may give affected test takers the opportunity to take the test again as soon as possible without charge.

TOEIC® Speaking and Writing Test Results

Reporting of Test Scores

For score reporting schedules, contact the local EPN office.

Passing the *TOEIC*® Speaking and Writing Tests

The *TOEIC*® Speaking and Writing tests are not the kind of tests that a test taker "passes" or "fails." Not every job or task requires the same level of English proficiency. Because they were developed specifically to meet the needs of the workplace, the TOEIC Speaking and Writing tests measure many levels of ability. The tests enable people to demonstrate what they currently can accomplish in English. The single, continuous scale also makes it possible for learners to set attainable goals and to measure their progress as their English improves.

Many companies use the TOEIC Speaking and Writing tests to set their own cut scores based on the levels of English necessary to carry out particular responsibilities. The test taker's company may require employees to have a minimum TOEIC score due to the corresponding level of English that is needed on the job. Many companies offer English training to help their employees reach target TOEIC Speaking and Writing test scores which reflect specific levels of proficiency based on professional need.

Rating of Speaking and Writing Responses

Speaking

The Speaking test responses are digitally recorded and sent to ETS's Online Network for Evaluation where they are scored by certified ETS raters. The Speaking test includes five different task types. The first four task types (Questions 1–10) are rated on a scale of 0 to 3 and the last task type (Question 11) is rated on a scale of 0 to 5.

Responses for the Read a Text Aloud tasks are scored on overall intelligibility of word- and sentence-level stress. For all other questions on the *TOEIC*® Speaking test, raters evaluate the test taker's abilities in task appropriateness, delivery, and language use. For task appropriateness, raters consider whether the test taker has addressed the task and conveyed the relevant information. They also consider whether the test taker has effectively synthesized the information in the integrated skills tasks. Raters evaluate if the delivery of the response is clear and smooth, and whether it is consistent throughout the response for overall intelligibility. Lastly, raters evaluate the range and accuracy of the test taker's vocabulary and grammar.

Writing

The Writing test responses are also sent to ETS's Online Network for Evaluation and scored by certified ETS raters. The Writing test includes three different task types. The first task type (Questions 1–5) is rated on a scale of 0 to 3, the second task type (Questions 6–7) is rated on a scale of 0 to 4, and the last task type (Question 8) is rated on a scale of 0 to 5.

The responses to the Write a Sentence Based on the Picture tasks are scored on grammar, completion of the task, and relevance of the responses to the pictures. The responses to the Respond to a Written Request tasks are scored on the quality and variety of the sentences, vocabulary, and organization. The response to the Write an Opinion Essay task is scored on grammar, vocabulary, organization, and whether the opinion is supported with reasons and/or examples.

Additional Resources:

The *TOEIC®* program takes measures to ensure scaled score comparability across forms to the best of our ability. *TOEIC®* Compendium papers regarding how ETS scores responses for the *TOEIC®* Speaking and Writing tests, rater reliability, and test form comparability are available at **https://www.ets.org/toeic/research/score-consistency.html** and **https://www.ets.org/toeic/research/validity-fairness.html**.

Information Reported on the Score Certificate

After a test taker has completed the *TOEIC®* Speaking and Writing tests, they will receive a Speaking test score and a Writing test score. Each is reported on a scale of 0 to 200 in increments of 10. There is no combined score. The scores indicate the general proficiency in speaking and writing that a test taker has demonstrated by their performance on the tests. In addition to scaled scores, test takers will receive a description of the English-language abilities typical of test takers scoring at their particular scaled score range. There are 8 possible descriptors for the Speaking test and 9 possible descriptors for the Writing test.

Tests are designed so that the earlier tasks require less proficiency in spoken and written English than the later tasks do. Performance on the higher-level tasks contributes more to the overall score than does a test taker's performance on lower-level tasks.

Score Interpretation

Reported scores for the Speaking Test should not be compared with reported scores for the Writing test. Information about *TOEIC®* Speaking and Writing scores can be found on **https://www.ets.org/toeic/research/score-consistency.html**.

Resources regarding how to interpret TOEIC Speaking and Writing scores can be found at **https://www.ets.org/toeic/research/validity-fairness.html**. Information about the appropriate use of TOEIC scores to make fair and equitable decisions is available at **https://www.ets.org/toeic/research/appropriate-test-use.html**.

If you have any questions about score interpretation, please contact your local EPN office. Your local EPN office can provide information and resources in your local language and provide assistance with standard setting.

Proficiency Descriptors

The proficiency descriptors correspond to a scaled score range and describe the types of general skills and proficiencies in spoken or written English that are common for most people who have achieved a score similar to the test takers. The descriptor associated with the score range that the test taker has achieved will help them understand the strengths and weaknesses of their ability to speak or write in English.

SPEAKING Scaled Score Ranges
0–30
40–50
60–70
80–100
110–120
130–150
160–180
190–200

WRITING Scaled Score Ranges
0–30
40
50–60
70–80
90–100
110–130
140–160
170–190
200

TOEIC® Speaking Test Proficiency Descriptors

The *TOEIC®* Speaking test is designed to give test takers and score users information about the ability of nonnative English speakers to use spoken English in the context of everyday life and the global workplace. The tasks on the test are related to three levels of claims about a test taker's ability:

Questions 1 – 4, Claim 1: These are the easiest task and provides information about a test taker's ability to produce some language that is intelligible to native and proficient nonnative English speakers.

