

TOEFL® Research **INSIGHT**

TOEFL Junior®
Framework and
Test Development

VOLUME 7

TOEFL® Research Insight Series, Volume 7: TOEFL Junior® 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 Junior Framework and Test Development

In order to promote the development of English ability early on—especially in countries where English is taught as a foreign language—English language instruction has become a regular part of school curricula in middle and high school. In some countries, in order to enjoy the benefits of early language learning, English instruction has been introduced at lower elementary school grades. Similarly, in places where English is used for everyday communication (such as the United States), English is taught as a second language to students whose first or native language is not English. In both English as a Foreign Language (EFL) and English as a Second Language (ESL) contexts, the main objective of English language teaching is to enable learners to communicate—opening the door to a wide range of personal, academic, and professional opportunities (Wolf & Butler, 2017). As the demand for high-quality EFL and ESL instruction increases, so does the need for appropriate, well-designed, and objective English proficiency measures for young learners. The TOEFL Junior test has been developed to address this need and provide relevant and reliable information to different stakeholders (e.g., teachers, students, parents) about the English language proficiency (ELP) of young, adolescent English learners worldwide.

The TOEFL Junior Test Framework

Originally launched in 2010, the TOEFL Junior program’s suite includes two tests that measure a range of communication skills in English:

- The *TOEFL Junior*® Standard test assesses students’ listening comprehension, language form and meaning, and reading comprehension.
- The *TOEFL Junior*® Speaking test measures a student’s ability to communicate orally in English in a school context.

The TOEFL Junior tests were designed to be curriculum independent. During their development, researchers and test designers surveyed curricula from various countries, interviewed English teachers, and identified key competencies and skills taught in English language education around the world (So, 2014). In other words, the TOEFL Junior tests are not based on any specific curriculum but rather on a global standard, measuring competencies, skills, and abilities needed for successful communication in ESL and EFL.

Target Population

The TOEFL Junior tests are typically used to assess the English language proficiency of students ages 11+. The majority of students who take them are 11–15 years of age. Their educational backgrounds and real-world experiences vary, but they typically have at least five full years of educational experience at the primary and/or middle school levels and some exposure to English instruction.

Test Purpose and Intended Uses

As part of the TOEFL Family of Assessments, the TOEFL Junior tests are intended to measure the communicative ability students need in order to participate in English-medium school settings. That is, the tests measure the English skills that students at the lower and intermediate levels of secondary school need to successfully navigate both social and academic situations in English-medium instructional environments.

The tests can be used for different purposes. First, they can be used to track students' progress over time, providing students, parents, and teachers with objective information about students' growing and developing English skills. Second, they can be used to support placement decisions. That is, the TOEFL Junior tests can serve as measurement tools that can help to place learners into programs and classes that teach English. Finally, they can be used to support instruction by providing information about learners' proficiency to teachers, parents, and other stakeholders (Gu et al., 2015; Papageorgiou & Cho, 2014). They also offer useful information that can be used to inform instruction.

Testing Format and Test Content

The TOEFL Junior Standard test can be delivered on paper or in digital form, while the TOEFL Junior Speaking test is a digitally delivered test.

All items included in the TOEFL Junior tests are designed to be reflective of activities that require language use in secondary-level, English-medium school settings. This is what we refer to as the “target-language use domain” for the TOEFL Junior tests.

This target-language use domain—language use in secondary-level, English-medium settings—can be further subdivided into three subdomains: academic, social-interpersonal, and navigational. The academic subdomain includes language activities performed to learn academic content in English, such as reading academic texts or summarizing a written or oral text. The academic subdomain targets more technical and formal (i.e., academic) language. In contrast, the social-interpersonal subdomain includes more personal and informal registers of language use, such as talking to friends at school. The navigational subdomain includes language activities that require students to navigate their way through a specific situation. For example, they need to be able to parse a school announcement and extract specific information or communicate with peers about how to do specific homework tasks. Each section of the TOEFL Junior tests includes items that represent one or more of the three target-language use subdomains (So et al., 2015).

TOEFL Junior Standard Test

This test includes three sections that measure listening comprehension, language form and meaning, and reading comprehension. Each test section consists of 42 multiple-choice questions.

