



Research Memorandum

ETS RM-24-02

Design Framework for the TOEFL Primary® Writing Test

Mikyung Kim Wolf
Michael Suhan
Mitchell Ginsburgh
Yoko Futagi
Feifei Li

January 2024



ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey
Lord Chair in Measurement and Statistics

ASSOCIATE EDITORS

Usama Ali
Senior Measurement Scientist

Beata Beigman Klebanov
Principal Research Scientist

Heather Buzick
Senior Research Scientist

Tim Davey
Director Research

Larry Davis
Director Research

Jamie Mikeska
Senior Research Scientist

Jonathan Schmidgall
Senior Research Scientist

Jesse Sparks
Senior Research Scientist

Klaus Zechner
Senior Research Scientist

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Gontz
Senior Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Design Framework for the TOEFL Primary® Writing Test

Mikyung Kim Wolf, Michael Suhan, Mitchell Ginsburgh, Yoko Futagi, and Feifei Li
Educational Testing Service, Princeton, New Jersey, United States

January 2024

Corresponding author: M. K. Wolf, E-mail: mkwolf@ets.org

Suggested citation: Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). *Design framework for the TOEFL Primary® Writing test* (Research Memorandum No. RM-24-02). ETS.

Find other ETS-published reports by searching the
ETS ReSEARCHER database.

To obtain a copy of an ETS research report, please visit
<https://www.ets.org/contact/additional/research.html>

Action Editor: Jonathan Schmidgall

Reviewers: Spiros Papageorgiou and Veronika Timpe-Laughlin

Copyright © 2024 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, TOEFL, TOEFL JUNIOR, and TOEFL PRIMARY are registered trademarks of Educational Testing Service (ETS). All other trademarks are the property of their respective owners.

Abstract

The TOEFL Primary® tests, designed as international English language proficiency assessments, have been widely used in over 120 countries since 2013. The tests are primarily intended to support the teaching and learning of young language learners aged 8 and older. Until recently, the TOEFL Primary tests measured only reading, listening, and speaking skills. In 2023, with the introduction of the TOEFL Primary Writing test, it became possible to assess all four language skills—reading, listening, speaking, and writing—through the TOEFL Primary tests. This research memorandum documents the underlying design principles behind the development of the TOEFL Primary Writing test and provides detailed information about the test. The directions for validity research concerning the TOEFL Primary Writing test are also discussed.

Keywords: English language proficiency, test development, TOEFL Primary®, validity, writing, young language learners

With the rapidly growing number of young students worldwide who learn English as an additional, second, or foreign language (EAL/ESL/EFL), international standardized English language proficiency assessments for young learners are in increased demand (Wolf & Butler, 2017). Meeting this demand, the TOEFL Primary® tests, part of the TOEFL® Young Students Series, have been used in more than 120 countries since 2013. Developed by ETS, the TOEFL Primary tests are designed for students primarily aged 8 and older. The overarching construct of the tests is to measure young EAL/EFL students' abilities to communicate in English in familiar and age-appropriate contexts. The intended uses of the tests include (a) assessing students' English language abilities to guide teaching and learning, (b) measuring students' progress in attaining English proficiency, (c) informing placement decisions for appropriate classes, and (d) evaluating performance against the international benchmark, specifically the Common European Framework of Reference for Languages (CEFR). Until recently, the tests included measures of reading, listening, and speaking skills (see ETS, 2019, 2023, for more details about the TOEFL Primary tests).

While the TOEFL Primary tests initially focused on foundational reading and oral language proficiency, many test users expressed interest in gaining more insight into students' writing proficiency and including a writing measure in the TOEFL Primary tests. To ensure that young students develop all aspects of their English language proficiency, assessing all four language skills is essential (McKay, 2006). To address the need, ETS recently developed the TOEFL Primary Writing test as part of the TOEFL Family of Assessments. The TOEFL Primary tests are administered in a modular manner, allowing test users to choose specific components for their needs. For example, reading and listening skills are assessed through the TOEFL Primary Step 1 or Step 2 test; speaking skills, through the TOEFL Primary Speaking test; and writing skills, through the TOEFL Primary Writing test. The TOEFL Primary Step 1 and Step 2 tests are available in both paper- and computer-based formats; the TOEFL Primary Speaking and Writing tests are available in digital format only, on either a computer or a tablet.