Questions 5 – 10, Claim 2: These tasks are at an intermediate-level of difficulty and provide information about a test taker's ability to carry out routine social and occupational interactions such as giving and receiving directions, asking for information, or asking for clarification.

Question 11, Claim 3: This is the most difficult task and provides information about a test taker's ability to create connected and sustained discourse appropriate to the typical workplace.

These claims are *hierarchical*; in other words, it is assumed that test takers who can create connected and sustained discourse can also carry out routine social interactions and can produce language that is intelligible to proficient nonnative English speakers. Likewise, it is assumed that some test takers who can carry out routine social and occupational interactions and produce intelligible language, may not be able to create connected and sustained discourse. The TOEIC Speaking test proficiency descriptors reflect this hierarchy:

• typical speakers in score ranges 160–180 and 190–200 successfully responded to all questions

• typical speakers in score range 130–150 are inconsistently able to support an opinion or provide a solution using connected and sustained discourse, but are able to carry out routine social and occupational interactions and are able to speak intelligibly

- typical speakers in score range 110–120 have limited ability to support an opinion or provide a solution using connected and sustained discourse, are inconsistently successful at routine interactions, but can produce language that is generally intelligible in familiar contexts

- typical speakers in score ranges 60–70 and 80–100 are generally unable to carry out routine social and occupational interactions, but can produce some language that is generally intelligible in familiar contexts

- typical speakers in score ranges 0–30 and 40–50 are generally not intelligible

The following detailed information regarding the TOEIC Speaking test proficiency descriptors may be helpful for those who use scores to make decisions in institutions, training programs, or companies.

Scale Score Range 190–200

Your scaled score is between 190 and 200. Test takers in this score range typically can create connected, sustained discourse appropriate to the typical workplace. When they express opinions or respond to complicated requests, their speech is highly intelligible. Their use of basic and complex grammar is good and their use of vocabulary is accurate and precise.

Test takers in this score range can also use spoken language to answer questions and give basic information.

Their pronunciation and intonation and stress are at all times highly intelligible.

Scale Score Range 160–180

Your scaled score is between 160 and 180. Test takers in this score range typically can create connected, sustained discourse appropriate to the typical workplace. They can express opinions or respond to complicated requests effectively. In extended responses, some of the following weaknesses may sometimes occur, but they do not interfere with the message:

- minor difficulties with pronunciation, intonation, or hesitation when creating language

- some errors when using complex grammatical structures

- some imprecise vocabulary

Test takers in this score range can also use spoken language to answer questions and give basic information.

When reading aloud, test takers in this score range are highly intelligible.

Scale Score Range 130–150

Your scaled score is between 130 and 150. Test takers in this score range typically are able to create a relevant response when asked to express an opinion or respond to a complicated request. However, at least part of the time, the reasons for, or explanations of, the opinion are unclear to a listener. This may be because of the following:

- unclear pronunciation or inappropriate intonation or stress when the speaker must create language

- mistakes in grammar

- a limited range of vocabulary

Most of the time, test takers in this score range can answer questions and give basic information. However, sometimes their responses are difficult to understand or interpret.

When reading aloud, test takers in this score range are intelligible.

Scale Score Range 110–120

Your scaled score is between 110 and 120. Test takers in this score range typically have limited success at expressing an opinion or responding to a complicated request. Responses include problems such as:

- language that is inaccurate, vague, or repetitive
- minimal or no awareness of audience
- long pauses and frequent hesitations
- limited expression of ideas and connections between ideas
- limited vocabulary

Most of the time, test takers in this score range can answer questions and give basic information. However, sometimes their responses are difficult to understand or interpret.

When reading aloud, test takers in this score range are generally intelligible. However, when creating language, their pronunciation, intonation, and stress may be inconsistent.

Scale Score Range 80–100

Your scaled score is between 80 and 100. Test takers in this score range typically are unsuccessful when attempting to explain an opinion or respond to a complicated request. The response may be limited to a single sentence or part of a sentence. Other problems may include:

- severely limited language use
- minimal or no audience awareness
- consistent pronunciation, stress, and intonation difficulties
- long pauses and frequent hesitations
- severely limited vocabulary

Most of the time, test takers in this score range cannot answer questions or give basic information.

When reading aloud, test takers in this score range vary in intelligibility. However, when they are creating language, they usually have problems with pronunciation and intonation and stress. For more information, check the "Read Aloud Pronunciation and Intonation and Stress ratings."

Scale Score Range 60–70

Your scaled score is between 60 and 70. Test takers in this score range typically can, with some difficulty, state an opinion, but they cannot support the opinion. Any response to a complicated request is severely limited. Most of the time, test takers in this score range cannot answer questions and give basic information. Typically, they have insufficient vocabulary or grammar to create simple descriptions.

When reading aloud, speakers in this score range may be difficult to understand. For more information, check the "Read Aloud Pronunciation and Intonation and Stress ratings."

Scale Score Range 40–50

Your scaled score is between 40 and 50. Test takers in this score range typically cannot state an opinion or support it. They either do not respond to complicated requests or the response is not at all relevant. In routine social and occupational interactions such as answering questions and giving basic information, test takers in this score range are difficult to understand.