Table 1. Structure of the TOEFL Junior Standard Test

Sections	Number of Items	Scale Scores	Testing Time
Listening Comprehension	42	200–300	40 min
Language Form and Meaning	42	200–300	25 min
Reading Comprehension	42	200–300	50 min
Total	126	600–900	1 hr 55 min

Listening and Reading

In the **Listening Comprehension** section, test takers listen to audio-recorded input. The listening stimuli include short conversations, classroom instructions, and academic lectures—which are reflective of the three subdomains of target language use (social-interpersonal, navigational, and academic). Tasks in each subdomain measure a number of listening subskills, including:

- Comprehending main ideas
- Identifying salient details
- Making inferences
- Making predictions
- Identifying the speaker’s purpose
- Understanding meaning conveyed by prosodic and idiomatic features

In the **Reading Comprehension** section, test takers are presented with academic texts that are typical of those used in middle schools, age-appropriate journalistic texts, and personal correspondence (e.g., letters or emails). The multiple-choice items that follow the reading passages measure the following reading subskills:

- Comprehending main ideas
- Identifying salient details
- Making inferences
- Discerning the meaning of low-frequency words or expressions from context
- Recognizing an author’s purpose or use of particular rhetorical structures
- Understanding figurative and idiomatic language from context

Language Form and Meaning

The **Language Form and Meaning** section assesses language skills that underlie and enable communication. To measure these enabling skills, such as grammatical and lexical knowledge, students are provided with reading passages from which four to eight words or phrases have been deleted. From a list of four possible answer options, test takers have to select the answer option that completes the sentence correctly, either from a grammatical perspective or meaning perspective. The grammar items in the assessment focus on subject-verb agreement, correct subject-object form, verb tense and aspect, active/passive voice, relative clauses, word order, and comparative/superlative forms. The language meaning-oriented items focus on the meaning of verbs, nouns, adjectives, determiners, adverbs, conjunctions, and prepositions.

TOEFL Junior Speaking Test

The TOEFL Junior Speaking test measures the speaking skills that students need in order to communicate in English-medium school contexts. This test consists of four tasks: read aloud, picture narration, and two listen-speak tasks—one academic and one nonacademic. All TOEFL Junior Speaking tasks, except the picture

narration task, involve integrating different language skills (i.e., students have to read a text or listen to oral language in order to complete the speaking part). These “integrated” tasks were chosen because they better reflect how language is used in the real world.

Table 2. Structure of the TOEFL Junior Speaking Test

Task Type	Items Per Form	Prep Time	Speaking Time	Total Time*	Maximum Score Points
Section Directions and Microphone Check	--	--	--	2 min 10 sec	--
Read Aloud	1	1 min	1 min	3 min 30 sec	4
6-Picture Narration	1	1 min	1 min	3 min 20 sec	4
Nonacademic Listen-Speak: Class Activity	1	45 sec	1 min	4 min 30 sec	4
Academic Listen-Speak: Course Content	1	45 sec	1 min	4 min 30 sec	4
Total	4	3 min 30 sec	4 min	18 min	16

*Note: Total time includes instructions, stimulus, and preparation time.

Test Development

ETS maintains a continuous and rigorous process of producing and reviewing new items and test content for the TOEFL Junior 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 Junior 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.

Additionally, the TOEFL program selects and trains outside item writers. Each year, ETS offers a summer institute in which candidates are trained in item writing for specific TOEFL assessments. At the end of the 6-week intensive training, the top candidates are selected and hired as external item writers. All external item writers have experience teaching ESL/EFL or related academic content areas.

Item Writing

In order to ensure that the test content is as comparable as possible across all administrations of the TOEFL Junior 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 Junior test development processes.

Item Review Process

All items used on TOEFL Junior 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 Junior test 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 then 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 an English language learner at the middle school level? 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 Junior 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 an appropriate 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) mandates 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 Junior test materials. TOEFL Junior 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 Junior 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 Junior Standard Test

All TOEFL Junior test items are pretested with at least 800 test takers before they can be used as actual scored test items. Pretest items are included in operational forms, and data are collected on real TOEFL test takers' ability to answer the items. Test takers cannot identify pretest items because they do not differ in any distinguishable way from 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 Junior Speaking Test

In operational administrations, the TOEFL Junior Speaking test does not contain embedded pretest items. Instead, ETS conducts small-scale tryouts of Speaking items among members of the test's target population. 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

To account for the different skills that are measured by the TOEFL Junior Standard and TOEFL Junior Speaking tests, the scoring procedures for the two tests are slightly different.

TOEFL Junior Standard Test

This test is scored locally by ETS's Preferred Network offices. First, "raw" scores are determined by counting the number of questions a student has answered correctly. There is no penalty for wrong answers. In a statistical procedure, these "raw" scores obtained on each section are then transformed into "scaled" scores on a standardized scale of 200–300 points. Transforming raw scores into scaled scores allows for comparison of scores across different test administrations.

TOEFL Junior Speaking Test

This test is scored by thoroughly trained human raters who use specific rating scales for each speaking task. The ratings scales can be viewed at www.ets.org/s/toefl_junior/pdf/toefl_junior_speaking_scoring_guides.pdf. The following language features are considered in assigning scores for each item:

- Read aloud: fluency and accuracy
- Picture narration: content, delivery, and language use
- Listen-speak tasks: content, delivery, and language use

For each one of the four tasks, test takers obtain a score between 0 and 4 points, with a maximum possible score of 16.

