ETS TOEFL

Quality Beyond Measure.

TOEFL Junior® Research Report
TOEFL JR–02
ETS Research Report No. RR–15-13

TOEFL Junior® Design Framework

Youngsoon So

Mikyung Kim Wolf

Maurice C. Hauck

Pamela Mollaun

Paul Rybinski

Daniel Tumposky

Lin Wang

June 2015

Discover this journal online at
Wiley Online Library
wileyonlinelibrary.com

The *TOEFL*® test was developed in 1963 by the National Council on the Testing of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program. In 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the *Graduate Record Examinations*® (*GRE*®) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education. The test is now wholly owned and operated by ETS.

ETS administers the TOEFL program under the general direction of a policy board that was established by, and is affiliated with, the sponsoring organizations. Members of the TOEFL Board (previously the Policy Council) represent the College Board, the GRE Board, and such institutions and agencies as graduate schools of business, two-year colleges, and nonprofit educational exchange agencies.

Since its inception in 1963, the TOEFL has evolved from a paper-based test to a computer-based test and, in 2005, to an Internet-based test, the *TOEFL iBT*® test. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. From 1977 to 2005, nearly 100 research l reports on the early versions of TOEFL were published. In 1997, a monograph series that laid the groundwork for the development of TOEFL iBT was launched. With the release of TOEFL iBT, a TOEFL iBT report series has been introduced.

Currently this research is carried out in consultation with the TOEFL Committee of Examiners (COE). Its members include representatives of the TOEFL Board and distinguished English as a second language specialists from academia. The committee advises the TOEFL program about research needs and, through the research subcommittee, solicits, reviews, and approves proposals for funding and reports for publication. Members of the TOEFL COE serve 4-year terms at the invitation of the Board; the chair of the committee serves on the Board.

Current (2014 – 2015) Members

TOEFL COE

Sara Weigle – Chair	Georgia State University
Yuko Goto Butler	University of Pennsylvania
Sheila Embleson	York University
Luke Harding	Lancaster University
Eunice Eunhee Jang	University of Toronto
Marianne Nikolov	University of Pécs
Lia Plakans	University of Iowa
James Purpura	Teachers College, Columbia University
John Read	The University of Auckland
Carsten Roever	The University of Melbourne
Diane Schmitt	Nottingham Trent University
Paula Winke	Michigan State University

Young Students Research Subcommittee

Yuko Goto Butler, University of Pennsylvania
Eunice Eunhee Jang, University of Toronto
Marianne Nikolov, University of Pécs

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

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's nonprofit Charter and Bylaws, ETS has and continues to learn from and to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

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

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

TOEFL Junior® Design Framework

Youngsoon So,[1] Mikyung Kim Wolf,[2] Maurice C. Hauck,[2] Pamela Mollaun,[2] Paul Rybinski,[2] Daniel Tumposky,[2] & Lin Wang[2]

1 Seoul National University, Seoul, South Korea
2 Educational Testing Service, Princeton, NJ

This paper presents the theoretical and empirical foundations of the *TOEFL Junior*® assessment and its development process. The TOEFL Junior test was developed to address the increasing need for objective measures of English language proficiency for young adolescent learners, who are being introduced to English as a second or foreign language at a much younger age than ever before. This paper presents the test purposes and intended uses, target population, target language use domains, and test constructs of the TOEFL Junior test. Also included is a description of the overall test structure and scoring system, which demonstrates how the constructs are operationalized. Finally, we outline research topics to support the interpretive argument of the use of the test. This document is expected to serve as a reference point during investigations of validity evidence to support the intended test uses over time.

Keywords English as a second or foreign language (ESL/EFL); young adolescent English learners; communicative language ability; English language proficiency (ELP); English for academic purposes; test specifications; interpretive argument of the test use

doi:10.1002/ets2.12058

This framework document describes the key elements of the *TOEFL Junior*® test and its development process. By documenting the design framework, including the construct definition and test characteristics, we demonstrate that our test design and development processes meet high professional standards in order to produce a quality assessment. The document also serves as a reference point during investigations of validity evidence to support the intended test uses over time.

The test purposes and intended uses, target population, target language use domains, and test constructs of TOEFL Junior are described in this framework. Also included is a description of the overall test structure and scoring system, demonstrating how the constructs are operationalized. Finally, we outline research topics to support the interpretive argument of the use of the test.

The generic name, TOEFL Junior, will be used inclusively in this document to refer to the two TOEFL Junior tests — the paper-based TOEFL Junior Standard test and the computer-delivered TOEFL Junior Comprehensive test — when the discussion applies equally to both tests. However, the specific name will be used when the discussion is only pertinent to that test, particularly with relation to the overall test and section structure and the scoring system of each test. The decision to develop two different versions of the test was made to reach a wider potential population of test takers and to provide stakeholders with the option to select a version that best meets their needs and serves their purposes. For example, whereas TOEFL Junior Comprehensive provides more information by measuring all four language skills, including speaking and writing (which are not measured in TOEFL Junior Standard), its cost and administration requirements may not make it the test of choice for all settings. On the other hand, TOEFL Junior Standard, exclusively consisting of selected-response items delivered by paper and pencil, has quicker score turnaround. Therefore, it can be used more flexibly, without requiring computers. Further information about how the two versions of the test differ is presented in later sections of this document.

Background

Generating a New Assessment

English proficiency is an increasingly important competence to develop for students worldwide. Mastery of English expands access to a range of educational, personal, and professional opportunities. As a result, in many education systems around the globe, English is a regular part of public school curricula. Whereas some countries introduce English

Corresponding author: Y. So, E-mail: youngsoon_so@snu.ac.kr

into the curriculum in secondary school, other public systems (and private ones as well) start English instruction at much lower grades (e.g., third grade in Korea, first grade in China). English as a foreign language (EFL) instructional programs are now attempting more ambitious learning objectives worldwide, with an emphasis on communicative language ability (cf. Bailey, Heritage, & Butler, 2014). This educational context increases the need for well-designed, objective measures of proficiency in English for young learners.

TOEFL Junior has been developed to address this need by providing much-needed information on the English language proficiency (ELP) attainment of young adolescent EFL learners worldwide.

As part of the TOEFL family of assessments, TOEFL Junior focuses on English learners' ability to communicate in an academic environment where English is the medium of instruction; that is, the test is intended to measure the communicative ability students need to participate in English-medium school settings. TOEFL Junior complements the existing university-level TOEFL assessments by assessing this proficiency at the middle school level.

English-medium instructional environments can take a range of forms, including (a) public or private schools in English-dominant countries (e.g., the United States, the United Kingdom, Canada, Australia); (b) international schools in non-English-dominant countries in which content instruction is delivered in English (e.g., International Baccalaureate World Schools); and (c) schools in any country that use either bilingual or content- and language-integrated learning approaches in which some content instruction is delivered in English. Although these instructional models are different in important respects, each calls for students to use English to learn new information in content areas. We also maintain that the traditional distinction between English as a second language (ESL) and EFL is of little importance in the aforementioned instructional models; the most relevant feature of all models is that English is used as an instructional language regardless of whether English is the language of communication outside of school. To differing degrees, these models also call for the use of English for nonacademic purposes, such as for social interactions, service encounters, and classroom management.

Proficiency for English-medium instructional environments may be aspirational for many EFL learners. For EFL learners with no specific plans to enter a program of instruction in English, TOEFL Junior will provide objective information about how their ELP relates to the standard embodied by the TOEFL Junior assessment. In providing an international benchmark for English learning, TOEFL Junior can serve as a general progress measure, providing students, parents, teachers, and schools with an objective measure of students' ELP.

Educational Significance

As the need to learn English increases, so does the need for appropriate measures to inform students of their English proficiency levels. Yet, relatively few international assessments are available for adolescent EFL students. Given the wide range of EFL contexts with varied standards and curricula, an international English proficiency assessment would be instrumental in providing some degree of standardized information about learners' proficiency levels. TOEFL Junior has been designed to measure students' proficiency levels and provide useful information about the stage of English proficiency they have attained. Students and teachers will also be informed of the various aspects of ELP needed to function in English-medium school settings. Thus, TOEFL Junior results have the potential to help English learners and their teachers set appropriate learning goals for the development of English proficiency.

Test Purpose and Intended Uses

TOEFL Junior is a measure of the English language ability of young students whose first language is not English and who are in the process of developing the proficiency required to participate in an English-medium instructional environment. The test measures language proficiency in situations and tasks representative of English-medium school contexts. Though some test tasks assess underlying enabling skills, such as grammatical and lexical knowledge, the main emphasis of the test is the measurement of communicative competence, that is, the ability to use language for communicative purposes. Test scores are intended to be used as indicators of the proficiency levels of students in the target population.

The following have been identified as appropriate intended uses of TOEFL Junior test scores for the target population: (a) to determine the ELP levels of students on the basis of their performance on tasks representative of English-medium instructional environments at the middle school level, (b) to support decisions regarding placement of students into

programs designed to increase their proficiency in academic and social English, and (c) to provide information about student progress in developing ELP over time.

Target Population

TOEFL Junior is designed for students for whom English is a foreign language and who aspire to participate in English-medium instructional environments at the middle school level. Test takers will typically range in age from 11 to 15 years. They are both male and female, with a wide variety of nationalities and native languages. Their educational backgrounds and real-world experiences will vary, but they are typically expected to have at least 5 full years of educational experience at the elementary and/or middle school level.

Identifying the Test Domains of Language Use

Identifying the characteristics of target language use (TLU) domains or situations is necessary to support the claim that test takers' performance in test tasks relates to their expected performance in real-life communicative situations. Normally, the closer the correspondence between TLU tasks and test tasks, the greater the validity of interpretations about a test taker's language proficiency based on his or her test performance (Bachman & Palmer, 1996, 2010). TLU descriptions thus provide useful guidelines for the development of item and task specifications. They can also serve as a basis for evaluating the authenticity and appropriateness of test content.

In the process of designing TOEFL Junior, a design team of Educational Testing Service (ETS) researchers, test developers, and consultants identified TLU tasks that middle school students are expected to perform in English-medium secondary school contexts by analyzing two main sources of data. First, English language standards/curricula and textbooks from Chile, China, France, Korea, and Japan were reviewed along with ELP standards for English learners in US middle schools (i.e., California, Colorado, Florida, New York, and Texas state standards and the WIDA consortium standards). Appendices A – D summarize the results of the curricula and standards reviews for each of the four language skills (listening, reading, speaking, and writing). The content for each skill has been categorized into three domains, which are discussed later in this section. Second, the existing academic literature on language used in academic contexts was reviewed. Research from the two aforementioned sources has identified important real-world tasks at the middle school level as well as skills needed to complete those tasks. It has also indicated that TLU tasks in an academic context can be categorized into three domains related to the purpose of language use. The three domains identified and considered in our test design are (a) social and interpersonal, (b) navigational, and (c) academic. In the following section, a brief summary of literature that supports our rationale for categorizing the three language use domains is provided. Next, the three domains are defined and illustrated with real-life language use examples.

Literature on the Language Demands of Academic Contexts

As mentioned earlier, the construct targeted by the TOEFL Junior test is the English ability needed to study in an English-medium middle school. Efforts to describe the language that students use in school can be traced back to Cummins's (1980, 1981) seminal work. Cummins differentiated social language ability, labeled as *basic interpersonal communication skills* (BICS), from more cognitively demanding, decontextualized language ability, which he labeled as *cognitive academic language proficiency* (CALP). Even though there have been critiques of the legitimacy of viewing language use (i.e., CALP) as decontextualized, the BICS – CALP categorization has had a significant influence on how we understand the language demands that students face in English-medium instructional environments. More importantly, Cummins's categories have spawned research that has sought evidence that academic language proficiency is distinguishable from the language proficiency needed for social and interpersonal purposes. In turn, this research has led to the definition and identification of the characteristics of academic language proficiency.

The research findings support the conclusion that the general language proficiency tests do not necessarily capture language skills needed for academic study. First, students do not necessarily perform equally well on (a) standardized content assessments (e.g., math, science, and social studies) given in English and (b) English language development (ELD) assessments mandated for all English learners attending US schools (Butler & Castellon-Wellington, 2005; Stevens, Butler, & Castellon-Wellington, 2000). Second, the language measured in ELD assessments does not adequately represent the

language used in standardized content assessments. In other words, existing ELD assessments, which many US states have used for identifying, classifying, and reclassifying English learners, have been found to be limited with respect to measuring the range of language ability required to take content assessments (Butler, Stevens, & Castellon-Wellington, 2007). Third, the language assessed in ELD assessments does not always accurately represent the language actually used in classes (Schleppegrell, 2001). These findings indicate that many widely used ELD assessments do not accurately measure the language ability required for students' participation in English-medium academic settings. If these findings support a conceptualization of academic language proficiency as distinct but related to general language proficiency, then the next question is how to characterize this ability.

Chamot and O'Malley (1994) defined academic English as "the language that is used by teachers and students for the purpose of acquiring new knowledge and skills . . . imparting new information, describing abstract ideas, and developing students' conceptual understanding" (p. 40). Although this definition provides a general concept of academic English, other researchers have explored more specific characteristics and expanded the definition of academic English. For instance, Schleppegrell (2001) identified specific linguistic features that are encountered in school-based texts (e.g., nominalizations, technical lexical choices). Scarcella (2003) further listed various features of academic English from discrete linguistic features (phonological, lexical, and grammatical features) and language functions (sociolinguistic features) to stylistic register (discourse features). In doing so, Scarcella attempted to establish a competence-based framework of academic English proficiency drawn from prior communicative competence research (e.g., Bachman, 1990; Bachman & Palmer, 1996; Canale & Swain, 1980). A fully comprehensive characterization of academic English, however, remains to be developed. Nonetheless, the evidence collected thus far shows that the difference between language used for general purposes and that used for academic purposes is relative, at both the linguistic and cognitive levels, with complex sentence structures and specialized vocabulary being used relatively more frequently in academic language (Bailey, 2007; Cummins, 2000).

However, it should be noted that the aforementioned literature on academic language proficiency does not undermine the importance of English for social and interpersonal purposes. Social language remains an important, foundational element of the language proficiency needed in school settings. Therefore, the TOEFL Junior test aims to measure the full range of language uses that students encounter in English-medium school settings. In other words, TOEFL Junior acknowledges the complex and multifaceted nature of the language that students need to learn in school contexts.