The present document describes the framework and design principles that underpin the development of the TOEFL Primary Writing test. This design framework document is intended to inform language educators and researchers about theoretical and practical backgrounds that

guide the overall test design as well as specific task design. This document also details the scoring and score report of the TOEFL Primary Writing test to help readers better understand the interpretations and inferences made from the test results. Additionally, we discuss future research areas to garner validity evidence to investigate claims about test quality and intended uses.

Design Principles: Theoretical and Practical Backgrounds

Writing is an essential literacy skill that school-aged children should acquire for academic success. As a productive mode of communication, it is an integral component of language learning curricula that focus on communicative ability (Cumming, 2012; Lee, 2016; Manchón, 2009). Writing is also a means to express and foster both creative and logical thinking. Further, it facilitates lifelong learning pathways, encompassing interpersonal/social, academic, and vocational purposes.

For young learners who are still developing writing skills in their first language (L1), writing in a second language (L2; used collectively to include EAL, EFL, and ESL in this document) can present both linguistic and cognitive challenges. Nevertheless, it is crucial for young L2 learners to learn how to write, not only to enhance their L2 learning but also to broaden their communicative horizons (Shin & Crandall, 2019; Williams, 2012). EAL/EFL curricula for primary and secondary schools commonly incorporate L2 writing in their scopes and sequences (Butler, 2015; Hasselgreen; 2013; Patekar, 2021). Despite the importance and benefits of developing L2 writing abilities, existing literature has highlighted a paucity of L2 writing assessments tailored to young learners (Rixon & Prošić-Santovac, 2019). The literature has also pointed to the lack of professional support for teaching L2 writing to young learners (Copland *et al.*, 2014; Lee & Yuan, 2021; Patekar, 2021). Thoughtfully designed assessments can be beneficial for guiding teaching and learning. The remaining section postulates the core design principles integrated into the TOEFL Primary Writing test.

Consider Young Learners' Developmental Characteristics

Designing an age-appropriate assessment and assessment tasks was the overarching principle for the TOEFL Primary Writing test. As mentioned earlier, the target test takers for the

TOEFL Primary tests are EAL/EFL students aged 8 and older (from around the third grade in primary school to the lower grades in secondary school). A large body of literature has discussed the unique characteristics of young learners to consider in the development and use of language assessments (e.g., Bailey, 2008; Butler, 2019; Hasselgreen, 2017; Inbar-Lourie & Shohamy, 2009; Jang *et al.*, 2017; McKay, 2006; Nikolov, 2016; Papp, 2018; Rea-Dickins, 2000; Wolf & Butler, 2017). Young learners are in the midst of developing their cognitive, social, and emotional capacities. These factors should be taken into consideration when developing assessments. For example, the complexity of tasks should match the level young learners can handle, or the tasks should come with appropriate, permissible guidance in assessment settings. It is desirable for the testing duration not to exceed the typical class period duration to which young learners are accustomed (e.g., 30–45 minutes), especially if the assessment is to be administered in one sitting. The topics and content of the tasks should be based on contexts familiar to young learners (e.g., family, school). Assessments should ideally be designed to foster positive experiences so that young learners maintain their motivation and cultivate positive attitudes toward L2 writing development.

Draw on English Language Proficiency Descriptors and the Common Communication Goals Expected of Young Learners

L2 writing instruction for young learners varies widely across different contexts (Bae & Lee, 2012; Butler, 2015; Coyle *et al.*, 2018; Geng *et al.*, 2022; Patekar, 2021). While some L2 curricula for young learners prioritize oral proficiency and delay the introduction of L2 writing until secondary school, others may introduce writing instruction earlier. It is important to have a clear understanding of the proficiency expectations for young learners in designing appropriate writing tasks. A review of widely used English language proficiency descriptors for L2 learners provides an insightful foundation for developing writing tasks, rubrics, and score reports. Additionally, in the development of the TOEFL Primary Writing test, CEFR descriptors pertinent to young learners (Council of Europe, 2018a, 2018b) were closely reviewed. The CEFR levels have been used to facilitate comparisons across different educational systems and curricula (Council of Europe, 2001). However, as Hasselgreen (2013) noted, the higher levels of the CEFR (e.g., B2, C1, C2) are deemed inappropriate for primary school students due to the

advanced cognitive complexity and sociolinguistic competence inherent in these levels. A review of other English proficiency descriptors for school-aged L2 learners from countries such as Australia, Canada, and the United States alongside the CEFR descriptors has resulted in various writing tasks for young learners. Yet, when designing assessment tasks, it is important to distill the most common communication goals or language functions expected of young learners. Our review identified describing and narrating as the primary communication goals of writing for young learners.