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 Junior Speaking section 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

In general, TOEFL Junior test scores are available for reporting for two years after the test date. Because English language proficiency can change rapidly with additional study, decisions involving scores will be most valid when the scores have been obtained as recently as possible. Institutions can choose to use TOEFL Junior test scores that are more than two years old. However, a student's proficiency in English may change considerably after two years. Each TOEFL Junior score report provides:

- Section and overall score levels to assist teachers, parents, and students in understanding what the scores mean
- A description of the English language abilities typical of test takers scoring around a particular scaled score level, to help identify strengths and areas for improvement
- The Common European Framework of Reference (CEFR) levels that correspond to the section test scores
- A Lexile® measure to help find books at the student's English reading level

CEFR Levels

The CEFR is a tool used internationally to understand and compare different stages of language learning. TOEFL Junior scores are mapped to CEFR levels to make it possible to compare learners' performances to an international standard, as well as provide a more detailed understanding of students' English proficiency levels. In 2010, a standard-setting study was conducted by ETS researchers to map TOEFL Junior Standard test scores to the CEFR. The purpose of the study was to provide policymakers with recommended minimum scale scores (cut scores) for the Listening Comprehension, Language Form and Meaning, and Reading Comprehension sections. As a result of additional research, the TOEFL Junior Standard cut scores for CEFR levels A2, B1, and B2 were revised in October of 2012. Research for recommended minimum scores for the TOEFL Junior Speaking test was also conducted in 2012 (Baron & Tannenbaum, 2011; Tannenbaum & Baron, 2015).

Table 3. Recommended TOEFL Junior Standard Cut Scores for CEFR levels A2, B1, and B2

Section	CEFR Level			
	Below A2	A2	B1	B2
Listening	Under 225	225–245	250–285	290–300
Language Form and Meaning	Under 210	210–245	250–275	280–300
Reading	Under 210	210–240	245–275	280–300

Table 4. Recommended TOEFL Junior Speaking Cut Scores for CEFR levels A2, B1, and B2

	CEFR Level			
	Below A2	A2	B1	B2
Speaking	Under 8	8–10	11–13	14–16

The Lexile Score

The Lexile Framework® for Reading is a scientific way to provide a score for a person’s reading ability. In a scientific study, TOEFL Junior Reading section test scores were mapped to Lexile scores. These Lexile scores are provided on TOEFL Junior Standard score reports. Lexile levels for many books can be found online (<https://fab.lexile.com/>), helping parents and teachers to select texts at an appropriate level of difficulty to support individual learners’ reading skill development in English.

Ongoing Oversight

Ongoing oversight is a key feature of the TOEFL Family of Assessments. The TOEFL Junior 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 Junior tests.

References

- Baron, P., & Tannenbaum, R. (2011). *Mapping the TOEFL® Junior™ test onto the Common European Framework of Reference*. ETS Research Memorandum (RM-11-07). Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (2014). *ETS standards for quality and fairness*.
- Educational Testing Service. (2016a). *ETS guidelines for fairness of assessments*.
- Educational Testing Service. (2016b). *ETS international principles for fairness of assessments*.
- Gu, L., Lockwood, J., & Powers, D. E. (2015). *Evaluating the TOEFL Junior® Standard Test as a Measure of Progress for Young English Language Learners*. ETS Research Report (RR-15-22). Princeton, NJ: ETS.
- Papageorgiou, S., & Cho, Y. (2014). An investigation of the use of TOEFL® Junior™ Standard scores for ESL placement decisions in secondary education. *Language Testing*, 31(2) 223–239.
- So, Y. (2014). Are teacher perspectives useful? Incorporating EFL teacher feedback in the development of a large-scale international English test. *Language Assessment Quarterly*, 11(3), 283–303.
- So, Y., Wolf, M. K., Hauck, M. C., Mollaun, P., Rybinski, P., Tumposky, D., & Wang, L. (2015). *TOEFL Junior design framework*. TOEFL Junior Research Report (RR-15-13). Princeton, NJ: Educational Testing Service.
- Tannenbaum, R., & Baron, P. (2015). *Mapping scores from the TOEFL® Junior Comprehensive test onto the Common European Framework of Reference (CEFR)*. ETS Research Memorandum (RM-15-13). Princeton, NJ: Educational Testing Service.
- Timpe-Laughlin, V. (2018). *A good fit?—Examining the alignment between the TOEFL Junior Standard test and the English as a foreign language (EFL) curriculum in Berlin, Germany*. ETS Research Memorandum (RM-18-11). Princeton, NJ: Educational Testing Service.
- Wolf, M. K., & Butler, Y. G. (2017). An overview of English language proficiency assessments for young learners. In: M.K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 3–21). New York: Routledge.