As noted previously, three domains of language use are identified and considered in the TOEFL Junior test design: social and interpersonal, navigational, and academic. These domains are based on Bailey and colleagues' extensive research on *school language* (Bailey, 2007; Bailey, Butler, LaFramenta, & Ong, 2004; Bailey & Heritage, 2008). In particular, Bailey and Heritage's (2008) tripartite categorization of school language has been found to be most consistent with what the test design team identified from its review of standards and curricula. Bailey and Heritage further divided academic English, which corresponds to CALP in Cummins's (1980, 1981) bipartite categorization, into school navigational language (SNL) and curriculum content language (CCL). They defined SNL as the language needed for classroom management, whereas CCL was defined as "the language used in the process of teaching and learning content material" (p. 15). SNL and CCL correspond to the navigational and the academic domain, respectively, in TOEFL Junior. More detailed discussion of how the three domains are defined and operationalized in the test is provided in the next section.

Target Language Use (TLU) Domains for TOEFL Junior

The TLU domain for TOEFL Junior (i.e., English-medium middle school environments) is divided into three subdomains—social and interpersonal, navigational, and academic—based on the rationales discussed in the previous section. It should be acknowledged that these three domains are fluid and cannot be clearly differentiated in all language use situations; the distinctions among the three domains can oversimplify the very complex process of language use. Note that in Figure 1, the lines representing the subdomains are dotted to symbolize the fuzzy boundaries among the domains. In addition, there is an overlap with respect to the characteristics of language required in each of the three domains. For example, there is likely to be a threshold level of grammatical knowledge that is fundamental for language use irrespective of the specific language use domain. However, despite its imperfections, we believe that this classification is effective for describing the wide range of language use activities in secondary-level English-medium school settings. Besides differing with regard to the functions or purposes of language activities, the three domains also differ in terms of the *characteristics* of language (e.g., word choice, complexity of sentence structures), which are discussed in more detail

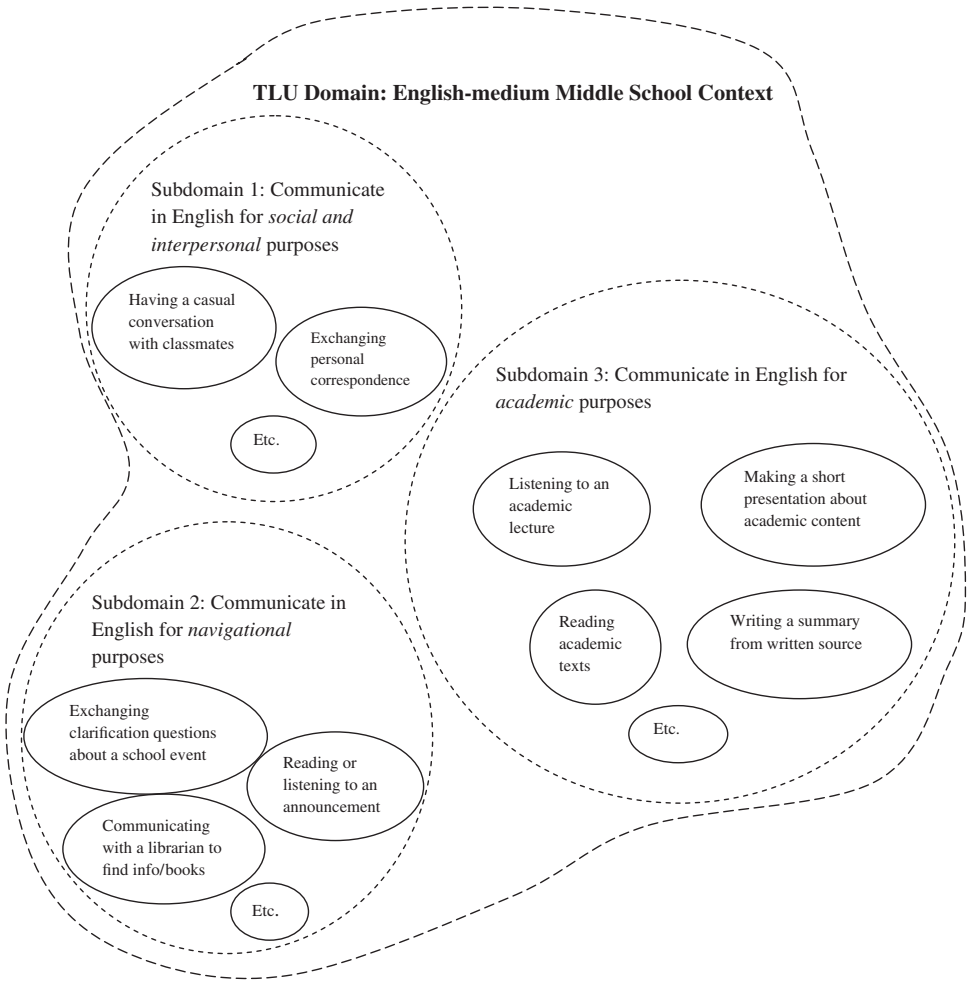


Figure 1 Defining the target language use (TLU) domain of the TOEFL Junior test.

in the section on construct definition. Finally, the academic subdomain is believed to play a more significant role than the other two domains in students' success in academic settings, and that is why the area representing the academic domain in Figure 1 is larger than those representing the other two domains. This interpretation is also reflected in the test blueprint, with more items tapping the academic domain than the other domains. Emphasizing the academic domain in the test is also believed to have a beneficial influence on test takers, motivating them to focus their language study on the areas that have been found in the academic English literature to be more difficult to master (Bailey, 2007; Cummins, 2000).

The three TLU subdomains are defined as follows.

Communicating in English for Social and Interpersonal Purposes

This subdomain encompasses uses of language for establishing and maintaining personal relationships. For example, students participate in casual conversations with their friends in school settings where they have to both understand other speaker(s) and respond appropriately. Students sometimes exchange personal correspondence with friends or teachers. The topics may include familiar ones, such as family, routine daily activities, and personal experiences. The tasks in this domain tend to involve informal registers of language use.

Communicating in English for Navigational Purposes

In school contexts, students communicate with peers, teachers, and other school staff about school- and course-related materials and activities but not about academic content. For example, students communicate about homework

assignments to obtain and/or clarify details. In some cases, they need to extract key information from school-related announcements. That is, students need to communicate to navigate school or course information. The second subdomain captures this specific purpose of communication.

Communicating in English for Academic Purposes

This subdomain entails language activities performed to learn academic content in English. Language functions such as summarizing, describing, analyzing, and evaluating are typically needed to learn academic content. The topics may be discipline related, including science, math, and social studies. Examples of this language use include comprehending ideas in lectures or class discussions, participating in short conversations about academic content in a class, comprehending written academic texts, and summarizing oral or written academic texts. Language used for this purpose typically involves more formal and technical registers with increased syntactic complexity.

Construct Definition

A Model of Language Knowledge

As discussed in the previous section, TOEFL Junior measures how successfully a test taker can complete test tasks that are designed to represent the range of communicative tasks encountered in English-medium middle schools. Among the many factors that may contribute to a test taker's success (e.g., cognitive ability, background knowledge, strategic competence), language ability—the target construct of the TOEFL Junior test—should be the main factor influencing successful test task completion.

As a framework for conceptualizing language ability, Bachman and Palmer's (2010) model of language knowledge provides the test design team with a useful framework of reference for designing individual test tasks and the test's organization. In particular, the breadth of the model makes it possible to (a) recognize the complex nature of the target construct, (b) identify specific component(s) of language knowledge that test tasks are designed to measure, (c) describe the specific features of reading/listening passages, and (d) specify the expected characteristics of the test takers' responses to speaking and writing test tasks.

As shown in Figure 2, Bachman and Palmer's (2010) model of language knowledge consists of two broad categories: *organizational knowledge* and *pragmatic knowledge*. Organizational knowledge refers to knowledge about the formal structure of a language; it is further divided into grammatical knowledge, which is needed to interpret and produce individual

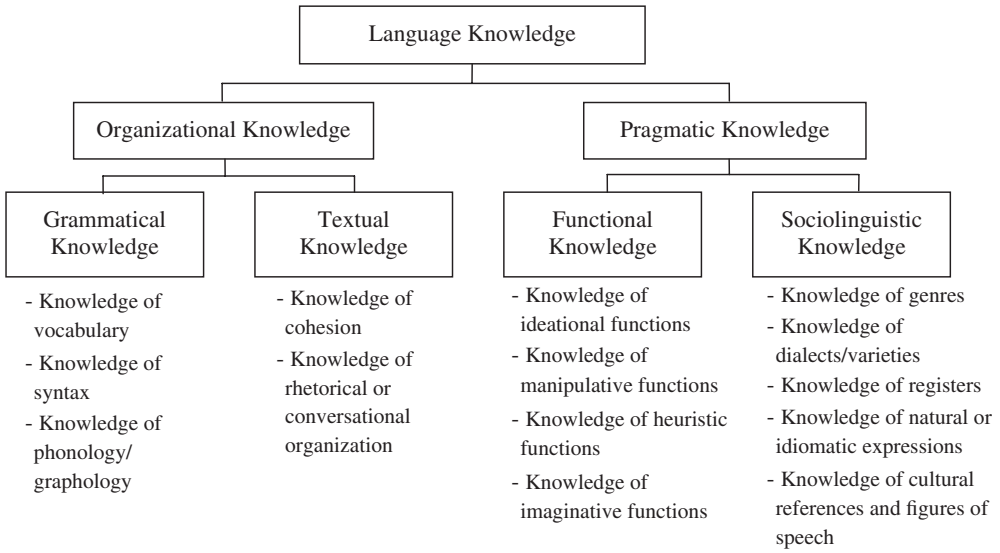


Figure 2 Bachman and Palmer's (2010) model of language knowledge. Adapted from *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*, by L. Bachman and A. Palmer, 2010, p. 45. Copyright 2010 by Oxford University Press.

sentences, and textual knowledge, which is needed to interpret and produce cohesive longer discourse. The second category, pragmatic knowledge, is the knowledge needed for a language user to produce and/or process language appropriately in relation to other variables such as the language users' intentions and situational factors. This category is further divided into functional and sociolinguistic knowledge.

It should be pointed out that not all of the areas of language knowledge in Figure 2 are considered appropriate or equally important for inclusion and measurement for the TOEFL Junior intended population. For example, knowledge of cultural references is inappropriate because it can be a source of between-group test bias. In addition, some areas of language knowledge form a fundamental basis for language users to perform certain tasks using language, whereas other areas require a certain level of mastery of the first type of knowledge to be appropriately used in context. The knowledge of words and sentence structures of a language (i.e., grammatical knowledge in the Bachman & Palmer, 2010, model) is an example of the former type of knowledge, whereas the ability to participate in a conversation appropriately by understanding the context-appropriate meaning of an utterance and responding to it appropriately (i.e., functional knowledge in the Bachman & Palmer, 2010, model) is an example of the latter type of knowledge, which requires a foundation in the former. In designing the TOEFL Junior test, the former type of language knowledge is categorized as *enabling skills* and is considered to be fundamental to any communicative language use. Therefore, except in the TOEFL Junior Standard Language Form and Meaning section (to be discussed later), enabling skills were considered in defining the language demands of communication tasks that students are likely to perform in TLU situations.

An example presented in the next section illustrates how this language knowledge model has informed the design of test tasks, and in particular, how it has helped to maximize their comparability to actual TLU tasks.

Linking Test Tasks to Target Language Use (TLU) Tasks

Upon reviewing TLU tasks that were identified through the curricula and standards review (see Appendices A – D), a set of TLU tasks was sampled to serve as the basis for the design of test tasks. Each section of the test was developed with tasks that, collectively, would provide evidence about a test taker's competence in communicating in English in all three of the TLU subdomains defined in the previous section. In operationalizing each potential task, efforts were made to ensure that the linguistic characteristics of each task stimulus (e.g., a listening passage) and its expected response (e.g., a spoken constructed-response) were as similar as possible to the language knowledge required to perform a similar task in a nonassessment situation in an English-medium middle school context, as represented in Figure 3.

The example in Figure 3 demonstrates how Bachman and Palmer's (2010) framework of language task characteristics guided test design. As illustrated in Figure 3, efforts were made to reproduce both the situational and the linguistic characteristics of the TLU tasks in the test tasks to the highest possible extent. In particular, in describing the linguistic characteristics of the input and expected responses, the test development team used the language model discussed in the previous section (Figure 2).

Organization of the Test Into Sections

A discussion of the organizational structures of the tests is provided in this section, with individual presentations of the two TOEFL Junior tests and the sections included in each. As summarized in Tables 1 and 2, two sections, listening and reading, appear in both tests, whereas other sections appear in only one of the tests. The language form and meaning section is only present in TOEFL Junior Standard, whereas the speaking and writing sections are included only in TOEFL Junior Comprehensive.

The decision to organize the test by modality (i.e., reading, listening, speaking, and writing) was made mainly because most curricula and textbooks currently in use are organized in this manner (Appendices A – D). It is expected, therefore, that stakeholders will find it useful to receive information about each modality. However, the design team also acknowledged that, in real life, multiple language modalities are often required to complete a single language use task. Hence, integrated tasks, which require multiple modalities (e.g., listening and speaking), are also included in the speaking and writing sections of the TOEFL Junior Comprehensive test. In addition, the decision was made to include the language form and meaning section in the TOEFL Junior Standard test to indirectly measure students' ability to use their knowledge of English grammar and vocabulary in speaking and writing, as these abilities cannot be easily operationalized in a constructed-response format on a paper-delivered test.

