

TOEFL® Research INSIGHT

TOEFL Primary®
Framework and
Test Development

VOLUME 8

TOEFL® Research Insight Series, Volume 8: TOEFL Primary® Framework and Test Development

Preface

The *TOEFL iBT*® test is the world's most widely respected English language assessment and is used for admissions purposes in more than 130 countries, including Australia, Canada, New Zealand, the United Kingdom, and the United States. Since its initial launch in 1964, the *TOEFL*® test has undergone several major revisions motivated by advances in theories of language ability and changes in English teaching practices. The most recent revision, the *TOEFL iBT* test, was launched in 2005. It contains a number of innovative design features, including integrated tasks that engage multiple skills to simulate language use in academic settings, and test materials that reflect the reading, listening, speaking, and writing demands of real-world academic environments.

In addition to the *TOEFL iBT* test, the *TOEFL* Family of Assessments has been expanded to provide high-quality English proficiency assessments for a variety of academic uses and contexts. The *TOEFL* Young Students Series (YSS) features the *TOEFL Primary*® and *TOEFL Junior*® tests, which are designed to help teachers and learners of English in school settings. The *TOEFL ITP*® program offers colleges, universities, and others affordable tests for placement and progress monitoring within English programs.

At ETS, we understand that scores from the *TOEFL* Family of Assessments are used to help make important decisions about students, and we would like to keep score users and test takers up-to-date about the research results that assure the quality of these scores. Through the publication of the *TOEFL Research Insight Series*, we wish to communicate to the institutions and English teachers who use any/all of the *TOEFL* tests about the strong research and development base that underlies the *TOEFL* Family of Assessments and to demonstrate our continued commitment to research.

Since the 1970's, the *TOEFL* test has had a rigorous, productive, and far-ranging research program. But why should test score users care about the research base for a test? In short, it is only through a rigorous program of research that a testing company can substantiate claims about what test takers know or can do based on their test scores, as well as provide support for the intended uses of assessments. Beyond demonstrating this critical evidence of test quality, research is also important for enabling innovations in test design and ensuring that the needs of test takers and test score users are persistently met. This is why ETS has made the establishment of a strong research base a fundamental feature underlying the evolution of the *TOEFL* Family of Assessments.

The *TOEFL* Family of Assessments is designed, produced, and supported by a world-class team of test developers, educational measurement specialists, statisticians, and researchers in applied linguistics and language testing. Our test developers have advanced degrees in fields such as English, language education, and applied linguistics. They also possess extensive international experience, having taught English on continents around the globe. Our research, measurement, and statistics teams include some of the world's most distinguished scientists and internationally recognized leaders in diverse areas such as test validity, language learning and assessment, and educational measurement.

To date, more than 300 peer-reviewed TOEFL research reports, technical reports, and monographs have been published by ETS, and many more studies on TOEFL tests have appeared in academic journals and book volumes. In addition, over 20 TOEFL related research projects are conducted by ETS's Research & Development staff each year and the TOEFL Committee of Examiners (COE), comprised of language learning and testing experts from the academic community, funds an annual program of TOEFL research by independent external researchers from all over the world.

The purpose of the *TOEFL Research Insight Series* is to provide a comprehensive yet user-friendly account of the essential concepts, procedures, and research results that assure the quality of scores for all members of the TOEFL Family of Assessments. Topics covered in these volumes include issues of core interest to test users, including how tests were designed, evidence for the reliability and validity of test scores, and research-based recommendations for best practices.

The close collaboration with TOEFL score users, English language learning and teaching experts, and university scholars in the design of all TOEFL tests has been a cornerstone to their success. Therefore, through this publication, we hope to foster an ever-stronger connection with our test users by sharing the rigorous measurement and research base and solid test development that continues to ensure the quality of the TOEFL Family of Assessments.