When reading aloud, speakers in this score range may be difficult to understand. For more information, check the "Read Aloud Pronunciation and Intonation and Stress ratings."

Scale Score Range 0–30

Your scaled score is between 0 and 30. Test takers in this score range typically left a significant part of the *TOEIC®* Speaking test unanswered. Test takers in this score range may not have the listening or reading skills in English necessary to understand the test directions or the content of the test questions.

Pronunciation Levels

Low

When reading aloud your pronunciation is not generally intelligible.

Medium

When reading aloud your pronunciation is generally intelligible with some lapses.

High

When reading aloud your pronunciation is highly intelligible.

Intonation and Stress Levels

(Intonation and Stress refer to your ability to use emphases, pauses, and rising and falling pitch to convey meaning to a listener.)

Low

When reading aloud your use of intonation and stress is generally not effective.

Medium

When reading aloud your use of intonation and stress is generally effective.

High

When reading aloud your use of intonation and stress is highly effective.

TOEIC® Writing Test Proficiency Descriptors

The *TOEIC®* Writing test is designed to give test takers and score-users information about the ability of nonnative English speakers to use written English in the context of everyday life and the global workplace. The tasks on the test are related to three levels of claims about a test taker's ability:

Questions 1 – 5, Claim 1: These are the easiest tasks and provide information about a test taker's ability to produce well-formed sentences (including subordination).

Questions 6 and 7, Claim 2: These tasks are at an intermediate-level of difficulty and provide information about a test taker's ability to produce multi-sentence-length text to convey straightforward information, questions, instructions, narratives, etc.

Question 8, Claim 3: This is the most difficult task and provides information about a test taker's ability to produce multi-paragraph-length text to express complex ideas, using, as appropriate, reasons, evidence, and extended explanations.

These claims are *hierarchical*; in other words, it is assumed that test takers who can produce multi-paragraph-length text to express complex ideas, can also convey straightforward information, and can produce well-formed sentences. Likewise, it is assumed that some test takers who can convey straightforward information and produce well-formed sentences, cannot create extended texts to support opinions. The TOEIC Writing test proficiency descriptors reflect this hierarchy:

- typical writers in score range 170–190 or with scaled score 200 successfully responded to all questions
- typical writers in score range 140–160 show weaknesses when writing an extended opinion, but are successful when conveying straightforward information, and can write sentences
- typical writers in score range 90–100 and 110–130 are not strong at supporting an opinion, but can usually convey straightforward information, and can write sentences
- typical writers in score range 50–60 and 70–80 cannot write an extended opinion, are mostly unsuccessful at conveying information, and show some weakness at sentence formation
- typical writers in score range 0–30 or with scaled score 40 have almost no ability to create English sentences and thus are unable to convey information or convey straightforward information or express an opinion

The following detailed information regarding the TOEIC Writing test proficiency descriptors may be helpful for those who use scores to make decisions in institutions, training programs, or companies.

Scale Score Range 200

Your scaled score is 200. Test takers with this scaled score typically can communicate straightforward information effectively and use reasons, examples, or explanations to support an opinion. When using reasons, examples, or explanations to support an opinion, their writing is well-organized and well-developed. The use of English is natural, with a variety of sentence structures and appropriate word choice, and is grammatically accurate. When giving straightforward information, asking questions, giving instructions, or making requests, their writing is clear, coherent, and effective.

Scale Score Range 170–190

Your scaled score is between 170 and 190. Test takers in this score range typically can communicate straightforward information effectively and use reasons, examples, or explanations to support an opinion. When giving straightforward information, asking questions, giving instructions, or making requests, their writing is clear, coherent, and effective. When using reasons, examples, or explanations to support an opinion, their writing is generally good. It is generally well-organized and uses a variety of sentence structures and appropriate vocabulary. It may also include one of the following weaknesses:

- occasional unnecessary repetition of ideas or unclear connections between ideas
- noticeable minor grammatical mistakes or incorrect word choices

Scale Score Range 140–160

Your scaled score is between 140 and 160. Test takers in this score range typically can effectively give straightforward information, ask questions, give instructions, or make requests but are only partially successful when using reasons, examples, or explanations to support an opinion. When attempting to explain an opinion, their writing presents relevant ideas and some support.

Typical weaknesses include:

- not enough specific support and development for the main points
- unclear connections between the points that are made
- grammatical mistakes or incorrect word choices

When giving straightforward information, asking questions, giving instructions, or making requests, their writing is clear, coherent, and effective.

Scale Score Range 110–130

Your scaled score is between 110 and 130. Test takers in this score range typically are partially successful when giving straightforward information or supporting an opinion with reasons, examples, or explanations. When giving straightforward information, asking questions, giving instructions, or making requests, the message omits important information or is partly unintelligible. When attempting to explain an opinion, their writing presents relevant ideas and some support.

Typical weaknesses include:

- not providing enough specific support and development for the main points
- unclear connections between the points that are made
- grammatical mistakes or incorrect word choices

Scale Score Range 90–100

Your scaled score is between 90 and 100. Test takers in this score range typically are at least partially successful when giving straightforward information. However, when supporting an opinion with reasons, examples, or explanations, they are mostly unsuccessful.