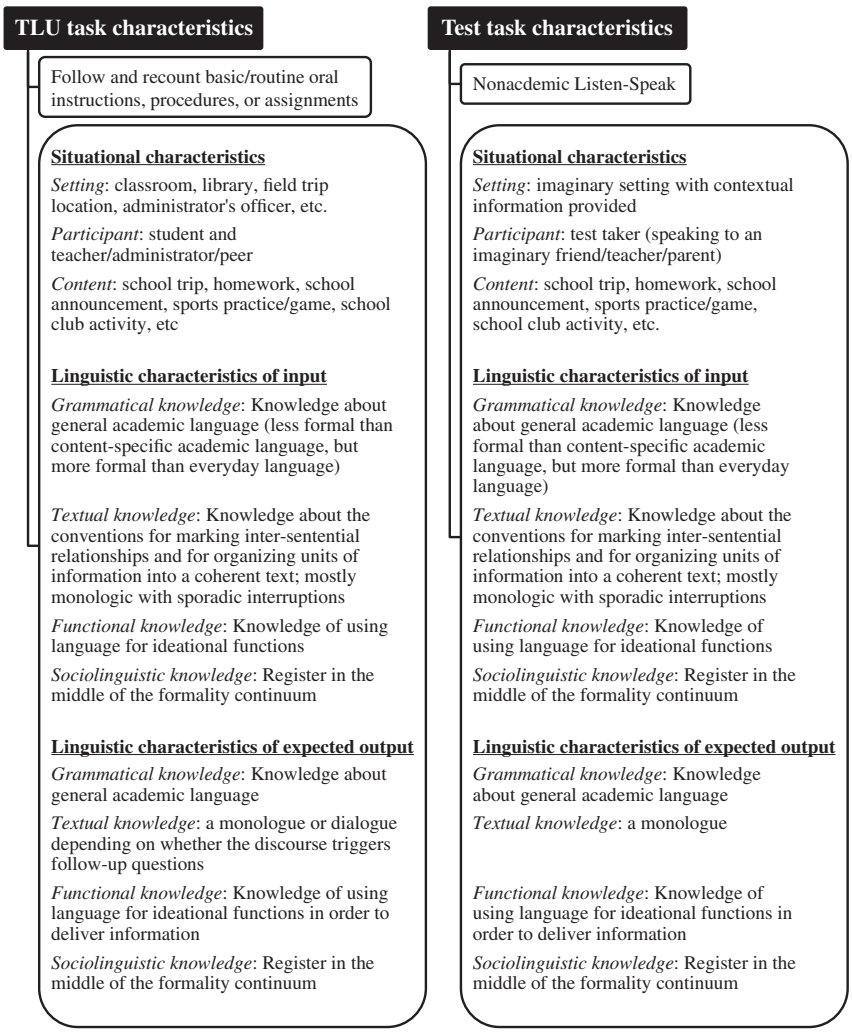


Figure 3 An example of linking a target language use (TLU) task to an assessment task.

Table 1 Overall Structure of TOEFL Junior Standard

Section	No. of items			Testing time
	Operational[a]	Variable[a]	Total	
Listening comprehension	30	12	42	40 min
Language form and meaning	30	12	42	25 min
Reading comprehension	30	12	42	50 min
Total	96	36	126	1 h 55 min

[a]The operational items are those that are considered for the official score reports of the test, whereas the variable items are those included in the test for trial purposes. In other words, students' responses to the variable items are reviewed to ensure that they can be used as operational items in the future.

Construct Definition by Section

This section presents detailed information about the definitions of the constructs for each of the test sections. The section is arranged in the order of language form and meaning, listening, reading, speaking, and writing, so that the first three sections are the ones included in the TOEFL Junior Standard test and the latter four sections (i.e., listening through writing) are the ones in the TOEFL Junior Comprehensive test. More information about how each of the two TOEFL Junior tests was operationalized is provided in the next section.

Table 2 Overall Structure of TOEFL Junior Comprehensive

	No. of items/tasks			
	Operational	Variable	Total	Testing time
Listening	28	8	36	35 min
Reading	28	8	36	40 min
Speaking	4	n/a	4	25 min[a]
Writing	4	n/a	4	40 min[a]
Total	64	16	80	140 min

[a]The testing time includes both administration time, which allows test takers to process the stimulus input and prepare for their responses, and response time, when test takers produce their responses.

Language Form and Meaning

This section, included in the TOEFL Junior Standard only, is differentiated from other sections in the TOEFL Junior test in that test items in the section aim to measure enabling skills required for communication, whereas items and tasks in the other sections measure the ability to apply such enabling skills in actual communicative tasks. Specifically, the items in this section assess the degree to which students can identify the structure of English and choose appropriate lexical units. The items are presented as gap-filling questions within the context of a cohesive paragraph. Therefore, students are required to take into account the context of an entire passage to answer the questions appropriately in the sections.

It should also be noted that this section intends to indirectly measure students' ability to use their grammar and vocabulary knowledge for communication in a test where the productive skills (i.e., speaking and writing) are not directly measured. In other words, the ability measured in this section has an association, at least to some extent, with students' ability to apply such knowledge of English grammar and vocabulary to speaking and writing tasks.

The items are divided into two categories: items targeting language meaning and items targeting language form. As explained in the following, vocabulary and grammar knowledge was measured in the context of a single paragraph, with the justification that the model of language knowledge (see A Model of Language Knowledge and Figure 2) can be better operationalized in a rich context than through decontextualized, individual sentences:

1. *The ability to identify an appropriate lexical item within context.* Students must be able to identify a word that semantically completes a sentence within the context of a paragraph.
2. *The ability to recognize a proper grammatical structure within context*. Students must be able to identify a proper structure needed to complete a grammatically accurate sentence in English.

Listening

TOEFL Junior assesses the degree to which students have the listening skills required to function in English-medium instructional environments. In such contexts, students are exposed to a wide range of aural input, for example, from personal conversations to lectures on academic content. Features specific to spoken discourse that distinguish it from written discourse include repetition, relatively complex verb structures, relatively little nominalization, and occasional performance disfluencies. Therefore, it is essential for successful participation in school that students gain familiarity with spoken discourse features and attain listening proficiency sufficient to comprehend different genres of spoken discourse. Moreover, to succeed in school, students need to understand the main ideas and important details, make inferences based on what is implied but not explicitly stated, make predictions based on what the speaker says, understand a speaker's purpose, and correctly interpret such features of prosody as intonation and contrastive stress. Three types of listening ability were defined to capture these skills and language features:

1. *The ability to listen for social and interpersonal purposes*. Students must be able to comprehend conversations on familiar topics about day-to-day matters that take place in a school setting, such as sharing experiences with their peers.
2. *The ability to listen for navigational purposes.* Students must be able to comprehend the language that teachers and other school staff produce for a range of purposes other than presenting academic content. This includes language that takes place both inside and outside of the classroom (e.g., in the school library or auditorium or on field trips)

and that fulfills a range of speech functions (e.g., providing directions, making announcements, giving reminders, issuing invitations, giving warnings).

3. *The ability to listen for academic purposes*. Students need to comprehend ideas presented in a lecture or discussion based on academic material. Though TOEFL Junior requires students to comprehend oral input such as that needed to learn new ideas in an English-medium classroom, it does not require subject-specific background knowledge in any given content area. In the domain of science, for example, such speech includes key terms, structures, and concepts that enable middle school students to access academic content (terms such as evidence and investigation, concepts such as making observations, reports on the results of an experiment, and a range of structures for expressing these concepts), but does not include specific content or concepts that would be taught as part of a specific science curriculum (e.g., photosynthesis or geotropism). However, it is construct-relevant to include such concepts in the assessment if they are presented, explained, and reinforced so that a proficient listener can learn their meanings from the academic speech contained in the stimulus.

Reading

TOEFL Junior assesses the degree to which students have mastered the reading skills required for English-medium instructional environments. The review of English language curricula and language objectives in reading (Appendix B) indicates that a wide range of reading subskills are expected of students, including understanding main ideas, identifying important details, and making inferences. In addition, the curricula and standards specify different types of text. A relationship was observed between text types and the three TLU subskills. Therefore, the three reading abilities to be measured in TOEFL Junior are defined as follows, according to text type:

1. *The ability to read and comprehend texts for social and interpersonal purposes*. Students should be able to read and comprehend written texts on familiar topics in order to establish or maintain social relationships. Text types for this purpose may include correspondence (e.g., e-mail, letters) and student writing. In addition, reading for personal pleasure (e.g., novels, periodicals) is included in this category.
2. *The ability to read and comprehend texts for navigational purposes.* Students need to be able to read and comprehend texts in order to identify key information from informational texts for future reference. Such texts include those containing school-related information, usually in less linear formats (e.g., directions, schedules, written announcements, brochures, and advertisements). Reading subskills that are particularly relevant to this type of reading include comprehending explicit meaning, identifying key information, and understanding steps and procedures.
3. *The ability to read and comprehend academic texts*. Students need to be able to read and comprehend academic texts in a range of genres (e.g., expository, biographical, persuasive, literary) across a range of subject areas (e.g., arts/humanities, science, social studies). They need to be able to read such texts at difficulty levels up to and including those typical of what is used in English-medium classrooms. In reading these texts, students need to be able to understand the main ideas and the key supporting information, to make inferences based on what is implied but not explicitly stated, and to understand key vocabulary (either from previous knowledge or from context) and cohesive elements within the text (i.e., referential relationships across sentences). Depending on the nature of the specific text, students may also need to understand an author's purpose, follow the logic and the intended meaning of basic rhetorical structures, and/or identify and understand figurative language. As with listening, reading texts will not require any specific background or prior knowledge but will sometimes require students to read in order to learn new information in an academic context.

Speaking

TOEFL Junior Comprehensive assesses the degree to which students have the speaking skills required by English-medium instructional environments. This includes three abilities:

1. *The ability to use spoken English for social and interpersonal purposes*. Students must be able to communicate orally in routine tasks and situations encountered in the school environment. For example, this includes the ability to communicate personal information, needs, and opinions on a wide range of familiar topics (such as hobbies, food, weather, and extracurricular events).

2. *The ability to use spoken English for navigational purposes to exchange classroom-related information*. Students must be able to engage in discussions and interactions on topics related to learning activities. This includes the ability to make requests, ask for assistance or information, participate in group activities, and convey simple directions and instructions.
3. *The ability to use spoken English for academic purposes to communicate about and demonstrate knowledge of academic course content*. Students must be able to participate in classroom activities to convey academic knowledge. This includes the ability to respond to oral questions about academic content and to convey information heard or read in an academic context.

Writing

TOEFL Junior Comprehensive assesses the degree to which test takers have the writing abilities required by English-medium instructional environments at the middle school level. This includes three types of ability:

1. *The ability to write in English for social and interpersonal purposes*. In English-medium instructional environments, students must be able to engage in written communication for the purposes of establishing and maintaining social and interpersonal relationships. This includes the ability to write effective informal correspondence to peers or teachers and the ability to recount events based on personal experience and observation.
2. *The ability to write in English for navigational purposes*. In school settings, students must be able to extract key school-related information from a variety of spoken or written stimuli and keep written records for future reference. For instance, students may need to take notes while listening to their teacher explain a class assignment or the steps of a science experiment. Students may also need to write simple, short summaries of school-related information (e.g., a field trip, announcements, directions, or procedures).
3. *The ability to write in English for academic purposes*. In English-medium instructional environments, students must be able to communicate in writing using appropriate written language on subject matters representing a range of content areas and genres. This includes the ability to produce connected text; to describe a process in an academic context; to understand and be able to summarize, synthesize, and paraphrase important and relevant information from spoken and written stimuli; and to integrate information from multiple academic spoken and/or written stimuli.

Operationalizing the Construct

In this section, the overall structures of TOEFL Junior Standard and TOEFL Junior Comprehensive are described, followed by a more detailed explanation of the structure of each section of the tests. In particular, this section describes how the constructs, the TLU subdomains, and the tasks are operationalized in TOEFL Junior.

Overall Structure of the Test

The overall structures of TOEFL Junior Standard and TOEFL Junior Comprehensive are presented in Tables 1 and 2, respectively, with information on the sections included and the number of items/tasks and the allotted time in different sections.

As briefly discussed in the introduction to this report, TOEFL Junior Standard consists of all selected-response questions and is delivered in paper-and-pencil format. On the other hand, TOEFL Junior Comprehensive is administered on a computer and consists of both selected-response and constructed-response questions. The receptive skills (i.e., listening and reading) are measured through selected-response questions and the productive skills (i.e., speaking and writing) are measured through constructed-response questions.

In each section of TOEFL Junior, with the exception of the language form and meaning section in TOEFL Junior Standard, items are selected to tap into the target construct in each of the three TLU subdomains: social and interpersonal, navigational, and academic. Details on the section structures in relation to the TLU subdomains and specific language skills are described in the following section.

Table 3 Structure of Language Form and Meaning Section

	No. of operational items
Language meaning	8 – 14
Language form	16 – 22
Total	30

Section Structures: Language Form and Meaning, Listening, Reading, Speaking, and Writing

This section is divided into five subsections, each of which focuses on one of the five sections appearing in TOEFL Junior: language form and meaning, listening, reading, speaking, and writing. The first three sections are included in TOEFL Junior Standard, and the latter four sections (starting from listening) are in TOEFL Junior Comprehensive. Because the listening and reading sections appear in both of the TOEFL Junior tests, they are discussed only once. However, it should be noted that there are slight differences in operationalizing the two sections in the two tests, such as the number of items. These differences are summarized in tables wherever applicable.

Language Form and Meaning

In this section, test takers are given reading passages in which words have been purposefully deleted so that students must fill in the blanks by choosing an answer among four options to complete the text appropriately. The passages can be one of the following types that middle school students are likely to encounter in their school lives: announcement, correspondence, advertisement, biographical, expository, or fiction.

Each reading passage contains four to eight items, depending on the type and length. Longer passages — usually expository, biographical, and fictional narrative texts — are eight to nine sentences in length and support six to eight items. Shorter passages (e.g., announcement, correspondence, and advertisement) are four to five sentences in length and support four questions. All items in this section measure knowledge of language meaning or form. The number of items targeting each of these constructs is summarized in Table 3.

The language meaning items ask students to choose, from a set of four options, the one correct word that semantically completes a sentence within the context of a passage. The language form items test a student's ability to recognize the proper structures needed to complete a grammatically accurate sentence in English. Both types of items, collectively, encompass a wide range of English vocabulary and grammar by including items targeting a variety of language categories. The vocabulary items encompass different parts of speech (noun, verb, adjective, adverb, determiner, conjunction, and preposition), and the grammar items include questions about sentence structure (e.g., correct subject and object forms, subject – verb agreement), verb form (e.g., tense and aspect), passive/active voice, relative clauses, word order, and comparative/superlative forms. Including a variety of language aspects was considered important in test design to ensure that students have a broad understanding of the English language. The language features measured in the section were chosen from among those that are taught in English curricula, and their difficulty was gauged by expert judgment.

Listening

In this section, test takers listen to aural stimuli and answer four-option multiple-choice questions presented after each stimulus. The number of questions per stimulus varies depending on the type of stimulus — three or four questions for short conversation stimuli, one question for classroom instruction stimuli, and four questions for academic listening stimuli. Note that the number of items of each type varies in TOEFL Junior Standard and TOEFL Junior Comprehensive, as presented in Table 4.