Dr. John Norris

Senior Research Director
English Language Learning and Assessment
Research & Development Division
Educational Testing Service (ETS)

TOEFL Primary Framework and Test Development

In countries where English is taught as a foreign language, it has become increasingly common for educational agencies to introduce English language instruction at the earliest grades. This trend reflects a growing understanding of the benefits of early language learning, as well as the reality of today's increasingly globalized world in which the ability to communicate in English is a bridge to highly valued opportunities in one's school, workplace, and personal life.

Moreover, in countries where English is a main or official language, learners for whom English is not their first or native language receive instruction in English as a second language (ESL). Often, this instruction also begins at the earliest grades in primary or elementary school. Given the increasing prevalence of English instruction in the early grades, it is critical to provide high-quality, objective English proficiency measures that are designed with attention to the unique needs of young learners. The TOEFL Primary tests, targeting English learners ages 8+, were introduced in 2013 to help fill this gap.

The TOEFL Primary Test Framework

The TOEFL Primary tests are the result of collaboration with leading experts from around the world. During the development of these tests, researchers surveyed existing scientific literature, curricula, standards, and textbooks to identify key English language knowledge, skills, and abilities (KSAs) and understand the unique challenges involved in assessing young English as a Foreign Language (EFL) students. The content and design of the tests were continuously modified and improved on the basis of the findings of a series of prototyping studies and insights from experts inside and outside ETS. As a standardized international assessment of English ability, the TOEFL Primary tests are not tied to any specific curriculum. Rather, the tests focus on communication skills and activities that are commonly found in EFL instruction for young EFL students.

Target Population

The TOEFL Primary tests were designed to serve children between 8 and 12 years of age who are both learning English in countries where English is a foreign language and have limited opportunities to use English, either inside or outside the classroom. Students in the tests' intended population are expected to possess a wide range of levels of English language proficiency, as they have different educational experiences and varied access to additional language learning support. The TOEFL Primary tests are designed to cover the wide range of proficiency levels represented among young EFL learners.

Test Purpose and Intended Uses

As independent measures of English communication skills in three areas—reading, listening, and speaking—the TOEFL Primary tests are intended to support teaching and learning by providing meaningful feedback that teachers can incorporate into their instruction. Test scores may be used to:

- assess the general English language proficiency of young students ages 8+
- obtain a snapshot of each student's ability in listening, reading, and speaking
- understand students' abilities in relation to a widely accepted international standard

It is not desirable to use TOEFL Primary test scores for high-stakes decisions, such as admitting students, evaluating teachers, and comparing or ranking individual students.

Test Content

The TOEFL Primary tests measure young EFL students' ability to communicate in English in three modalities—reading, listening, and speaking. The *TOEFL*® Primary Reading and Listening test sections must be taken together as a single test. The *TOEFL*® Primary Speaking test is taken as an independent test. Each test focuses on the ability to use English in accomplishing communication goals in familiar and age-appropriate contexts. Thus, the test tasks used in the TOEFL Primary tests were designed to resemble real-life language use situations that students are likely to encounter in learning English, as well as measure enabling language knowledge and skills that support the development of communication ability.

TOEFL Primary Reading Test

This section measures the ability to use English to achieve the following **communication goals**:

- Identify people, objects, and actions
- Understand commonly occurring nonlinear written texts (e.g., signs, schedules)
- Understand written directions and procedures
- Understand short, personal correspondence (e.g., letters)
- Understand simple, written narratives (e.g., stories)
- Understand written expository or informational texts about familiar people, objects, animals, and places

To achieve these goals, young EFL students need the following **enabling knowledge and skills**:

- Recognize the written English alphabet and sounds associated with each letter
- Identify words based on sounds
- Recognize the mechanical conventions of written English
- Recognize basic vocabulary
- Process basic grammar
- Identify the meaning of written words through context
- Recognize the organizational features of various text types

TOEFL Primary Listening Test

This section measures test takers' ability to use English to achieve the following **communication goals**:

- Understand simple descriptions of familiar people and objects
- Understand spoken directions and procedures
- Understand dialogues or conversations
- Understand spoken stories
- Understand short informational texts related to daily life (e.g., phone messages, announcements)
- Understand simple teacher talks on academic topics