When giving straightforward information, asking questions, giving instructions, or making requests, the message omits important information or is partly unintelligible.

When attempting to explain an opinion, significant weaknesses that interfere with communication occur, such as:

- not providing enough examples, explanations, or details to support the opinion or they are inappropriate
- inadequate organization or connection of ideas
- limited development of ideas
- serious grammatical mistakes or incorrect word choices

Scale Score Range 70–80

Your scaled score is between 70 and 80. Test takers in this score range typically have some developing ability to express an opinion and to give straightforward information. However, communication is limited.

When attempting to explain an opinion, significant weaknesses that interfere with communication occur, such as:

- not providing enough examples, explanations, or details to support the opinion or they are inappropriate
- inadequate organization or connection of ideas
- limited development of ideas
- serious grammatical mistakes or incorrect word choices

When giving straightforward information, asking questions, giving instructions, or making requests, the responses do not successfully complete the task because of:

- missing information
- missing or obscure connections between sentences and/or
- many grammatical mistakes or incorrect word choices

Test takers in this score range have some ability to produce grammatically correct sentences, but they are inconsistent.

Scale Score Range 50–60

Your scaled score is between 50 and 60. Test takers in this score range typically have limited ability to express an opinion and to give straightforward information. When attempting to explain an opinion, the responses show one of the following serious flaws:

- serious disorganization or underdevelopment of ideas
- little or no detail or irrelevant specifics
- serious and frequent grammatical mistakes or incorrect word choices

When giving straightforward information, asking questions, giving instructions, or making requests, the responses do not successfully complete the task because of:

- missing information
- missing or obscure connections between sentences and/or
- many grammatical mistakes or incorrect word choices

Test takers in this score range have some ability to produce grammatically correct sentences, but they are inconsistent.

Scale Score Range 40

Your scaled score is 40. Test takers with this scaled score typically have only very limited ability to express an opinion and give straightforward information. When attempting to explain an opinion, the responses show one of the following serious flaws:

- serious disorganization or underdevelopment of ideas
- little or no detail or irrelevant specifics
- serious and frequent grammatical mistakes or incorrect word choices

At level 2, test takers cannot give straightforward information. Typical weaknesses at this level include:

- not including any of the important information
- missing or obscure connections between ideas
- frequent grammatical mistakes or incorrect word choices

Test takers with this scaled score are unable to produce grammatically correct sentences.

Scale Score Range 0–30

Your scaled score is between 0 and 30. Test takers in this score range typically left part or parts of the *TOEIC®* Writing test unanswered. Test takers in this score range may need to improve their reading ability in order to understand the test directions and the content of test questions.

Statistical Characteristics of the *TOEIC*® Speaking and Writing Tests

Reliabilities

Reliability is a measure of test score consistency across different forms of a test. The reliability of a test ranges from .00 (no reliability) to 1.00 (perfect reliability). Tests do not have perfect reliability because test takers perform differently on one occasion or test form than on another for reasons that may or may not be related to the purpose of the test. For example, a test taker may be more familiar with the content of questions on one test form than on another form. The error produced by these inconsistencies is called error of measurement.

The most direct way to evaluate reliability is to calculate the correlation between the scores of test takers responding to two different test forms on one day. Because no one takes more than one form a day, test reliability is estimated using data from a single administration or a single form (e.g., internal consistency reliability). This type of reliability estimation method reflects the consistency of test taker performance on individual questions within each test form. Based on data from June 2022 to September 2022 administrations in the Public or Institutional Testing Program, the internal consistency reliability of the Speaking test is around 0.80. The internal consistency reliability of the Writing test is around 0.79.

Standard Error of Measurement

As indicated earlier, error of measurement occurs when a test taker performs differently on one occasion or test form than on another for reasons that may or may not be related to the purpose of the test. Although it is never possible to determine the exact error of measurement in a given score, the standard error of measurement (SEM) can be used to estimate the expected variation of each test taker's observed scores around their "true score." The true score can be thought of as the average score a test taker would earn on a very large number of forms. Based on the same dataset used for reliability estimates, the SEM is approximately 11 scale score points for the Speaking test, and 15 scale score points for the Writing test.

A test taker's Speaking score would be expected to fall between 11 scale score points above and 11 scale score points below their true Speaking test score 68% of the time. A test taker's Writing score would be expected to fall between 15 scale score points above and 15 scale score points below their true Writing test score 68% of the time. Another way of interpreting the SEM is that approximately 68% of test takers receiving scale scores of 100 for the Speaking test will have true scores between 89 and 111. Similarly, because the SEM for the Writing test is 15, approximately 68% of test takers receiving scale scores of 100 for the Writing test will have true scores between 85 and 115.

Correlation Between Scores

There is a relationship between Speaking and Writing test scores. Based on the data from June 2022 to September 2022 administrations in the Public or Institutional Testing Program, the correlation between Speaking and Writing scale scores is around 0.72.

Due to the error of measurement associated with the scale scores for Speaking and Writing tests, the observed correlation is usually underestimated (attenuated). The "disattenuated correlation" between two measures provides an estimate of the correlation between the two measures if there was no error of measurement. Based on the test reliability of 0.80 for Speaking and 0.79 for Writing as reported above, thedisattenuated correlation is around 0.91, which indicates that the tests measure somewhat different, but related skills.