Table 4 summarizes the relationships among stimulus type, TLU domain, and the subskills to be measured by a stimulus. As shown in the first two columns of the table, there is a one-to-one correspondence between stimulus type and the TLU subdomain that each stimulus type is targeting. The short conversations, classroom instruction, and academic listening stimuli are intended to measure test takers' ability to communicate for social and interpersonal, navigational, and academic subdomains, respectively.

Table 4 Listening Section Structure

Stimulus/input	Target language use subdomain	Subskills measured	No. of operational items	
			Standard	Comprehensive
Short conversations	Social and interpersonal	• Comprehending the main idea • Identifying salient details • Making inferences • Making predictions • Identifying speaker's purpose • Understanding a meaning conveyed by prosodic features	11 – 12	12
Classroom instruction (monolog)	Navigational	• Comprehending the main idea • Identifying salient details • Making inferences • Making predictions	6 – 7	8
Academic listening (monolog/discussion)	Academic	• Comprehending the main idea • Identifying salient details • Making inferences • Making predictions • Identifying speaker's purpose	12	8
Total			30	28

In addition, each listening item aims to measure either (a) one of the common subskills or (b) one of the domain-specific subskills. The common subskills refer to listening abilities that can be measured in any of the three TLU subdomains. For example, a question about main idea can be based on a listening stimulus in any of the TLU subdomains. The domain-specific subskills are operationalized to be measured in one or two of the TLU subdomains only. Specifically, the ability to identify speaker's purpose is operationalized for both the social and interpersonal and academic domains, whereas the ability to understand a meaning conveyed by prosodic features is operationalized exclusively for the social and interpersonal domain.

Reading

In the reading section, as in the listening section, test takers are presented with reading materials and then with four-option multiple-choice questions. As summarized in Table 5, each stimulus type taps into one of the three TLU subdomains. In addition, each reading comprehension item is designed to measure one of the seven common subskills, which are listed in the third column of the table. Finally, as shown in the last column of Table 5, some stimulus types may not be included in a given operational test form. The single exception is the expository stimulus type: Every operational form includes two eight-item sets, each with an expository stimulus.

Speaking

The speaking section consists of four tasks, as summarized in Table 6. In each task, the total time, shown in the last column of the table, represents time provided for test takers to (a) process the stimulus input, either linguistic, nonlinguistic, or both; (b) prepare for their responses (i.e., preparation time); and (c) record their responses (i.e., speaking time).

As shown in Table 6, each speaking task is designed to measure the test takers' ability to communicate in one of the three TLU subdomains. It should be noted that all of the tasks except the picture narration task require test takers to understand language input, either written or spoken, to successfully complete the task, as shown in the integrated skills column in

Table 5 Reading Section Structure

Stimulus/input	Target language use subdomain	Subskills measured	No. of operational items	
			Standard	Comprehensive
Correspondence	Social and interpersonal	• Comprehending the main idea	0 or 4	0 or 4
Nonlinear text	Navigational	• Identifying important supporting factual information	0 or 4	0 or 4
Journalism	Navigational	• Making inferences	0 or 8	0 or 8
Expository	Academic	• 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	16	16
Total			30	28

Table 6 Speaking Section Structure

Task	Target language use subdomain	Integrated skills	Preparation time	Speaking time	Total time
Read Aloud	Academic[a]	Reading, speaking	1:00	1:00	3:30
Picture Narration	Social and interpersonal	n/a	1:00	1:00	3:20
Nonacademic Listen – Speak	Navigational	Listening, speaking	0:45	1:00	4:30
Academic Listen – Speak	Academic	Listening, speaking	0:45	1:00	4:30

[a]Opinions may differ with respect to the appropriateness of categorizing the read aloud task as academic, because this task does not tap directly into communicative skills but rather targets enabling skills (e.g., accuracy of pronunciation and intonation and fluency) that form the basis for all speaking tasks. While acknowledging this perspective, the read aloud task is categorized as academic in the test design framework because the classroom is the most common context in which students are asked to read text aloud. In other words, this task is one of the important tasks that students are commonly expected to perform in an academic context.

Table 7 Writing Section Structure

Task	Target language use subdomain	Integrated skills	Writing time	Total time
Editing	Navigational/Academic	Reading, writing	5:00	5:30
E-mail	Social and interpersonal	Reading, writing	7:00	7:30
Opinion	Social and interpersonal/Academic	n/a	10:00	10:30
Listen – Write	Academic	Listening, writing	10:00	14:30

Table 6. This was a conscious decision intended to ensure that three of the four tasks measure integrated language skills for communication, better reflecting language use in the real world.

Writing

The writing section consists of four tasks. The tasks and the time allowed for each task are summarized in Table 7. In this section, the total time includes both time for test takers to process the stimulus input and time to produce their written responses. Unlike in the speaking section, time for test takers to prepare for their responses is not separately assigned in the writing section; instead, test takers use their response time for planning their writing (e.g., outlining), composing their responses, and finally, proofreading what they have written.

Table 8 Scores on the TOEFL Junior Standard Score Report

Section/overall	Reported score range (increments of 5 for the section scale scores)	CEFR level and can-do statements	Additional information
Listening	200 – 300	Below A2, A2, B1, B2	Lexile score 510L – 1150L
Language form and meaning	200 – 300	Below A2, A2, B1, B2	n/a
Reading	200 – 300	Below A2, A2, B1, B2	n/a
Overall score level	1 – 5	n/a	Overall performance descriptor and CEFR profile for the three sections

Note. CEFR = Common European Framework of Reference.

Table 9 Scores on the TOEFL Junior Comprehensive Score Report

Section/overall	Reported score range (increments of 1)	CEFR level and can-do statements	Additional information
Reading	140 – 160	Below A2, A2, B1, B2	Lexile score 510L – 1150L
Listening	140 – 160	Below A2, A2, B1, B2	n/a
Speaking	0 – 16	Below A2, A2, B1, B2	n/a
Writing	0 – 16	Below A2, A2, B1, B2	n/a
Overall score level	1 – 6	n/a	Overall performance descriptor and CEFR profile for the four skills

Note. CEFR = Common European Framework of Reference.

As in the speaking section, each writing task is designed to measure the test takers' ability to communicate in one or two of the three TLU subdomains, and three of the four writing tasks require the integration of language skills for their successful completion.

Scoring System

This section describes how each of the scores included on the score report were developed to provide reliable, meaningful, and accessible information about test takers' performance. In developing scores, the following considerations were taken into account: current practices in establishing score scales, results from the pilot study, and potential uses of the reported scores.

A score report for both the TOEFL Junior Standard test and the TOEFL Junior Comprehensive test contains the following information: overall score level, section scores for each of the sections, a Common European Framework of Reference (CEFR; Council of Europe, 2009) level for each test section, can-do statements that describe what students can typically do at the scored CEFR level, and a Lexile score on the reading section. The can-do statements included in the score reports are adapted from the CEFR can-do statements (Council of Europe, 2009) and modified to make them more appropriate for the language use required for the target age group of the test. See Appendix E for a sample TOEFL Junior Comprehensive score report. In addition, Tables 8 and 9 summarize the scores that are provided on the score reports for TOEFL Junior Standard and TOEFL Junior Comprehensive, respectively.

As summarized in Tables 8 and 9, the CEFR levels reported for each test section represent four levels: below A2 (the lowest performance level measured by the test), A2, B1, and B2 (the highest performance level measured by the test). These levels were established through standard-setting studies that ETS conducted separately for the two TOEFL Junior tests.[1] Finally, for the reading section, another auxiliary score, the Lexile measure, is reported. The Lexile score is provided so that a student can easily identify reading materials at an optimal level of difficulty to improve his or her reading skills. Information about the relationship between performance on the TOEFL Junior Reading section and the Lexile measure can be found in MetaMetrics (2012).

In the next three sections, more detailed explanations are provided for the following three test development procedures: (a) section scores, (b) overall score levels and performance descriptors, and (c) scoring rubrics for the speaking and writing tasks. It should be noted that the last subsection, which is about the scoring rubrics, is relevant only to TOEFL Junior Comprehensive.

Table 10 Number of Items, Raw Scores, and Scale Scores for the Two TOEFL Junior Tests

		Raw score		Scale score	
Section	No. of items	Range	Increments	Range	Increments
TOEFL Junior Standard					
Listening	30	0–30	1	200–300	5
LFM[a]	30	0–30	1	200–300	5
Reading	30	0–30	1	200–300	5
TOEFL Junior Comprehensive					
Reading	28	0–28	1	140–160	1
Listening	28	0–28	1	140–160	1
Speaking	4	0–16	1	n/a	n/a
Writing	4[b]	0–16	0.5	n/a	n/a

[a]Language form and meaning. [b]One of the four writing tasks—editing—has two individual items. The average of the scores from the two items is the score for the editing task. This procedure results in 0.5 increments in the raw writing scores.

Section Scores

Section Raw Scores

Table 10 summarizes information about the number of items in each section and the range and increments of raw and scaled scores. For the sections composed of selected-response items—language form and meaning, listening, and reading—test takers earn one score point for each item answered correctly, while no points are earned for incorrect responses or no response at all. As indicated in the table, the raw scores (i.e., the number of items answered correctly) are converted to scaled scores (discussed in the next section of this report), and only scaled scores are included in the score report. For the speaking and writing sections that consist of four constructed-response tasks each, each response is scored by a human rater on a holistic rubric scale of 0 to 4 (discussed in the section titled Scoring Rubrics of the Speaking and Writing Tasks). In particular, with reference to the descriptions in the scoring rubrics, the meanings of raw scores in the speaking and writing sections can be more easily interpreted than can the meanings of raw scores on the selected-response items. Therefore, it was deemed unnecessary to convert speaking and writing scores into scaled scores, and raw scores are reported for these two sections.

Considerations for Scaled Score Development

It is a common assessment practice that scaled scores, instead of raw scores, are reported in order to ensure that scores are comparable across test forms that may not have the same difficulty level (Kolen, 2006). As a best practice, scaled scores are created from raw scores with appropriate statistical adjustments for form difficulty; this enables scaled scores to hold their meaning over time and across different test forms. A variety of guidelines have been discussed in educational measurement literature about best practices for creating appropriate and meaningful scaled scores (Dorans, 2002; Kolen, 2006). The following essential guidelines were considered in creating scaled scores for TOEFL Junior:

1. Use distinctive scales that do not overlap with other scales, either between the two TOEFL Junior tests or with any other ETS tests, to avoid confusion and misuses.
2. Make every item or raw score point in the meaningful raw score range count toward a scaled score point, if possible, to avoid loss of information that results from converting multiple raw score points to a single score point on the scale.
3. Ensure that for every scaled score point, there is at least one item or one raw score point to avoid the unjustified differentiation of test takers.

It is worth emphasizing that the first point was considered particularly important in the score scale development for the two TOEFL Junior tests. As discussed in the previous sections of this report, the two versions were developed to provide stakeholders with options to choose from as suited to their needs and purposes. However, we did not want the test scores from one version to be misinterpreted or misused in contexts where the use of the other version seemed more appropriate. This consideration provided the main rationale for developing different score scales for the two TOEFL Junior tests.

In light of these considerations, scales for the selected-responses sections in the TOEFL Junior Standard and Comprehensive tests were developed. One difference in the scales is that the resulting scaled scores range from 200 to 300 in increments of 5 in the TOEFL Junior Standard test, whereas they range from 140 to 160 in increments of 1 in the TOEFL Junior Comprehensive test (see Table 10). Scores on any new test form will be equated and then reported on their respective scales.

Determining the Speaking and Writing Scales

The speaking and writing sections each have four constructed-response items. Being few in number, these items are susceptible to memorization. This means that pretesting the constructed-response items would pose a test security risk. Consequently, conventional score equating that requires pretesting of items is not feasible for constructed-response items. In many testing programs that use constructed-response items only, conventional score equating is not performed. Instead of conventional score equating, quality control is maintained by trying out new items in small-scale sessions before they are used in the test,[2] as well as through rigorous training of human raters and monitoring of their performance. These quality control methods are used to ensure quality and stability in the meaning of scores for the speaking and writing sections of the TOEFL Junior Comprehensive test.

Because the speaking and writing section scores will not be equated, the scores are not strictly comparable, psychometrically speaking, across test forms, despite the aforementioned quality control measures that have been put in place. To avoid any incorrect impression on the part of stakeholders that the speaking and writing scores are comparable across forms, as the reading and listening section scores are, it was decided that the speaking and writing scales would be made clearly distinguishable from the reading and listening scales. In addition, to maximize the interpretability of the speaking and writing scales, speaking and writing scores are reported so as to be clearly associated with the performance levels that the scoring rubrics describe.

Both the speaking and writing scaled scores range from 0 to 16 in increments of 1. The four previously mentioned guidelines were followed in setting the scales. For speaking, each scaled score is associated with one and only one raw score. For writing, half points are rounded to the next higher whole number when calculating scaled scores (e.g., raw score 3 is set to scaled score 3; raw score 3.5 is set to scaled score 4). Because each speaking and writing response is scored on a 0–4 rubric scale (see Appendices F and G for scoring rubrics for the speaking and writing tasks) and the section score is the sum of the four item scores, dividing a scaled score by 4 yields a value that is compatible with the average item score; the corresponding scoring rubrics for this average item score may assist in understanding the typical characteristics of performance at this average item score level.

Overall Score Levels and Performance Descriptors

Based on the section scores explained earlier, total scaled scores were calculated, by either summing the section scores (TOEFL Junior Standard) or developing a different total score scale (TOEFL Junior Comprehensive). However, there is a limit to the amount of information that a numeric, total scaled score can provide about a test taker's language performance across different sections of a test. This fact becomes particularly clear in light of the fact that many possible combinations of section scores could arrive at the same total scaled score. To overcome this limitation of total scaled scores, it was decided that overall score levels would be reported instead. The overall score levels are band scores, as discussed in the next subsection. They are intended to help test users better understand the test results and better interpret their meanings. The following two steps were followed in developing the overall score levels and level descriptors: (a) developing band levels and (b) developing performance descriptors. More details about the procedures can be found in Papageorgiou, Morgan, and Becker (2014) for the TOEFL Junior Standard test and in Papageorgiou, Xi, Morgan, and So (in press) for the TOEFL Junior Comprehensive test.