To achieve these goals, young EFL students need the following **enabling knowledge and skills**:

- Recognize and distinguish English phonemes
- Comprehend commonly used expressions and phrases
- Understand very common vocabulary and function words
- Identify the meaning of spoken words through context
- Understand basic sentence structure and grammar
- Understand the use of intonation, stress, and pauses to convey meaning
- Recognize organizational features of conversations, spoken stories, and teacher talks

TOEFL Primary Speaking Test

This test measures test takers' ability to use English to achieve the following **communication goals**:

- Express basic emotions and feelings
- Describe people, objects, animals, places, and activities
- Explain and sequence simple events
- Make simple requests
- Give short commands and directions
- Ask and answer questions

To achieve these goals, young EFL students need the following **enabling knowledge and skills**:

- Pronounce words clearly
- Use intonation, stress, and pauses to pace speech and convey meaning
- Use basic vocabulary and common and courteous expressions
- Use simple connectors (e.g., and, then)

Test Structure and Format

Depending on school curricula and other factors, young students acquire their English abilities at different times and in different ways. The TOEFL Primary program offers 3 tests to measure a range of skills in each modality. The Reading and Listening tests are available at two difficulty levels (Step 1 and Step 2). The Speaking test is a single-level test that both Step 1 and Step 2 test takers can take.

The TOEFL Primary Reading and Listening tests include three-option multiple-choice items, pictures, and a variety of text types in order to keep students engaged and focused while taking the test.

Table 1. TOEFL Primary Reading and Listening Test — Step 1*

Test	Number of Questions	Number of Examples	Total Number of Questions	Time
Reading Section	36	3	39	30 min
Listening Section	36	5	41	30 min

*Paper or digitally delivered

Table 2. TOEFL Primary Reading and Listening Test — Step 2*

Test	Number of Questions	Number of Examples	Total Number of Questions	Time
Reading Section	36	1	37	30 min
Listening Section	36	3	39	30 min

*Paper or digitally delivered

The TOEFL Primary Speaking test includes 7 constructed-response items that are presented in a scenario. During the Speaking test, students speak to multiple fictional virtual characters. Animations, playful characters, and whimsical content are used to keep students engaged and elicit more spontaneous and natural responses.

Table 3. TOEFL Primary Speaking Test*

Test	Number of Questions	Time
Speaking	7	20 min

*Paper or digitally delivered

Test Development

ETS maintains a continuous and rigorous process of producing and vetting new items and test content for the TOEFL Primary tests.

Content Development Staff

The TOEFL program maintains high standards for test content developers, using only carefully selected, highly qualified staff to write items and create content for the TOEFL Primary tests. All members of the test development staff are thoroughly trained in the process of authoring quality items. In addition, they all have formal university-level training in language learning or related subject areas. The majority of ETS's test development staff hold graduate-level degrees from English-medium universities and have taught at schools or universities internationally.

Item Writing

In order to ensure that the test content is as comparable as possible across all administrations of the TOEFL Primary tests, each item writer follows detailed item writing guidelines when creating test questions and other test content, such as reading passages or lectures. They make sure test questions and content:

- are clear and coherent;
- are culturally accessible and appropriate;
- are at an appropriate level of difficulty;
- do not require background knowledge in order to be comprehensible;
- align with ETS fairness guidelines; and
- contain sufficient testable content.

These principles are fundamental to all TOEFL Primary test development processes.

Item Review Process

All items used on the TOEFL Primary tests are subject to a rigorous review process, including content, fairness, and editorial reviews.