Mapping *TOEIC*® Speaking & Writing Scores to the Common European Framework of Reference

The Common European Framework of Reference for Languages (CEFR) is a structure used to describe the language proficiency of foreign-language learners across Europe. It was designed to facilitate communication in the field of language development by providing a common basis for describing the skills needed to reach various levels of language proficiency. The CEFR is widely used by teachers, curriculum designers, and testing organizations in their efforts to promote language competence, and it is gaining acceptance as the standard for grading the language proficiency of second language learners. Its six levels are clustered in three bands:

A1 - A2 (Basic User)

B1 - B2 (Independent User)

C1 - C2 (Proficient User)

In 2006, ETS assembled a panel of 22 English-language experts representing 10 European countries, who were asked to "map" scores from the (then) new *TOEIC*® Speaking and Writing tests onto the six levels of the CEFR. That is, they were asked to judge the English skills necessary to reach each of the specific levels on the Framework, as demonstrated by *TOEIC*® test scores. Linkages were determined by means of expert judgment using recognized standard-setting procedures: in particular, the [modified] test taker response selection method for constructed-response questions (see, Brandon, 2004; Hambleton, Jaeger, Plake, & Mills, 2000; Hambleton & Pitoniak, 2006). A detailed description of the study methods, as well as the detailed results, can be found in Tannenbaum & Wylie (2008).

In short, the study identified the minimum *TOEIC*® Speaking and Writing scores required for learners to be categorized in each of the CEFR levels, as shown in the table below.

Test Section	A1	A2	B1	B2	C1	C2
TOEIC Writing	30	70	120	150	200	—
TOEIC Speaking	50	90	120	160	200	—

Table entries are minimum scores needed in order to be categorized in the level specified. Entries are missing for the C2 level because the experts involved in the study did not believe that the *TOEIC*® Speaking and Writing sections were challenging enough to recommend placing test takers in the highest (C2) level.

It is not surprising perhaps that not all levels of the CEFR could be linked to TOEIC Speaking and Writing scores. Indeed, a test that adequately covered ALL of the CEFR levels would need to be quite long in order to enable reliable classifications at each of the six distinctly different CEFR levels of proficiency. Therefore, any single test that claims to have established *reliable* linkages to all CEFR levels should be viewed cautiously, and the evidence supporting its claims should be scrutinized carefully.

Institutions, employers, English programs, and learners can interpret the recommended minimum scores on the TOEIC Speaking and Writing measures by referring to the global descriptors and other tables in the CEFR: **http://www.coe.int/t/dg4/linguistic/CADRE_EN.asp**.

Although the mapping of TOEIC Speaking and Writing scores to CEFR levels may be useful in helping to interpret the meaning of scores, it is important to remember the following: Although the results represent the recommendations of informed experts using recognized standard setting approaches, they should be considered guidelines only, not absolute, unequivocal standards: Neither test scores nor the judgments of experts are perfectly reliable. Other independent evidence to support the results of this study is desirable. This evidence might include, for example, comparisons of performance on TOEIC Speaking and Writing with test taker self-ratings of language proficiency and with instructor ratings of test taker language proficiency to confirm the test score levels recommended by experts.

Validity Studies

One kind of evidence that has proven useful in establishing the meaning, or validity, of *TOEIC*® scores has come from test takers themselves in the form of self-assessments of their own language skills (for example, Powers, Kim, and Weng, 2008). There is a good deal of very solid research on the use of self-assessments, and it comes from a variety of fields – higher education (Falchikov & Boud, 1989), personality research (Ackerman, 2002), occupational psychology (Mabe & West, 1982), and language learning (Ross,1998), to name but a few. Self-assessments have been shown to be valid in a variety of contexts, especially in the assessment of language skills. In this regard, Upshur (1975) noted that language learners often have more complete access to the full spectrum of their successes and failures than do third-party assessors, who may hold a much narrower view of an individual's language skills. Similarly, Shrauger and Osberg (1981) noted that people are often active observers of their own behavior, and they often have extensive data on which to base their judgments (often much more than do external evaluators). But just as self-assessments enjoy considerable justification for their use, clearly they also have a number of potentially serious limitations. Chief among them, of course, is that people don't always perceive themselves objectively. They may tend to exaggerate their skills and abilities when reporting them—either intentionally or unintentionally—or they may be inclined to present themselves in socially desirable ways. It is probably unwise therefore to trust the results of self-assessments unconditionally. Nonetheless, we've felt that, all in all, self-assessments constitute a useful criterion against which to judge the validity of TOEIC scores.