Developing Overall Score Levels

The main goal of this step was to determine the number of overall score levels and to set cut scores to classify test takers into levels both meaningfully and reliably. In the process, the following criteria were applied for TOEFL Junior Standard and TOEFL Junior Comprehensive, respectively. We note that the types of data considered were different, primarily because

Table 11 Overall Score Levels, Performance Descriptors, and Common European Framework of Reference (CEFR) Profiles for TOEFL Junior Standard

Overall score level	Label	Overall performance descriptor	CEFR profile
		These descriptions represent performance in middle schools which use English for instruction. A typical student at this level:	A typical student at this level achieved these section-level CEFR scores:
5	Superior	Consistently demonstrates comprehension of complex written and spoken materials, drawing on knowledge of complex language structures and vocabulary.	B2 for all sections
4	Accomplished	Often demonstrates comprehension of complex written and spoken materials, drawing on knowledge of complex language structures and vocabulary.	B1 for all sections
3	Expanding	Demonstrates comprehension of some complex written and spoken materials and most basic materials, drawing on knowledge of basic language structures and vocabulary.	Mostly B1 for all sections, but occasionally A2
2	Progressing	Occasionally demonstrates comprehension of basic written and spoken materials, drawing on knowledge of basic language structures and vocabulary.	Mostly A2 for all sections, but occasionally A1 for reading and listening
1	Emerging	Can comprehend some very basic written and spoken texts, drawing on knowledge of basic language structures and vocabulary, but needs to further develop these language skills and comprehension abilities.	Mostly A1 for listening and reading; mostly A2 for language form and meaning

of the difference in structure between the two tests. However, the general procedures for the development of band levels were the same across the two tests.

In the development of overall score levels for the TOEFL Junior Standard test, which happened after the development of these levels for the TOEFL Junior Comprehensive test, it was decided that the number of overall score levels for the two TOEFL Junior tests should differ so as to prevent any misuse of the results, such as making direct comparisons between the score levels of the two tests (see the section titled Considerations for Scaled Score Development). The scores of 4,977 students who took one of the two operational test forms of TOEFL Junior Standard in 2012 were used to develop the overall score levels.

For TOEFL Junior Comprehensive, the following data, collected from the 2,931 students who participated in the 2011 TOEFL Junior Comprehensive pilot administrations, were taken into consideration: (a) the means and standard deviations of the total scaled scores for each raw score point on the speaking and writing sections; (b) the means and standard deviations of the listening and reading section scores for each raw score point on the speaking and writing sections; and (c) the CEFR profiles of the four sections for each total scaled score—this information was also collected from a separate standard-setting study that set TOEFL Junior Comprehensive cut scores for the CEFR levels.

For each of the tests, three proposals were developed to set the number of overall score levels and cut scores, and then the reliability of each proposal was estimated using RELCLASS (Livingston & Lewis, 1995). In addition, the CEFR profiles of the band levels for each solution were examined to provide an initial understanding of how proficiency progresses from lower to higher bands. A five-score-level solution (Table 11) and a six-score-level solution (Table 12) were finally selected for TOEFL Junior Standard and TOEFL Junior Comprehensive, respectively.

Developing Overall Score-Level Performance Descriptors

After making final decisions about the overall score levels for each of the TOEFL Junior tests, assessment specialists and researchers collaborated to develop performance descriptors that capture a typical student's language proficiency within each overall score level. Following is the information that was taken into account in developing the performance descriptors: (a) the means and standard deviations of each of the test sections by overall score level; (b) the characteristics

Table 12 Overall Score Levels, Performance Descriptors, and Common European Framework of Reference (CEFR) Profiles for TOEFL Junior Comprehensive

Overall score level	Label	Overall performance descriptor	CEFR profile
		These descriptions represent performance in middle schools, which use English for instruction. A typical student at this level:	A typical student at this level achieved these section-level CEFR scores:
6	Excellent	Consistently demonstrates the skills needed to communicate successfully at a high level in complex interactions and while using complex materials.	B2 for all sections
5	Advanced	Often demonstrates the skills needed to communicate successfully at a high level in complex interactions and while using complex materials.	B1 or B2 for reading and listening; B1 for speaking and writing
4	Competent	Demonstrates the skills needed to communicate successfully in some complex situations and in most simple interactions and while using basic materials.	B1 for reading and listening; B1 or A2 for speaking and writing
3	Achieving	Usually demonstrates the skills needed to communicate successfully in simple interactions and while using basic materials.	A2 or B1 for listening; A2 for reading, speaking, and writing
2	Developing	Occasionally demonstrates the skills needed to communicate successfully in simple interactions and while using basic materials.	A2 for reading and listening; below A2 for speaking and writing
1	Beginning	Demonstrates some basic language skills but needs to further develop those skills in order to communicate successfully.	Below A2 for all sections

of reading and listening items answered correctly by students at different levels; (c) the test performance of US middle school students (both English learners and native English speakers), reported in Wolf and Steinberg (2011); (d) descriptors of the proficiency scales of the CEFR to which the test scores are mapped; (e) typical profiles of students across the test sections; and (f) the rubrics used to score the writing and speaking tasks (TOEFL Junior Comprehensive only). Tables 11 and 12 summarize the results of the procedures used to define meaningful and reliable overall score levels with reference to the total scaled scores and to develop performance descriptors for each of the overall score levels for TOEFL Junior Standard and TOEFL Junior Comprehensive.

The Relationship of Overall Score Levels Between the Two TOEFL Junior Tests

Despite the potential usefulness, relative to numeric scores, of reporting overall score levels and accompanying performance descriptors, there exists a potential for misuse of the score levels. One of these potential misuses would be to claim that results from the two TOEFL Junior tests are equivalent. To prevent this unjustified use, different numbers of overall score levels (five for TOEFL Junior Standard and six for TOEFL Junior Comprehensive) were developed for the two TOEFL Junior tests, as discussed earlier. In addition, empirical evidence was collected to illustrate why the aforementioned misuse is not warranted. Table 13 shows the relationship of the overall score levels between the tests. The results in the table were produced as part of the study that developed the overall score levels for TOEFL Junior Standard (Papageorgiou et al., 2014).

What needs to be emphasized, as shown in the table, is that there is not a one-to-one correspondence in the overall score levels between the two tests. Instead, there is a probabilistic relationship between the overall score levels of the two tests. For example, for students who received the highest overall score level (Level 5) on the TOEFL Junior Standard, half of them are projected to receive Level 6 (the highest level on TOEFL Junior Comprehensive), while the remaining students are projected to obtain either Level 5 or 4. Furthermore, as explained in previous sections, the two TOEFL Junior tests measure different constructs and are composed of different sections with different structures.

Table 13 Percentage in Each TOEFL Junior Comprehensive Overall Score Level Conditional on TOEFL Junior Standard Overall Score Level

TOEFL Junior Standard overall score level	TOEFL Junior Comprehensive overall score level					
	6	5	4	3	2	1
5	50%	33%	15%	0%	1%	0%
4	4%	33%	54%	7%	1%	1%
3	1%	4%	49%	36%	9%	1%
2	0%	0%	6%	39%	41%	14%
1	0%	0%	1%	6%	28%	65%

Note: Adapted from *Development of Overall Score Levels and Performance Descriptors for the TOEFL Junior Standard Test*, by S. Papageorgiou, R. Morgan, and V. Becker, 2014.

For these two reasons, overall score levels should not be compared directly between the two tests. Rather, stakeholders should choose the test that best fits their needs and interests. For example, if the primary need of a score user is to track the developmental progress of students in a language learning program that values the balanced development of all of the four language skills, TOEFL Junior Comprehensive would be expected to provide more useful information for this specific use.

Scoring Rubrics of the Speaking and Writing Tasks (TOEFL Junior Comprehensive Only)

The speaking and writing scoring rubrics were developed in a multistage process. A small-scale prototype study was conducted with English learners in the United States in 2010 to trial prototype items and gather indicators of different levels of performance on the items. Experts experienced in evaluating the speaking and writing abilities of nonnative English speakers (e.g., *TOEFL iBT*® test certified raters) analyzed responses to the prototype items, and the results were used to formulate descriptors for the initial sets of scoring rubrics. Pilot study results were then used to further refine the scoring rubrics for each of the tasks and to establish benchmark and calibration samples for rater training. It should also be noted that speaking and writing ability, respectively, were considered the constructs to be measured and scored in the integrated speaking and writing items (see Tables 6 and 7). In other words, to avoid cases in which listening or reading stimulus comprehension difficulty compromises test takers' ability to complete the integrated tasks, the reading and listening stimuli of the integrated items were written so as to be lower in comprehension difficulty than the texts used as stimuli in the listening and reading sections.

Developing Scoring Rubrics for the Speaking Tasks

The scoring rubrics with which a test taker's spoken responses are to be evaluated were developed in three stages. First, test takers' responses representing a wide range of speaking proficiency levels were sampled from responses collected during the prototyping stage. Second, raters with extensive experience in scoring TOEFL iBT and/or the *TOEIC*® tests were recruited to participate in the rubric development study. Third, raters were trained to rank order the sampled responses according to three dimensions: oral production, syntax and vocabulary, and content. In addition, the raters rank ordered the responses on overall fluency, a more holistic evaluation of speaking performance. Specific features for each dimension, and for overall fluency, include the following:

Oral Production

- Pronunciation is clear.
- Intonation and stress effectively convey meaning.
- Pacing is appropriate.
- Occasional errors do not interfere with communication.

Syntax and Vocabulary

- Sentence and phrase types vary effectively.
- Word form and word choice are correct.

- Word choice is appropriate to context (e.g., representative of academic context).
- Occasional errors do not interfere with communication.

Content

- Content is full and relevant.
- Content is mostly accurate.
- Content/idea(s) is clearly connected.

Overall Fluency

- Expression is fluid.
- Intelligibility is high.
- Ideas progress clearly (coherence).

Finally, in addition to rank ordering samples, the raters were asked to provide written descriptions of each test taker's performance to justify its ranking. During this process, a scoring rubric of 0 – 4 was developed.

The data from this rubric development study were then finalized based on the pilot administration with a larger sample of test takers from different countries. The final versions of the scoring rubrics are provided in Appendix F.

Developing Scoring Rubrics for the Writing Tasks

The process of creating the writing rubrics was similar to the process used for the speaking rubrics. First, test takers' responses representing a wide range of writing proficiency levels were sampled from responses collected during the prototyping stage. Responses were selected only from the four items whose specifications were similar to those of the items that were selected to be piloted. Second, raters with extensive experience in scoring TOEFL iBT and/or TOEIC writing items were recruited to participate in the rubric development study.

The raters were trained to rank order the sample responses according to four dimensions: content, syntax, vocabulary, and mechanics/conventions. In addition, raters were trained to rank order the responses in terms of overall writing quality. Finally, in each category, raters were asked to list the features of each response that they considered to be most salient, the goal being to provide a rationale for the rankings assigned as well as to support the creation of detailed feature descriptors for the rating scale. Based on these results, scoring rubrics on a 0 – 4 scale were developed for each of the four writing tasks and later refined based on the additional response samples collected during the pilot administrations around the world. The resulting rubrics are presented in Appendix G.

Interpretive Argument and Supporting Research

To support the adequacy and appropriateness of TOEFL Junior scores for the intended test uses outlined earlier, collecting diverse sources of validity evidence is essential. The framework for gathering evidence to validate TOEFL Junior test score interpretation and use is based on the interpretive argument structure approach (Bachman & Palmer, 2010; Kane, Cooks, & Cohen, 1999; Mislevy, Steinberg, & Almond, 2003; Toulmin, 2003). Chapelle, Enright, and Jamieson (2008) provided a comprehensive account of how the interpretative argument approach was utilized as a validation framework for TOEFL iBT test score interpretation and use. In this framework, various types of inferences are made based on warrants or statements that connect test scores to their meanings and uses. To back up the warrants supporting each inference, evidence needs to be collected. Table 14 illustrates inferences, warrants, and types of research needed to yield supportive evidence for validating TOEFL Junior uses. The test design team has referred to the framework to collect validity evidence at different test development stages, and this effort will continue to provide research support to ensure that the TOEFL Junior scores are interpreted and used validly. The penultimate column of the table indicates whether each area of research was addressed at the time of test development, has been conducted subsequent to the introduction of the test, or has yet to be completed. In addition, a reference is provided in the last column if the documentation is publicly available.

Table 14 TOEFL Junior Interpretive Argument and Research Topics

Inferences in the interpretive argument	Warrant	Research topics	Status[a]	Reference
Domain description	Items and tasks in TOEFL Junior represent skills and abilities required for students to have a successful experience in an English-medium middle school context.	• Review of various ESL/EFL standards and textbooks	A	
		• English teacher evaluation of tasks	A	So (2014)
Evaluation	Scores on TOEFL Junior reflect the targeted language abilities and skills.	• Determining difficulty and discrimination levels of test items	B	Young, Morgan, Rybinski, Steinberg, and Wang (2013)
		• Developing scoring rubrics for constructed-response tasks	A	
		• Exploring automated scoring feasibility and identify models	C	
Generalization	Similar scores are expected to be obtained across different test forms and testing occasions.	• Estimating rater reliability and test reliability	A	
		• Monitoring rater and test reliability	C	
		• Developing overall score levels and descriptors for TOEFL Junior	A	Papageorgiou et al. (2014); Papageorgiou et al. (in press)
Explanation	TOEFL Junior scores are attributed to a construct of academic English proficiency.	• Investigating the factor structure of the TOEFL Junior Comprehensive test	B	Gu (2015)
		• Investigating the ability of the TOEFL Junior test to measure students' progress over time	B	
Extrapolation	The construct of academic English proficiency as assessed by TOEFL Junior accounts for students' performance in the real-world context.	• Investigating the relationship between TOEFL Junior and another English proficiency measure	C	
		• Validating the ability of TOEFL Junior to measure/monitor progress in English proficiency in secondary education	B	

Table 14 Continued

Inferences in the interpretive argument	Warrant	Research topics	Status[a]	Reference
Utilization	TOEFL Junior scores are useful for making decisions about placement and for tracking student progress in English proficiency.	• Developing overall score levels and descriptors for TOEFL Junior Standard	A	Papageorgiou et al. (2014); Papageorgiou et al. (in press)
		• Developing overall score levels and descriptors for TOEFL Junior Comprehensive	A	
		• Linking TOEFL Junior performance to the CEFR levels	A	Baron and Tannenbaum (2011)
		• Linking the Reading section of the TOEFL Junior with the Lexile framework	A	MetaMetrics (2012)
		• Validating the intended uses of TOEFL Junior Standard scores for placement decisions in secondary education	B	Papageorgiou and Cho (2014)
		• Understanding users' needs for TOEFL Junior scores	C	
		• Validating the TOEFL Junior Comprehensive score levels and section descriptors using operational data	C	

Note. EFL = English as a foreign language; ESL = English as a second language; CEFR = Common European Framework of Reference.