Content Review

Before an item is considered fit for operational use, it has to pass a rigorous quality control process that consists of two key review stages: content review and fairness review. Upon completion of the first rough draft of an item, the item writer sends the item into content review. At the content review stage, different TOEFL Primary assessment development specialists will answer the item like a test taker and then independently revise the item to improve quality. Each change is documented in the comments section of the database for subsequent reviewers. Ultimately, the item writer revises the item based on the commentary provided. Multiple iterations of content review are conducted until all review comments are addressed and no further issues are flagged. The reviews focus on questions such as these:

- Is the language in the test materials clear? Is it accessible to a nonnative speaker of English in our target population? Is it age appropriate?
- Is the content of the stimulus accessible to nonnative speakers who lack specialized knowledge about a given topic?

For multiple-choice questions, reviewers also consider the following factors:

- the appropriateness of the point tested
- the uniqueness of the answer or answers
- the clarity and accessibility of the language used
- the plausibility and attractiveness of the incorrect answer choices

For constructed-response items in the TOEFL Primary Speaking test, the review process is similar but not identical. Reviewers tend to focus on accessibility, clarity in the language used, and how well they believe the particular Speaking item will generate a fair and scorable response. It is also essential that reviewers judge each Speaking item to be comparable with others in terms of difficulty. Expert judgment plays a major role in deciding whether a Speaking item is acceptable and can be included in an operational test.

Fairness Review

After an item has successfully passed the content review stage, it enters fairness review—a process that ensures that items are fair and equitable to test takers of all cultural and ethnic backgrounds.

The *ETS Standards for Quality and Fairness* (ETS, 2014) mandate fairness reviews. This fairness review must take place before a test item is administered to test takers. All ETS test developers undergo fairness training (in addition to item writing training) soon after their arrival at ETS. As part of their training, item writers become familiar with the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2016a) and the *ETS International Principles for Fairness Review of Assessments* (ETS, 2016b) and use them when developing and reviewing test content. Although fairness issues are considered at each stage of the development process, they are particularly focused on at the fairness review stage.

During fairness review, specially trained fairness reviewers conduct an independent review of all TOEFL Primary test materials. TOEFL Primary test developers may not perform this official fairness review; the official fairness reviewer is typically a test developer who works on other ETS tests. In this way, the fairness review is more objective. When fairness reviewers find unacceptable content in the test materials, they issue a fairness challenge. A content reviewer must then work with the fairness reviewer to resolve the challenge to the satisfaction of both reviewers. For rare cases in which the reviewers cannot reach agreement, a panel of both content and fairness reviewers decides on the issues at hand and comes to a resolution.

Editorial Review

All TOEFL Primary test materials also receive an editorial review. The purpose of this review is to ensure that language in the test materials is clear, concise, and consistent. Editors ensure that established ETS test style is followed. All suggestions for changes need to be approved by the content specialist for the given test section.

Item Pretesting and Tryout

TOEFL Primary Reading and Listening Test

All TOEFL Primary multiple-choice test items are pretested with a large number of test takers. Pretest items are included in operational forms, and data are collected on real TOEFL Primary test takers' ability to answer the items. Test takers cannot identify pretest items because they do not differ in any distinguishable way from the operational (i.e., scored) items on the test. Pretesting items allow test developers to identify poorly functioning items and revise or exclude them from the operational item pool. Test developers review data from item pretesting and use the information to refine their understanding of what makes a good test item.

TOEFL Primary Speaking Test

In operational administrations, the TOEFL Primary Speaking test does not contain embedded pretest items. Instead, ETS conducts small-scale tryouts of Speaking items with young English language learners. Test developers review and evaluate spoken responses to these tryout questions, using expert judgment to determine which prompts are likely to elicit valid and scorable responses from test takers across the range of proficiency levels. These viable prompts are the ones that appear in operational test forms.

Scoring

TOEFL Primary Reading and Listening Tests

These tests are scored locally by ETS's Preferred Network offices. The test scores are determined by the number of questions a student has answered correctly. There is no penalty for wrong answers. The number of correct responses on each section is then converted to a scaled score of 101–109 points for Step 1 and 104–115 for Step 2. This is done using a statistical procedure that takes “raw” scores obtained on each section and transforms or adjusts them to a standardized scale. Transforming raw scores into scaled scores allows for comparison of scores across different test administrations.