In the fall of 2008, the *TOEIC*® Speaking test was taken by approximately 3,500 test takers in Korea and Japan. Approximately 1,500 test takers also took the new *TOEIC*® Writing test. These test takers also completed a self-assessment questionnaire, administered in their first language, which requested self-evaluations of their ability to perform (either easily, with little difficulty, with some difficulty, with great difficulty, or not at all) each of a variety of practical, everyday language tasks in English. Moderately strong relationships of TOEIC scores with test takers' self-reports were noted for both the speaking and writing language domains. TOEIC Speaking scores and self-assessments of the ability to perform speaking tasks correlated reasonably strongly (r = .54), as did TOEIC Writing scores and reports of the ability to perform everyday writing tasks (r = .52). As a measure of the relationship between two variables, these correlations (which can range from -1.00 to +1.00) are meaningful practically by virtue of several commonly used metrics. For instance, by conventional standards, correlations of .50 and above in social science research have been considered to be "large" (Cohen, 1988). Moreover, the correlations observed here compare very favorably with those typically observed in validity studies that have used other kinds of criteria such as course grades, faculty ratings, and degree completion. For example, in a recent large-scale meta-analysis by Kuncel and Hezlett (2007), the most predictable of several criteria considered was first-year grade average. First year grades correlated, on average, about .45 with a variety of different kinds of tests.

Equally important, for each of the language tasks, higher-performing *TOEIC*® test takers were much more likely than lower-scoring test takers to report that they could perform the task. Test taker assessments for selected, representative tasks are shown in the following tables for each test section. (Complete details for all 36 speaking tasks and all 29 writing tasks are available in Powers, Kim, Yu, Weng, and VanWinkle, 2009).

More recently, researchers investigated whether *TOEIC*® Speaking test scores were good predictors of test takers' communicative effectiveness (Schmidgall & Powers, 2021) and whether *TOEIC*® Writing test scores were good predictors of the "functional adequacy" of test takers' writing skills (Schmidgall & Powers, 2020), as judged by professionals in the international workplace. For each study, professionals in 10 countries participated in a survey in which they evaluated the communicative effectiveness of test takers (or functional adequacy of test-takers' writing) in different communicative scenarios. The results provide evidence that TOEIC Speaking and Writing test scores are strong predictors of professionals' judgments of communicative effectiveness.

Table 1: Percentages of TOEIC Test Takers, by Speaking-Score Level, Who Indicated They Could Perform Various English-Speaking Tasks Easily or With Little Difficulty

I can:	Score Level						
	0–50	60–70	80–100	110–120	130–150	160–180	190–200
Make/change/cancel an appointment to see a person	19	32	43	65	78	91	100
Telephone a company to place (or follow-up) an order for an item	16	22	34	56	67	83	96
Have "small talk" with a guest about topics of general interest (e.g., the weather) before discussing business	10	24	35	57	69	83	94

Table 2: Percentages of TOEIC Test Takers, by Writing-Score Level, Who Indicated They Could Perform Various English-Writing Tasks Easily or With Little Difficulty

I can:	Score Level					
	0–80	90–100	110–130	140–160	170–190	200
Write an email requesting information about hotel accommodations	12	37	48	70	81	91`
Write clear directions on how to get to my office	14	23	37	58	69	86
Translate documents (e.g., business letters, manuals) into English	11	18	24	39	54	81

References

Ackerman, P.L. (2002). What do we really know about our abilities and knowledge. *Personality and Individual Differences*, *33*, 587–605.

Brandon, P.R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education*, *17*, 59–88.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*, 395–430.

Hambleton, R.K., & Pitoniak, M.J. (2006). Setting performance standards. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport, CT: Praeger Publishers.

Hambleton, R.K., Jaeger, R.M., Plake, B.S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, *24*, 355–366.

Kuncel, N.R., & Hezlett, S.A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080.

Mabe, P.A., & West, S.G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*, 280–296.

Powers, D.E., Kim, H.-J., & Weng, V. (2008). *The redesigned TOEIC® (listening and reading) test: Relations to test-taker perceptions of proficiency in English*. (ETS Research Rep. RR-08-56). Princeton, NJ: ETS.

Powers, D.E., Kim, H.-J., Yu, F., Weng, V., & VanWinkle, W. (2009). *The TOEIC® speaking and writing tests: Relations to test-taker perceptions of proficiency in English*. (ETS Research Rep. RR-09-18). Princeton, NJ: ETS.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, *15*, 1–20.

Schmidgall, J., & Powers, D.E. (2020). TOEIC Writing test scores as indicators of the functional adequacy of writing in the international workplace: Evaluation by linguistic laypersons. *Assessing Writing, 46*.

Schmidgall, J., & Powers, D.E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC Speaking test scores from linguistic laypersons. *Language Testing, 38*(2), 302–325.

Shrauger, J.S., & Osberg, T.M. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin*, *90*, 322–351.

Tannenbaum, R., & Wiley, E.C. (2008). *Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology*. (ETS Research Rep. RR-08-34, TOEFL iBT Research Report TOEFL iBT-08). Princeton, NJ: ETS.

Upshur, J. (1975). Objective evaluation of oral proficiency in the ESOL classroom. In L. Palmer & B. Spolsky (Eds.), *Papers on language testing 1967–1974* (pp. 53–65). Washington, DC: TESOL.

Score Validity

Test takers are likely to achieve improved scores over time if, between tests, they study English or increase their exposure to native spoken English. Thus, improvement in scores may not indicate an irregularity in the test itself or its administration. However, *TOEIC®* Speaking and Writing test score users that note such inconsistencies as high TOEIC Speaking and Writing test scores and apparent weak English proficiency should refer to the photo on the official score certificate for confirmation of the test takers identification. Score users should notify the local EPN office if there is any such evidence or they believe the scores are questionable for other reasons.