[a]A = a topic addressed in the test development process before the launch of the test; B = a topic investigated subsequent to the introduction of the test; C = an ongoing operational procedure or a topic being currently investigated or to be investigated. For the topics in the A and B categories, a reference has been provided in case the documentation is publicly available at the time of writing this document.

Notes

1 Details about the relationship between TOEFL Junior scores and the CEFR levels in each of the TOEFL Junior tests can be found on the TOEFL Junior website at https://www.ets.org/toefl_junior/scores_research/standard/cefr (for TOEFL Junior Standard) and at http://www.ets.org/toefl_junior/scores_research/comprehensive/cefr/ (for TOEFL Junior Comprehensive).

2 This trialing process is different from pretesting because trial items are administered to students who are believed to represent the target test-taker population. Conversely, pretest items are administered to actual test takers at the time when they are taking an operational test.

References

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, England: Oxford University Press.

Bailey, A. L. (2007). Introduction: Teaching and assessing students learning English in school. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 1–26). New Haven, CT: Yale University Press.

Bailey, A. L., Butler, F. A., LaFramenta, C., & Ong, C. (2004). *Towards the characterization of academic English in upper elementary science classrooms* (Final Deliverable to OERI Contract No. R305B960002; CSE Technical Report No. 621). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bailey, A. L., & Heritage, H. M. (2008). *Formative assessment for literacy, Grades K–6: Building reading and academic language skills across the curriculum*. Thousand Oaks, CA: Corwin Press.

Bailey, A. L., Heritage, H. M., & Butler, F. A. (2014). Developmental considerations and curricular contexts in the assessment of young language learners. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 423–439). Malden, MA: Wiley-Blackwell.

Baron, P. A., & Tannenbaum, R. J. (2011). *Mapping the TOEFL Junior test onto the Common European Framework of Reference* (ETS Research Memorandum No. RM-11-07). Princeton, NJ: Educational Testing Service.

Butler, F. A., & Castellon-Wellington, M. (2005). Students' concurrent performance on tests of English language proficiency and academic achievement. In *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives* (CSE Technical Report No. 663). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Butler, F. A., Stevens, R., & Castellon-Wellington, M. (2007). ELLs and standardized assessments: The interaction between language proficiency and performance on standardized tests. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 27–49). New Haven, CT: Yale University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1–47.

Chamot, A. U., & O'Malley, J. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. London, England: Routledge.

Council of Europe. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf

Cummins, J. (1980). The construct of proficiency in bilingual education. In J. E. Alatis (Ed.), *Georgetown University round table on languages and linguistics: Current issues in bilingual education* (pp. 81–103). Washington, DC: Georgetown University.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3–49). Los Angeles, CA: National Dissemination and Assessment Center.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.

Dorans, N. J. (2002). *The recentering of SAT scales and its effects on score distributions and score interpretations* (Research Report No. 2002–11). New York, NY: College Board.

Gu, L. (2015). Language ability of young English language learners: Definition, configuration, and implications. *Language Testing*, *32*, 21–38.

Kane, M., Cooks, T., & Cohen, A. (1999). Validating measure of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17.

Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 155–186). Westport, CT: Praeger.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197.

MetaMetrics. (2012). *Linking the reading section of the TOEFL Junior with the Lexile framework*. Durham, NC: Author.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–62.

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*, 223–239.

Papageorgiou, S., Morgan, R., & Becker, V. (2014). *Development of overall score levels and performance descriptors for the TOEFL Junior Standard test*. Manuscript submitted for publication.

Papageorgiou, S., Xi, X., Morgan, R., & So, Y. (in press). Developing and validating band levels and descriptors for reporting overall examinee performance. *Language Assessment Quarterly*.

Scarcella, R. (2003). *Academic English: A conceptual framework* (Technical Report No. 2003–1). Santa Barbara: University of California, Linguistic Minority Research Institute.

Schleppegrell, M. J. (2001). *The language of schooling*. Mahwah, NJ: Lawrence Erlbaum.

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*, 283–303.

Stevens, R., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs* (CSE Technical Report No. 552). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge, England: Cambridge University Press.

Wolf, M. K., & Steinberg, J. (2011). *An examination of United States middle school students' performance on TOEFL Junior* (ETS Research Memorandum No. RM-11-15). Princeton, NJ: Educational Testing Service.

Young, J. W., Morgan, R., Rybinski, P., Steinberg, J., & Wang, Y. (2013). *Assessing the test information function and differential item functioning for TOEFL Junior Standard* (TOEFL Young Learners Research Report No. TOEFL-YL-01). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2013.tb02324.x

Appendix A

Summary of the Curricula and Standards Reviews: Listening

The curricula and standards reviews indicate that the language use in the three TLU subdomains, that is, social and interpersonal, navigational, and academic, differs more by the genres of spoken discourse than by the listening subskills. In other words, the listening subskills required overlap across the three TLU subdomains, which can be seen in the following table.

Table A1 Common Listening Subskills in Multiple Subdomains

Subskill	Examples from ELP standards	TLU subdomain
Understanding the main idea and supporting details	"Identify and explain the main ideas and some details of texts" (CA)	Social and interpersonal, navigational, and academic subdomains
Identifying important details	"Listen and gain information for a variety of purposes, such as summarizing main ideas and supporting details" (FL)	Social and interpersonal, navigational, and academic subdomains
Making inferences or predictions	"Understand implicit ideas and information in increasingly complex spoken language commensurate with grade-level learning expectations" (TX)	Social and interpersonal, navigational, and academic subdomains
Interpreting prosodic features such as intonation and contrastive stress	"Distinguish sounds and intonation patterns of English with increasing ease" (TX)	Social and interpersonal subdomain
Understanding a speaker's purpose	"Identify speaker attitude and point of view" (MI)	Social and interpersonal and academic subdomains

Note. ELP = English language proficiency; TLU = target language use; CA = California; FL = Florida; MI = Michigan; TX = Texas.

Unlike the listening subskills, which are commonly applied to all subdomains, the types of spoken discourse are found to differ across subdomains. The genre, topic/content, and linguistic characteristics of spoken discourse required in each subdomain are summarized in the following table.

Table A2 Types of Spoken Discourse in Each Subdomain

Genre	Topic/content	Characteristics of input/stimuli
Social and interpersonal subdomain		
Conversations	Personal • Feelings • Opinions • Experiences • Events	Form: a dialog/multiparty conversation Length: a number of turn-taking sentences Language characteristics: • Lexical: basic, familiar vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: expressing feeling/opinions; narrating; delivering information; describing
Navigational subdomain		
Directions Announcements	Class-related • Field trip • Homework • School announcement	Form: a monolog Length: a sentence to several sentences Language characteristics: • Lexical: basic, familiar, academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent monolog • Pragmatic: delivering information; describing; instructing; reminding; announcing; requesting
Academic subdomain		
Lectures Academic discussions	Academic content-related • Science • Social studies • Literature • Math	Form: a monolog/multiparty discussion Length: sustained discourse about an academic topic Language characteristics: • Lexical: academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent discourse about a given topic • Pragmatic: summarizing, describing, analyzing, and evaluating

Appendix B

Summary of the Curricula and Standards Reviews: Reading

As in listening, common reading subskills are found to be required in all of the three TLU subdomains. The following table summarizes the reading subskills. Note that all of the reading subskills summarized in the table apply to all three subdomains: social and interpersonal, navigational, and academic.

Table B1 Reading Subskills Common to all Subdomains

Subskill	Examples from ELP standards
Understanding the main idea	"Identify and explain the main ideas and some details of texts" (CA) "Identify important details, essential message, and main idea of a text" (FL)
Identifying important details	"Listen and gain information for a variety of purposes, such as summarizing main ideas and supporting details" (FL)
Making inferences or predictions	"Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings" (NY)

Table B1 Continued

Subskill	Examples from ELP standards
Inferring the meaning of a word from context/understanding figurative and idiomatic language from context	"Employ phonemic awareness, inference, contextual clues, synonyms and antonyms relationships to analyze words and text" (FL) "Apply knowledge of word relationships, such as roots and affixes, to derive meaning from literature and texts in content areas" (CA)
Recognizing an author's purpose	"Identify speaker attitude and point of view" (MI)

Note. ELP = English language proficiency; CA = California; FL = Florida; MI = Michigan; NY = New York.

The three TLU subdomains are found to differ by the genres of reading materials that students are required to understand for different purposes. The topics and the linguistic characteristics are also found to change with different genres.

Table B2 Types of Written Genre in Each Subdomain

Genre	Topic/content	Characteristics of input/stimuli
Social and interpersonal subdomain		
Correspondence (e.g., e-mails and letters)	Personal • Feelings • Opinions • Experiences • Events	Form: a written letter, e-mail, social media site post, text message Length: varied: a few words to multiple paragraphs Language characteristics: • Lexical: mostly basic vocabulary; idiomatic expressions • Grammatical: simple to complex sentences • Discourse: a coherent text • Pragmatic: using appropriate register; delivering information; explaining; describing
Navigational subdomain		
Nonlinear text (e.g., schedules and announcements) Brochures Journalism	Class-related • Field trip • Homework • School announcement	Form: chart, graph, poster, flyer including a written text, advertisement, brochure, graphic (schedule) Length: varied: phrases, a few sentences to multiple paragraphs Language characteristics: • Lexical: basic to some academic vocabulary • Grammatical: simple to complex sentences • Discourse: fragments to simple sentences • Pragmatic: delivering information
Academic subdomain		
Text about an academic topic	Academic content-related • Science • Social studies • Literature • Math	Form: a written text Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent text • Pragmatic: describing; analyzing; comparing; contrasting; evaluating; commenting

Appendix C

Summary of the Curricula and Standards Reviews: Speaking

For the productive language skills (speaking and writing), the specific language tasks can be described more explicitly than for the receptive skills. Therefore, the summary tables for the two productive skills are organized somewhat differently from the tables for the receptive skills.

Table C1 Types of Speaking Tasks in Each Subdomain

Task	ELP standards	Characteristics of input	Characteristics of expected response
Social and interpersonal subdomain **Topic/content: Personal topics such as feelings, greetings, hobbies, experiences, and events**			
Establish social contact: greetings and farewells; introductions; giving thanks (France, Korea, Japan)	"Respond to and offer greetings, compliments, introductions or farewells with teachers or peers in L1 and L2" (WIDA) "Independently use common social greetings and simple repetitive phrases (e.g., "Good morning, Ms. ___")" (CA)	Form: a spoken dialog Length: number of turn-taking sentences Language characteristics: • Lexical: basic, familiar; contextualized formulaic expressions • Grammatical: mostly simple sentence structures • Discourse: a coherent dialog • Pragmatic: greeting; interpersonal conversation	Form: a spoken dialog Length: number of turn-taking sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic, familiar, routine vocabulary; contextualized formulaic expressions • Grammatical: mostly simple sentence structures • Discourse: a coherent dialog • Pragmatic: greeting; interpersonal conversations
Ask and respond to questions concerning basic needs or request help (Korea)	"Request and provide information and assistance orally for personal, social, and academic purposes" (NY) "Ask questions or exchange information with teachers or peers in L1 and L2" (WIDA) "Orally communicate basic needs" (CA)	Form: a spoken dialog, a written text Length: varied Language characteristics: • Lexical: mostly basic and some formulaic language appropriate to context • Grammatical: simple to complex sentences • Pragmatic: delivering information • Discourse: a coherent dialog	Form: sustained utterances of varying lengths Language characteristics: • Phonological: produce clear pronunciation • Lexical: mostly basic with some idiomatic usage appropriate to the context • Grammatical: some variation in grammatical constructions of phrases and sentences • Discourse: a coherent monolog • Pragmatic: delivering information; requesting; clarifying; describing
Describe routine everyday experiences, people, events (France)	"State basic personal information and preferences" (CO)	Form: a spoken dialog with aids of visual images (a picture or sequence of pictures)	Form: a spoken dialog or a monolog

Table C1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
Communicate with peers about daily life (Korea, Japan)	"Describe, participate in, or recommend a favorite activity, book, song, or other interest" (NY)	Length: number of turn-taking sentences Language characteristics: • Lexical: basic, familiar vocabulary; formulaic expressions • Grammatical: simple to complex sentences • Discourse: a coherent dialog	Length: number of turn-taking sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic, familiar vocabulary, some formulaic expressions • Grammatical: simple to complex sentences • Discourse: a coherent dialog, usually recounting events in an appropriate sequence • Pragmatic: describing; delivering information; narrating
Give brief comments on the views of others (France, Chile)	"Recognize and communicate personal and multiple points of view within and among groups" (NY) "Express opinions, ideas" (TX)	Form: a spoken dialog Length: number of turn-taking sentences Language characteristics: • Lexical: basic, familiar, non-familiar vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: using appropriate register; explaining; clarifying; commenting	Form: a spoken dialog Length: number of turn-taking sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic, familiar, nonfamiliar vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: using appropriate register; explaining; clarifying; commenting; supporting
Navigational subdomain **Topic/content: School-related topics such as field trip, homework, school announcement, and extracurricular activities**			
Ask and respond to questions to obtain or clarify information	"Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning" (NY) "Ask and respond to questions to obtain, clarify, and extend information" (CO) "Ask questions or exchange information with teachers or peers in L1 and L2" (WIDA)	Form: a spoken dialog Length: a few turn-taking sentences Language characteristics: • Lexical: basic, familiar to academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: delivering information	Form: a spoken dialog Length: a few turn-taking sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic, familiar to academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: delivering, identifying and finding factual information