TOEFL Primary Speaking Test

Responses to the TOEFL Primary Speaking test are scored by human raters at ETS using scoring rubrics with either a 0–3 point scale or a 0–5 point scale, depending on the task type. The range of speaking scores is 0–27. The scoring rubrics were developed based on performance data collected in the pilot and field test administrations of the test. The rubrics identify three major dimensions that are taken into consideration—language use, content, and delivery—with each dimension considered in relation to the clarity of overall meaning. Despite the fact that three dimensions are considered, only one “holistic” score is assigned to each response.

Scoring Speaking responses presents challenges that multiple-choice testing does not. Whereas multiple-choice tests can be scored objectively, rating speaking performances relies on human judgment. ETS supports scoring quality and consistency for the TOEFL Primary Speaking test in a number of ways:

- Raters must be qualified. In general, they must be experienced teachers, ESL or EFL specialists, or in possession of other relevant experience. In addition to teaching experience, ETS prefers raters who have master's degrees and experience assessing spoken and written language.
- If they have the formal qualifications, raters are then trained. ETS trains raters using a web-based system. Following their training, raters must pass a certification test in order to be eligible to score. To assure reliability of constructed-response scoring, ETS monitors raters continuously as they score.
- Nonnative speakers of English may be raters and, in fact, contribute a much-needed perspective to the rater pool, but they must pass the same certification test as native-speaking raters.
- The scoring process is centralized, and it is performed separately from the test center administration in order to ensure that test data are not compromised. Through centralized, separate scoring, each scoring step is closely monitored to ensure its security, fairness, and integrity.
- ETS uses its patented Online Network for Evaluation to distribute test takers' responses to raters, record ratings, and monitor rating quality constantly.

At the beginning of each rating session, raters must pass a calibration test for the specific task type they will rate before they proceed to operational scoring. Scoring leaders—the scoring session supervisors—monitor raters in real time, throughout the day. These supervisors also regularly work as raters on different scoring shifts and are subject to the same monitoring. No rater, no matter how experienced, scores without supervision. ETS test developers also monitor rating quality and communicate with scoring leaders during rating sessions.

For each administration, ETS's Online Network for Evaluation sends Speaking responses to multiple independent raters for scoring. Each test taker's responses are scored by more than one rater. When a discrepancy between raters arises, it is resolved by a third rater.

Score Reporting

After taking the TOEFL Primary tests, students receive score reports and certificates of achievement, while schools and teachers receive a group-level score report for their students. These reports provide detailed and comprehensive information regarding students' performance on the test.

Individual score reports provide a variety of information:

- **Numeric scores** for each skill to help measure progress
- **Band level and performance descriptors** to provide meaningful descriptive information and recommend next steps that students can take to improve their English language abilities
- **Lexile® scores** to help identify reading materials that match students' current reading levels
- **Common European Framework of Reference (CEFR) levels** to help interpret students' abilities in relation to an international English language proficiency standard

A group-level score report is available for teachers and schools to view their students' performance and keep track of progress.

Ongoing Oversight

Ongoing oversight is a key feature of the TOEFL Family of Assessments. The TOEFL Primary tests undergo regular internal audits every three years. The auditors evaluate compliance with ETS's Standards for Quality and Fairness and report directly to the ETS Board of Trustees on any issues they may find.

Additionally, the COE provides guidance and oversight for research and development related to all tests in the TOEFL Family of Assessments. The COE is a panel of 12 experts from around the world, each of whom has achieved professional recognition in an academic field related to learning and testing English as a second or foreign language. The TOEFL YSS Subcommittee of the COE consists of three individuals with specialized expertise in the education and assessment of young learners. They advise on research and development efforts related to the TOEFL Primary tests.

References

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Educational Testing Service. (2016a). *ETS guidelines for fairness of assessments*. Princeton, NJ: Author.

Educational Testing Service. (2016b). *ETS international principles for fairness of assessments*. Princeton, NJ: Author.