Apparent irregularities reported by score users or brought to the attention of the EPN office by test takers or test administrators who believe that misconduct has taken place, are investigated. Such reports are reviewed, statistical analyses are conducted, and scores may be canceled by ETS or the EPN office as a result.

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

Test Fairness

ETS is committed to ensuring that our tests and test scores are of the highest quality and as free of bias as possible. All of ETS's products and services—including individual test questions, assessments, instructional materials, and publications—are evaluated during development to ensure that they

- are not offensive or controversial
- do not reinforce stereotypical views of any group
- are free of racial, ethnic, gender, socioeconomic, and other forms of bias
- are free of content believed to be inappropriate or derogatory toward any group

All of ETS's tests and other products undergo rigorous, formal reviews to ensure that they adhere to the fairness guidelines, which can be found on the web at **www.ets.org**:

- ETS Standards for Quality and Fairness
    - Every test ETS produces must meet the exacting criteria of our *Standards for Quality and Fairness*. These standards reflect ETS's commitment to producing fair, valid, and reliable tests that can stand up to the most intense scrutiny, either in the courtroom or in the court of public opinion.
    - The ETS Office of Professional Standards Compliance audits each ETS testing program to ensure its adherence to the *ETS Standards for Quality and Fairness*.

- ETS Fairness Review
    - The *ETS Fairness Review Guidelines* identify aspects of test questions that might hinder people in various groups from performing at optimal levels.
    - Fairness reviews are conducted by specially trained reviewers.

- ETS International Principles for Fairness Review of Assessments
    - The *International Principles for Fairness Review of Assessments* was written to ensure that tests created under the guidance of ETS for a country outside the United States are fair and appropriate for test takers within the country.

Every member of the ETS staff responsible for writing and reviewing test questions undergoes extensive training in the fairness review process. ETS's commitment to fairness helps distinguish ETS from other assessment companies.

Policies and Guidelines for Score Use

These policies and guidelines are designed to provide information about the appropriate use of *TOEIC®* Speaking and Writing test scores for corporations or schools that use the scores in making hiring, evaluation, and career promotion decisions, as well as school-related decisions. They are also intended to protect test takers from unfair decisions that may result from inappropriate uses of scores. Adherence to the guidelines is important.

The TOEIC Speaking and Writing tests are designed to assess English Language proficiency relevant to today's international market. As measures with known statistical properties and high-quality technical characteristics, the scores from these tests, when used properly, can improve the hiring, evaluation, and career promotion decision processes of local and multi-national corporations, and other organizations where English communication skills are critical job requirements.

As more and more students go to colleges to acquire job skills in the global marketplace, schools are also increasingly using TOEIC Speaking and Writing scores to evaluate student proficiency in English.

The *TOEIC®* program and its authorized local EPN office have a particular obligation to inform users of the appropriate uses of TOEIC Speaking and Writing scores and to identify and try to rectify instances of misuse. To this end, the following policies and guidelines are available to all TOEIC Speaking and Writing test takers, institutions, and organizations that are recipients of TOEIC Speaking and Writing scores.

Policies

In recognition of their obligation to ensure the appropriate use of *TOEIC®* Speaking and Writing test scores, the TOEIC program and its authorized local EPN office developed policies designed to make score reports available only to approved recipients, to encourage these organizational score users to become knowledgeable about the validity of the tests, to protect the confidentiality of test takers' scores, and to follow up on cases of possible misuse of scores. The policies are discussed below.

Confidentiality

TOEIC Speaking and Writing test scores, whether for an individual or aggregated for an institution, are confidential and can be released only by authorization of the individual or institution or by compulsion of legal process. ETS recognizes test takers' right to privacy with regard to information that is stored in data or research files held by ETS and its local EPN office, and its responsibility to protect test takers from unauthorized disclosure of the information.

Testing Irregularities

"Testing irregularities" refers to irregularities in connection with the administration of a test, such as equipment failure, improper access to test content by individuals or groups of test takers, and other disruptions of test administrations (natural disasters and other emergencies). When testing irregularities occur, ETS and its local EPN office may decline to score the test or may cancel the test score. When it is appropriate to do so, ETS or the EPN office may give affected test takers the opportunity to take the test again as soon as possible without charge.

Identification Discrepancies

When, in the judgment of ETS, the EPN office, or test center personnel, there is a discrepancy in a test taker's identification, the test taker may be dismissed from the test center. In addition, ETS or the EPN office may decline to score a test or may cancel the scores and the test fee will be forfeited.

Misconduct

When ETS, the EPN office, or test center personnel find that there is misconduct in connection with a test, the test taker may be dismissed from the test center. ETS or the EPN office may decline to score the test or may cancel the test scores. Test takers whose scores are canceled because of misconduct will forfeit their test fees. Misconduct includes, but is not limited to, noncompliance with the Test Center Procedures and Regulations on pages 10–13 of the Examinee Handbook.

Plagiarism

ETS or the EPN office reserves the right to cancel the scores of test takers when, in their judgment, there is evidence that an essay or speaking response includes, for example, text or ideas that are substantially similar to that found in other TOEIC responses, or the quotation or paraphrasing, without attribution, of language or ideas from published and unpublished sources. Such responses do not reflect the independent speaking or compositional writing skills that the test seeks to measure.