Table C1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
Recount details of school-related events or basic/routine oral instructions, procedures, or visually supported assignments (France)	"Convey and organize information, using facts, details, illustrative examples" (NY) "Restate or paraphrase visually supported information from multimedia (e.g., in newspapers, magazines or broadcasts)" (WIDA) "Restate information and identify main idea and some details using sentences, e.g., oral presentations, texts, media, etc." (CO)	Form: event Length: a few turn-taking sentences Language characteristics: • Lexical: basic, familiar vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent dialog • Pragmatic: delivering information	Form: a spoken monolog; dialog Length: a sentence to a paragraph Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic, familiar vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent monolog or dialog • Pragmatic: delivering information; recounting the facts with some amount of accurate details
Academic subdomain **Topic/content: Academic content-related such as science, social studies, literature, and math**			
Read aloud short academic texts, typical of middle school course content, to the class (Korea)	"Approximate pronunciation, rhythm, stress, and intonation of English" (CO) "Consistent use of sounds, intonation, pitch" (CA)	Form: a written text Length: a sentence to a paragraph Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent text	Form: reading aloud Length: a sentence to a paragraph Language characteristics: • Phonological: produce clear and accurate pronunciation, reading fluency, and accurate pace and rhythm/intonation/stress • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent monolog • Pragmatic: delivering information
Summarize or paraphrase academic information from a written and/or spoken source, including salient details	"Demonstrate the ability to paraphrase and/or summarize" (FL) "Restate and paraphrase visually supported information from multimedia" (WIDA) "Restate information and identify the main idea and some details" (CO)	Form: lecture, along with written texts and visual images or pictures about the topic Length: varying; could be up to 20–30 min Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences, with characteristics of academic language (e.g., complex noun phrases, passive construction) • Discourse: a coherent monolog; coherence devices	Form: a monolog Length: a paragraph to multiple paragraphs Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic to academic vocabulary, incorporating key vocabulary appropriately • Grammatical: variation in sentence structures, depending on contexts • Discourse: a coherent monolog; transitional devices; providing main ideas with some elaboration • Pragmatic: delivering information with details; explaining

Table C1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
Recount an event or an experience using words that establish chronological sequence, i.e., conveying chronology of events or stages of developments in classroom lessons in history, math, or science (France, CEFR)	"Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures, including chronological order, cause and effect, problem and solution, and general-to-specific presentation" (NY) "Discuss, share, reflect on, develop, and express opinions and evaluations about a variety of experiences, ideas, and information" (NY)	Form: a monolog (lecture); visual images Length: varied Language characteristics: • Lexical: basic to academic vocabulary, incorporating technical terminology • Grammatical: simple to complex sentences, with language features of academic language • Discourse: coherence devices • Pragmatic: clarifying; delivering information	Form: a monolog Length: several sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic to academic vocabulary, incorporating key vocabulary appropriately • Grammatical: variation in sentence structures, depending on contexts • Discourse: a coherent monolog; transitional devices • Pragmatic: delivering information; clarifying
On academic topics, explain, compare, and contrast (France)	"Discuss, compare/contrast and state differences, describe, defend, explain causes and effects" (WIDA) "Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects" (NY) "Discuss, share, reflect on, develop, and express opinions and evaluations about a variety of experiences, ideas, and information" (NY)	Form: a spoken dialog; monolog (lecture) Length: number of turn-taking sentences; multiple paragraphs Language characteristics: • Lexical: general to technical academic vocabulary • Grammatical: simple to complex sentences, academic language characteristics • Discourse: a coherent dialog • Pragmatic: explaining clarifying; defining; describing	Form: a spoken dialog; a monolog Length: number of turn-taking sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic to academic vocabulary, incorporating key vocabulary appropriately • Grammatical: variation in sentence structures, depending on contexts • Discourse: a coherent dialog • Pragmatic: explaining; comparing; commenting; organizing
Link and convey information from more than one source	"Select information appropriate to the purpose of the investigation, relate ideas from one written or spoken source to another, and exclude nonessential information" (NY) "Research, organize, and effectively deliver a basic speech to inform, persuade, or defend, based on retelling, speech drafting, and teacher prompting/coaching" (FL)	Form: a spoken dialog, along with a related written text or slides Length: multiple paragraphs Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences, academic language features • Discourse: a coherent monolog, dialog	Form: a spoken monolog Length: a paragraph to multiple paragraphs Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic to academic vocabulary, incorporating key vocabulary appropriately • Grammatical: variation in sentence structures, depending on contexts

Table C1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
		• Pragmatic: delivering information	• Discourse: a coherent monolog • Pragmatic: delivering information; describing; explaining; comparing
Prepare and deliver a simple presentation on an assigned academic topic (Korea)	"Make simple presentations in small groups using single words and phrases and visual support" (CO) "Prepare and deliver presentations on ideas, premises, images, or content areas, including a purpose, point of view, introduction, coherent transition, and appropriate conclusions" (CA)	Form: a spoken dialog along with a written text or slides Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences; academic language features • Discourse: a coherent monolog, dialog • Pragmatic: delivering information	Form: a spoken monolog Length: several sentences Language characteristics: • Phonological: produce clear pronunciation • Lexical: basic to academic vocabulary, incorporating key vocabulary appropriately • Grammatical: variation in sentence structures, depending on contexts • Discourse: a coherent monolog • Pragmatic: delivering information; describing; explaining

Note: (Korea), (France), (Japan), (China), (Chile) indicate sources of international curricula or textbooks, and the state names in the English language proficiency (ELP) standards column (CA = California; CO = Colorado; FL = Florida; NY = New York; TX = Texas) indicate the state where the specific standards were found.

Appendix D

Summary of the Curricula and Standards Reviews: Writing

The following summary table is organized in the same way that the speaking summary was organized. One difference is that there are writing tasks that can be applied to all of the three TLU subdomains. These are categorized as enabling skills in the table.

Table D1 Types of Writing Tasks in Each Subdomain

Task	ELP standards	Characteristics of input	Characteristics of expected response
Enabling skills			
Use correct writing conventions	"Use conventional spelling, capitalization, and punctuation" (CO) "Edit for basic conventions of writing and revise for appropriate word choice and organization" (CO)		
Edit and revise written products	"Revise writing to clarify meaning and improve the word choice, mechanics, organization, and consistent point of view" (CA) "Modify sentence structure, paragraphs, and word choices using resources and reference material" (FL)		
Social and interpersonal subdomain			
Topic/content: Personal topics such as feelings, greetings, hobbies, experiences, and events			
Communicate basic personal information to others through e-mails, letters, etc. (France)	"Use a variety of print and electronic forms for social communication and for writing to or for self" (NY) "Communicate in writing using a variety of genres, e.g.: narrative, content area reports, letter writing, poetry, autobiography" (CO)	Form: a written text, an oral presentation, an event/trip, a dialog Length: a few sentences to multiple paragraphs Language characteristics: • Lexical: basic vocabulary • Grammatical: simple to complex sentences • Pragmatic: describing; delivering information	Form: a written text, e-mail, essay Length: a few sentences to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: describing; informing; reporting basic facts
Navigational subdomain			
Topic/content: School-related topics such as field trip, homework, school announcement, and extracurricular activities			
Complete various forms (Japan, Chile)	"Use a variety of print and electronic forms" (NY) "Make lists associated with school life from visuals and word/phrase banks (e.g., subjects, classes, activities)" (WIDA)	Form: a written text, table, chart, graphic organizer Length: varied Language characteristics: • Lexical: basic vocabulary	Form: written text Length: varied: a single word to a few sentences Language characteristics: • Lexical: basic and academic vocabulary

Table D1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
		• Grammatical: simple to complex sentences • Pragmatic: delivering information	• Grammatical: simple to complex sentences • Pragmatic: delivering information; reporting basic facts
Academic subdomain **Topic/content: Academic content-related such as science, social studies, literature, and math**			
Write using a variety of genres (Korea)	"Communicate in writing using a variety of genres, e.g.: narrative, content area reports, letter writing, poetry, autobiography" (CO) "Write essays in a variety of information/expository forms (e.g., procedures experiments assembly instructions)" (FL)	Form: a written text, an oral presentation, an event/trip, a dialog Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: delivering information	Form: a written text, story, letter, essay Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: delivering information; describing
Take notes from a lecture or oral presentation	"Take notes during lectures or from movies and reading material" (FL) "Collect information from various sources (e.g., dictionary, library books, research materials) and take notes on a given topic" (CA)	Form: an oral presentation/lecture, visual images (pictures, charts, tables) Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent text • Pragmatic: delivering information; describing; explaining	Form: fragments of text Length: varied Language characteristics: • Lexical: basic to academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: inferring; synthesizing; describing; delivering information; reporting; explaining
Write a description of events	"Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures, including chronological order, cause and effect, problem and solution, and general-to-specific presentation" (NY)	Form: journal writing, a science report Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary	Form: a written text Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary

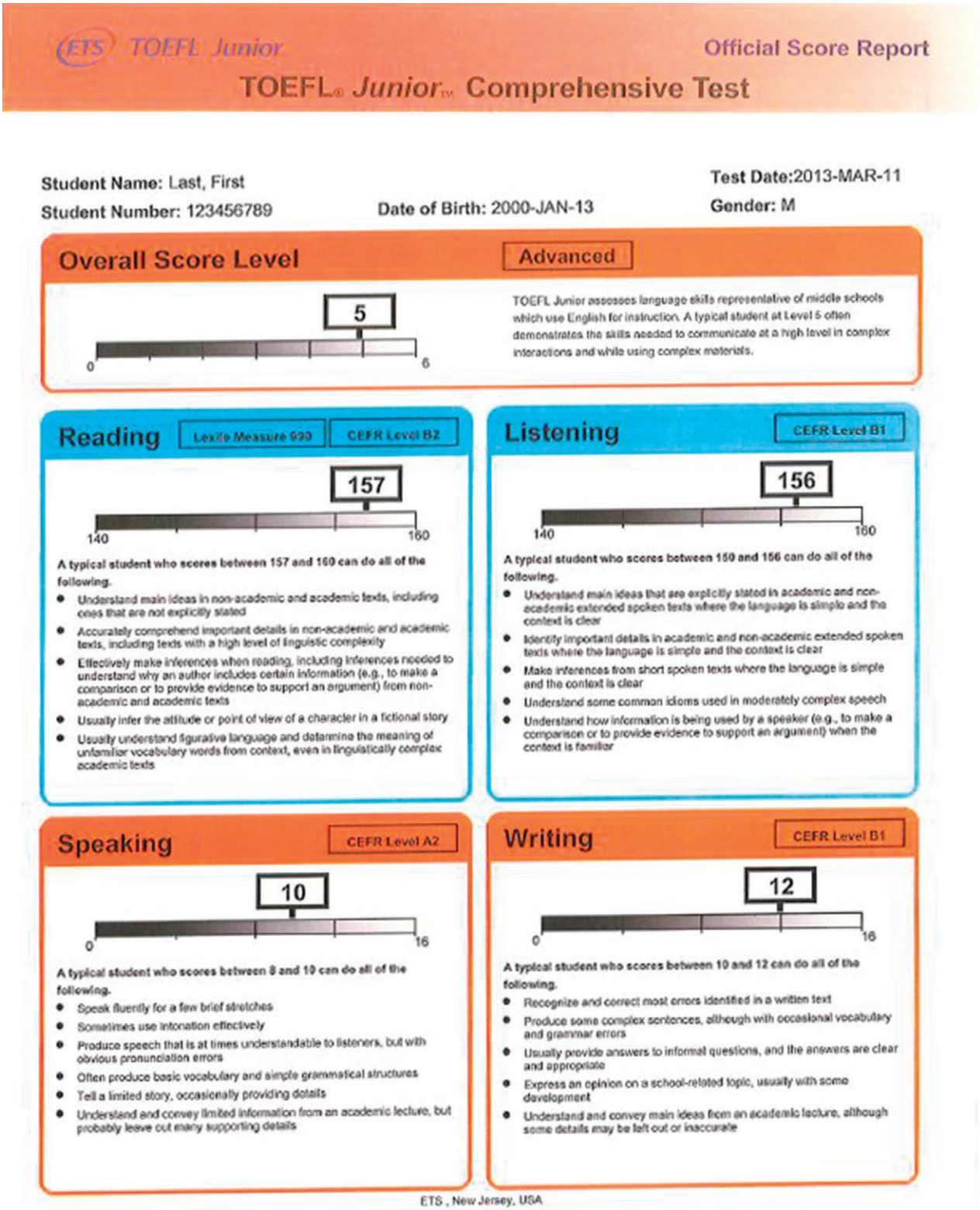
Table D1 Continued

Task	ELP standards	Characteristics of input	Characteristics of expected response
		• Grammatical: simple to complex sentences • Pragmatic: delivering information	• Grammatical: simple to complex sentences • Discourse: a coherent text • Pragmatic: delivering information; narrating; synthesizing; describing; comparing
Write a paragraph with main ideas and supporting reasons, details and/or examples	"Write essays that contain a topic sentence or controlling idea, supporting details or arguments for the validity of the proposed idea" (FL) "Write paragraphs and short expository compositions on content area topics, e.g.: compare/contrast, cause/effect, problem/solution that include an introductory statement, supporting details and conclusion" (CO)	Form: a written text, a short talk/lecture, visual images (pictures, charts, tables) Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: mostly basic and some academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: delivering information	Form: a written text, book report Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: mostly basic and some academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: delivering information; summarizing
Write persuasive/opinion compositions	"Write persuasive and expository compositions that include a clear thesis, describe organized points of support, and address a counterargument" (CA)	Form: a written text, an oral presentation Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Discourse: a coherent text	Form: a persuasive written text Length: a paragraph to multiple paragraphs Language characteristics: • Lexical: basic and academic vocabulary • Grammatical: simple to complex sentences • Pragmatic: explaining; persuading; comparing; evaluating; justifying

Note: (Korea), (France), (Japan), (China), (Chile) indicate sources of international curricula or textbooks, and the state names in the English language use (ELP) standards column (CA = California; CO = Colorado; FL = Florida; NY = New York; TX = Texas) indicate the state where the specific standards were found.

Appendix E

TOEFL Junior Comprehensive Sample Score Report



Understanding Your Score Report

Overall Score Level, Section Scores, and Performance Descriptors

Your score report provides both an overall score level and individual section scores. Overall score levels range from 1-6 and are accompanied by overall performance descriptors. The Reading and Listening section scores range from 140-160. The Speaking and Writing section scores range from 0-16; these scores are linked to the rubrics used to score Speaking or Writing tasks to help you better understand the meaning of your scores. Section scores are accompanied by section performance descriptors. For more detailed information about the score scales and performance descriptors, please visit www.ets.org/toefl_junior.

The following table shows all of the TOEFL Junior Comprehensive overall score levels, overall performance descriptors, and CEFR profiles.

TOEFL Junior Overall Score Level	Overall Performance Descriptor These descriptors represent performance in middle schools which use English for instruction. A typical student at this level:	CEFR Profile A typical student at this level achieved these section-level CEFR scores:
6 Excellent	consistently demonstrates the skills needed to communicate successfully at a high level in complex interactions and while using complex materials.	B2 for all sections (Reading, Listening, Speaking, and Writing)
5 Advanced	often demonstrates the skills needed to communicate successfully at a high level in complex interactions and while using complex materials.	B1 or B2 for Reading and Listening; B1 for Speaking and Writing
4 Competent	demonstrates the skills needed to communicate successfully in some complex situations and in most simple interactions and while using basic materials.	B1 for Reading and Listening; B1 or A2 for Speaking and Writing
3 Achieving	usually demonstrates the skills needed to communicate successfully in simple interactions and while using basic materials.	A2 or B1 for Listening; A2 for Reading, Speaking, and Writing
2 Developing	occasionally demonstrates the skills needed to communicate successfully in simple interactions and while using basic materials.	A2 for Reading and Listening; below A2 for Speaking and Writing
1 Beginning	demonstrates some basic language skills but needs to further develop those skills in order to communicate successfully.	Below A2 for all sections (Reading, Listening, Speaking, and Writing)

NOTE: For students who did not attempt the test, an overall score level of 0 is reported.

CEFR Levels

The Common European Framework of Reference (CEFR) is a a widely-used tool for understanding different stages of language learning. It covers six proficiency levels: A1, A2, B1, B2, C1 and C2, with C2 being the highest level. The performance descriptors for each section of TOEFL Junior are adapted from the CEFR level descriptors to reflect the TOEFL Junior test content and the age of the test takers. For more detailed information on how TOEFL Junior test scores relate to the CEFR, please visit www.ets.org/toefl_junior.

Lexile Measure

A Lexile measure matches English-language readers with appropriate texts. Use your Lexile measure to find books at your reading level. To begin, visit toefljunior.lexile.com.

Recommended Score Validity -- 2 Years

TOEFL Junior measures a student's current level of English language proficiency and is not designed to be a predictor of future TOEFL iBT scores. Because language proficiency is likely to change over time, institutions are advised not to use test scores that are older than two years.

Copyright © 2012 by Educational Testing Service. All rights reserved. ETS, the ETS logos, and TOEFL Junior are registered trademarks of Educational Testing Service (ETS) in the United States and other countries throughout the world.

LEXILE and the LEXILE logo are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad..

94596-94596 • SR812E59 • Printed in U.S.A. • I.N. 301525

768890

Appendix F

Scoring Rubrics for Speaking Tasks

TOEFL Junior Speaking Scoring Guide
Task 1: Read Aloud

Score	Fluency and Accuracy Descriptors
4	**A typical response at this level is characterized by the following:** • Reading is mostly fluid and intelligible • Words are grouped in meaningful phrases with effective pauses. Punctuation is marked appropriately throughout • Intonation varies to match text provided • Speech is clear and distinct with only minor mispronunciations, substitutions, or omissions • Rate of speech is mostly appropriate
3	**A typical response at this level is characterized by the following:** • Reading is fairly fluid and intelligible • Words are generally grouped in meaningful phrases with only minor lapses. Punctuation is usually marked appropriately • Intonation may seem flat/monotone at times • Speech is clear and distinct most of the time; some mispronunciations, substitutions, or omissions may be noticeable but do not impact overall intelligibility • Rate of speech is mostly appropriate; occasional variation *may* cause minor lapses in intelligibility
2	**A typical response at this level is characterized by the following:** Reading is noticeably choppy and unintelligible at times; sometimes is read word-by-word without meaningfully grouped phrases. Punctuation may not be marked at times. • Intonation may often be flat or monotone • Speech is clear and distinct at times but noticeable mispronunciations, substitutions, omissions, and self-corrections interrupt the flow and may impact overall intelligibility • Rate of speech is inappropriate at times (resulting in choppy pace or slurred words and mispronunciations) • Note: A response at this level may also be marked by numerous substitutions, omissions, and attempts to paraphrase, rather than read, sections of the text.
1	**A typical response at this level is characterized by the following:** • Reading is hard to follow and mostly unintelligible, with multiple starts and stops. Reading may be incomplete • Intonation is rarely used effectively • Frequent errors in pronunciation and stress; words may not be comprehensible • Self-corrections are ineffective most of the time. Substitutions may alter meaning substantially • Slow rate of speech • Punctuation rarely marked
0	No attempt to respond OR No English in the response OR Response is off topic OR Insufficient language to evaluate

TOEFL Junior Speaking Scoring Guide
Task 2: Six-Picture Narration

Score	Content, Delivery and Language Use Descriptors
4	**A typical response at this level is characterized by the following:** • Story is full, relevant to the pictures, and includes some detail and elaboration. Only minor lapses in content or coherence. Events unfold evenly and the sequence is easy to follow. • Overall fluidity of expression is evident; fairly smooth and confident rate of speech; little hesitancy. Errors of pronunciation, stress, and intonation may occur but rarely obscure meaning. • Grammar and word choice are varied, appropriate to the task, and effectively used to convey meaning clearly. Errors rarely obscure meaning. Use of connecting devices helps to link events for the listener.
3	**A typical response at this level is characterized by the following:** • Story is mostly complete but may include some noticeable lapses in content or coherence. Description of some key events may lack detail or elaboration. Some details may be confusing to the listener. • Mostly fluid expression but some hesitancy and choppiness may be noticeable. Errors in pronunciation, stress, and intonation may occasionally obscure meaning. Able to sustain speech to complete the story. • Some limitations and errors in grammar and word choice are noticeable but meaning is rarely obscured. Use of connecting devices to link events may be limited.
2	**A typical response at this level is characterized by the following:** • Limited development of the story. Most events are recounted with little to no elaboration or detail. Some events or details may be difficult to follow for listeners not familiar with the pictures. Limited development and cohesion may cause listener to fill in the gaps. • May sustain speech throughout but pace may be slow, choppy, or hesitant throughout. Errors in pronunciation, stress, and intonation occasionally impact intelligibility and flow. • May struggle to convey the story due to limitations in grammar and word choice. May rely on mostly simple grammatical constructions and basic vocabulary. These limitations and errors may result in vague or unclear meaning at times.
1	**A typical response at this level is characterized by the following:** • Very limited development of story; may be incomplete. Story lacks detail or elaboration. • Generally unable to sustain speech throughout to complete a story. Frequent errors in pronunciation, stress, and intonation impact intelligibility. • Most utterances are characterized by errors. Vocabulary is limited and often inaccurate.
0	No attempt to respond OR No English in the response OR Response consists of a repetition of the prompt OR response is off topic

TOEFL Junior Speaking Scoring Guide
Tasks 3 and 4: Listen-Speak

Score	Content, Delivery and Language Use Descriptors
4	**A typical response at this level is characterized by the following:** • Content is full and appropriate to the task. Key information is conveyed coherently and accurately with some elaboration and detail although minor errors may occur. Connection among ideas is clear. • Speech is mostly clear and fluid with occasional imperfections. Minor errors of pronunciation, stress, and intonation do not interfere with understanding. Mispronunciation of key content words may occur but rarely obscure meaning. • Grammar and word choice are varied, appropriate to the task, and effectively convey meaning. Occasional errors of word form and grammar may occur but rarely obscure meaning.
3	**A typical response at this level is characterized by the following:** • Content is mostly complete and appropriate to the task. Most key information is conveyed accurately but supporting details and elaboration are limited or lacking; minor inaccuracies or omissions may be evident. Response is fairly cohesive with minor lapses. • Response is mostly fluid and sustained with some lapses or imperfections evident. Pronunciation, stress, and intonation errors are noticeable but do not usually interfere with understanding. May struggle with pronunciation of unfamiliar content words. • Basic grammar and vocabulary are usually controlled although minor errors may occur. Awkward or inappropriate phrases may occur as the speaker attempts new constructions but these do not cause major misunderstandings.
2	**A typical response at this level is characterized by the following:** • Development is mostly limited to some (or all) of the main facts, presented one by one. Relies on the listener to make the connections between facts most of the time. Key information may be vaguely expressed or incomplete. Some misunderstanding of the talk may be evident. Some key information may be omitted or inaccurate. • Response may be fluid at times but speaker struggles to sustain. Response may be characterized by slow, choppy, or hesitant delivery. Errors in pronunciation, word stress, and intonation are evident and may interfere with understanding at times. • Lacks sufficient range and control of grammar and vocabulary to provide a concise summary of information. May rely on basic vocabulary to convey meaning. Linguistic errors are evident, may be systematic, and occasionally interfere with understanding.
1	**A typical response at this level is characterized by the following:** • Content is incomplete and/or lacks development. Information conveyed is limited and may be vague or inaccurate. • Struggles to sustain speech to complete the task (or may sustain speech for only brief segments at a time, stopping and starting often) Pronunciation errors are evident but speaker may be understandable at times to the sympathetic listener. • May rely heavily on basic, high-frequency vocabulary or familiar, rehearsed phrases to convey content. Vocabulary is limited and often inaccurate. There may be little use of modifiers. Struggles to construct grammatical utterances beyond a few words.
0	No attempt to respond OR No English in the response OR Response consists of a repetition of the prompt OR response is off topic

Appendix G

Scoring Rubrics for Writing Tasks

TOEFL Junior Writing Scoring Guide
Task 1: Edit

Score	Language Use
4	• Corrects all four errors
3	• Corrects three errors
2	• Corrects two errors
1	• Corrects one error
0	Attempts to correct all errors but does so incorrectly OR Makes no attempt to correct errors, only copies words from the stimulus, consists of only unrelated content, consists of keystroke characters, or is written in a foreign language

TOEFL Junior Writing Scoring Guide
Task 2: E-mail

Score	Development and Language Use Descriptors
4	**A typical response at this level is characterized by the following:** • responds to all questions in the e-mail, directly or indirectly • is coherent • shows lexical variation appropriate for the task • displays a varied sentence structure appropriate for the task • may contain minor errors but they do not interfere with meaning
3	**A typical response at this level is characterized by the following:** • responds to most of the questions in the e-mail, directly or indirectly • is generally coherent • shows some lexical variation appropriate for the task • may display variation in sentence structure appropriate for the task • may contain some errors that occasionally interfere with meaning
2	**A typical response at this level is characterized by the following:** • responds to some questions in the e-mail • may be incoherent at times • shows little lexical variation (e.g., vocabulary is simple and repetitive), or often uses vocabulary incorrectly • may show little control of sentence structures • may contain errors that frequently interfere with meaning
1	**A typical response at this level is characterized by the following:** • responds minimally to questions in the e-mail • is generally incoherent • displays limited vocabulary that may be used incorrectly • uses mostly incorrect sentence structures • displays many errors that seriously interfere with meaning
0	Only copies words from the prompt, rejects the prompt, is completely off topic, consists of keystroke characters, is written in a foreign language, or is blank

TOEFL Junior Writing Scoring Guide
Task 3: Opinion

Score	Development and Language Use Descriptors
4	**A typical response at this level is characterized by the following:** • states a position on the topic • provides support for the position, with specific details and/or examples • is mostly well organized and coherent • shows lexical variation appropriate for the task • displays a varied sentence structure appropriate for the task • may contain minor errors but they do not interfere with meaning or clarity
3	**A typical response at this level is characterized by the following:** • states a position on the topic • provides support for the stated position, but may have difficulty doing so fully • is generally well organized, with an occasional lapse of clarity when connecting ideas • shows some lexical variation appropriate for the task • may display some variation in sentence structure appropriate for the task • may contain some errors that occasionally interfere with meaning
2	**A typical response at this level is characterized by the following:** • states a position on the topic, but provides inadequate/incomplete support, OR • only vaguely implies a position on the topic, and provides inadequate/incomplete support • connections between ideas are attempted, but are sometimes unclear or missing • shows little lexical variation (e.g., vocabulary is simple and repetitive), or frequently uses vocabulary incorrectly • shows little variation in sentence structure (e.g., sentences are mostly simple and short), and shows little control of sentence structures • may contain errors that frequently interfere with meaning
1	**A typical response at this level is characterized by the following:** • states a position but provides incoherent or no support OR • does not state a position, or makes only a minimal connection to the prompt and provides minimal or no support • is generally unorganized and incoherent • displays extremely limited vocabulary that is frequently used incorrectly • uses mostly incorrect sentence structures • displays many errors that seriously interfere with meaning
0	Only copies words from the prompt, rejects the prompt, is completely off topic, consists of keystroke characters, is written in a foreign language, or is blank

TOEFL Junior Writing Scoring Guide
Task 4: Listen-Write

Score	Development and Language Use
4	**A typical response at this level is characterized by the following:** • accurately provides all key points • provides support using relevant details from the talk • is mostly well organized and coherent • shows lexical variation appropriate for the task • displays a varied sentence structure appropriate for the task • may contain errors but they do not interfere with meaning or clarity
3	**A typical response at this level is characterized by the following:** • accurately provides most key points • provides some supporting details from the talk • is generally organized, with an occasional lapse of clarity when connecting ideas • shows some lexical variation appropriate for the task • may display some varied sentence structure appropriate for the task • may contain errors that occasionally interfere with meaning
2	**A typical response at this level is characterized by the following:** • provides some accurate content from the key points • provides minimal or no supporting details from the talk • connections between ideas are attempted but are often unclear or missing • shows little lexical variation (e.g., vocabulary is simple and repetitive), or frequently uses vocabulary incorrectly • shows little variation in sentence structure (e.g., sentences are mostly simple and short), or shows little control of sentence structures • may contain errors that frequently interfere with meaning
1	**A typical response at this level is characterized by the following:** • provides minimal or no content from the key points • does not provide details beyond those shown in the visuals • provides incoherent or no support for any of the points • is generally unorganized and incoherent • displays extremely limited vocabulary that is frequently used incorrectly • uses mostly incorrect sentence structures • displays many errors that seriously interfere with meaning
0	Only copies words from the prompt, rejects the prompt, is completely off topic, consists of keystroke characters, is written in a foreign language, or is blank

Suggested citation:

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 No. 02, ETS Research Report No. RR-15-13). Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12058

Action Editor: Donald E. Powers

Reviewers: This report was reviewed by the Research Subcommittee of the TOEFL Committee of Examiners.

ETS, the ETS logo, GRADUATE RECORD EXAMINATION, GRE, TOEFL, TOEFL IBT, TOEFL JUNIOR, the TOEFL JUNIOR logo, and TOEIC are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/