Test Score Data Retention

TOEIC® Speaking and Writing test scores measure a test takers' English proficiency at the time the tests are administered. Test scores are retained in the *TOEIC*® database for only two years from the test date and scores more than two years old are not reported. Individuals who took the *TOEIC*® test more than two years ago must retake the test if they want scores. While all information that could be used to identify an individual is removed from the database after two years, anonymous score data and other information that can be used for research or statistical purposes are retained.

Score Verification

If the score user believes that a score does not accurately reflect the test taker's ability in English, the score user may contact the local EPN office to verify score results. Scores can be verified up to two years after the date of the test administration.

Score Cancellation

ETS or the local EPN office reserve the right to take all action—including but not limited to barring test takers from future testing and/or canceling scores—for failure to comply with test administration regulations or the test administrator's directions. If a test taker's scores are canceled, a letter will be sent from the EPN office. The test taker will not receive a refund.

Guidelines

All organizational users of *TOEIC®* Speaking and Writing test scores have an obligation to use the scores in accordance with the guidelines below, i.e., using multiple criteria, accepting only official TOEIC Speaking and Writing test scores, etc. Organizations have a responsibility to ensure that all individuals using TOEIC Speaking and Writing test scores are aware of these guidelines and to monitor the use of the scores, correcting instances of misuse when they are identified. The *TOEIC®* program and its authorized local EPN office are available to assist institutions in resolving score-misuse issues.

Use Multiple Criteria

Regardless of the decision to be made, multiple sources of information should be used to ensure fairness and to balance the limitations of any single measure of knowledge, skills, or abilities. These sources may include graduate or undergraduate grade point average, years of experience in the target position, and recommendations from past supervisors and colleagues. When used in conjunction with other criteria, *TOEIC®* scores can be a powerful tool in making hiring, evaluation, promotion, or school related decisions.

Accept Only Official *TOEIC®* Speaking and Writing Official Score Certificates and Score Reports

The only official certificates and reports of *TOEIC®* Speaking and Writing test scores are those issued by ETS or by the local EPN office. If an organization administers TOEIC Speaking and Writing tests internally, with ETS knowledge and approval, it can obtain and keep score reports of that test. Scores obtained from other sources should not be accepted. If there is a question about the authenticity of an official score certificate, the question should be referred to the local EPN office, who will then verify the accuracy of the scores and whether an official score certificate was issued.

Maintain Confidentiality of *TOEIC®* Speaking and Writing Test Scores

All individuals who have access to *TOEIC®* Speaking and Writing test scores should be aware of the confidential nature of the scores and agree to maintain their confidentiality. Policies should be developed and implemented to ensure that confidentiality is maintained.

Appropriate Uses and Misuses of *TOEIC®* Speaking and Writing Test Scores

The suitability of *TOEIC®* Speaking and Writing tests for a particular use should be explicitly examined before using test scores for that purpose. The following lists of appropriate uses of TOEIC Speaking and Writing test scores are based on the policies and guidelines outlined above. The list is meant to be illustrative, not exhaustive, in nature. There may be other appropriate uses of TOEIC Speaking and Writing test scores, but any uses other than those listed on the next page should be discussed in advance with *TOEIC®* program staff and its local EPN office to determine their appropriateness.

If a use other than those appropriate uses listed on the next page is contemplated, it will be important for the user to validate the use of scores for that purpose. TOEIC program staff and its authorized local EPN office will provide advice on the design of such validity studies.

Appropriate Uses

Used along with other sources of information, and provided all applicable guidelines are followed, *TOEIC*® Speaking and Writing test scores are suitable for the uses described below.

- Hiring of applicants for an open position within a corporation or organization where workplace/everyday-life English is a required job skill

- Placement of applicants or test takers within a corporation or organization where workplace/everyday-life English is a required job skill

- Promotion of test takers within a corporation or organization where workplace/everyday-life English is a required job skill

- Measurement of workplace/everyday-life English proficiency levels of students in schools

- Measurement of individuals' progress in workplace/everyday-life English proficiency levels over time

Misuses

- Any use of test scores for purposes other than those listed should be discussed with ETS before being implemented.

Contacting the *TOEIC*® Program

For more information, visit us on the web at **www.ets.org/toeic** or contact your local ETS Preferred Network (EPN) member.

A list of local EPN offices can be found on the web at **www.ets.org/toeic**.

If there is no EPN office in your country, please contact us at:

Mail:

TOEIC Program
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Phone: 1-609-771-7170
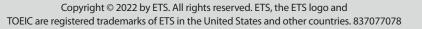
Fax: 1-973-735-1903 or 1904

E-mail: TOEIC@ets.org

For more than 40 years as an industry leader, the *TOEIC*® program has set the global standard for assessing English-language communication skills needed in the workplace and everyday life. The TOEIC assessments are the most widely used around the world with 14,000+ organizations across more than 160 countries trusting TOEIC scores to inform the decisions that matter.

At nonprofit ETS, our mission is to advance quality and equity in education by providing fair and valid assessments, research and related services. Our products and services measure knowledge and skills, promote learning and performance, and support education and professional development for all people worldwide.

www.ets.org