Provide Engaging and Interesting Contexts for Writing Tasks

Previous research indicates that L2 writing poses challenges for many young learners (Bui & Luo, 2021; Copland *et al.*, 2014; Lee *et al.*, 2018). Students must not only generate and organize new ideas coherently but also transcribe them using L2 knowledge that is still in development. Thus, for young learner L2 writing instruction and assessment, it is important to provide engaging tasks with familiar topics. By doing so, assessment tasks can better elicit young learners' L2 writing samples (Hasselgreen, 2005; Rea-Dickins, 2000). The TOEFL Primary Writing test achieves the goal of providing engaging tasks by following a scenario-based assessment approach. One scenario-based assessment design feature is a purposeful sequencing of a set of items embedded in a thematic scenario (Purpura, 2016; Sabatini *et al.*, 2020). The sequence is typically intended to follow steps a skilled learner would take to solve a problem. In each step of the scenario, specific knowledge and skills are assessed. Moreover, the scenario simulates an authentic context for test takers. Writing is a social and cultural act, shaped for the purposes and audience within a context (Cushing Weigle, 2002; Lee, 2016). By providing a contextual scenario with avatar characters, the tasks in the TOEFL Primary Writing test aim to engage young learners in writing activities for meaningful communication.

Utilize Scaffolding Techniques to Model Instruction

In educational settings, scaffolding is widely used as a type of support that guides a learner toward completing tasks. This concept originates from Vygotsky's (1978) notion of the zone of proximal development, the distance between the actual developmental level where a learner can solve a problem without guidance and the level of potential development where a

learner may be able to solve a challenging problem with guidance or support. Hence, with appropriate scaffolding, the learner's capability can be increased. This concept is especially important for young learners, who may be able to better demonstrate their abilities when scaffolding is provided. In line with this notion, scaffolding is considered an effective instructional strategy for language learners (Echevarria *et al.*, 2004; Gibbons, 2002). Echevarria and her colleagues (2004) categorized three types of scaffolding for language learners: (a) verbal scaffolding such as providing guiding questions or sentence frames, (b) procedural scaffolding that involves presenting the task structure such as grouping students and modeling, and (c) instructional scaffolding related to tools used for instruction, such as visual aids and graphic organizers. Past empirical research has also demonstrated the benefits of scaffolding in the assessment of language learners, particularly young learners (Choi *et al.*, 2019; Poehner, 2013; Wolf *et al.*, 2016). In the TOEFL Primary Writing test, a mix of instructional, verbal, and procedural scaffolding techniques (e.g., visual supports, word banks, and presenting one guiding question at a time) were integrated into the design of an extended writing task to aid young learners in completing the task. This design not only mirrors good instructional practice but also aims to bring about positive effects on teaching and learning in settings where the TOEFL Primary Writing test is used.

Incorporate Technology to Prepare Students for 21st Century Skills

While it is important for young language learners to acquire L2 paper-and-pencil writing skills, the TOEFL Primary Writing test is designed to keep pace with the current digital age and prepare students for 21st century skills. The integration of computers and technology in English language learning and education is becoming increasingly prevalent (Chun *et al.*, 2016; Thorn & May, 2017), necessitating students to be able to type in English and navigate technology features. Chun and her colleagues (2016) asserted that “it is not possible to ‘opt out’ of using technology: It is so pervasive and so interwoven with human activity that to teach language without some form of technology would create a very limited and artificial learning environment—if it were even possible at all” (p. 65). They further noted that technology reshapes the way people use language and interact with each other for communication. In K–12 education settings in the United States, academic standards explicitly call on students' use of

technology. For example, one third-grade writing standard states: “With guidance and support from adults, use technology to produce and publish writing (using keyboarding skills) as well as to interact and collaborate with others” (Council of Chief State School Officers & National Governors Association, 2010, p. 21). Moreover, computer-based writing is expected not only to guide students’ writing processes by keeping records of multiple drafts easily but also to deliver feedback more efficiently. Beyond these technological advantages, the computer-based format of the TOEFL Primary Writing test enables the presentation of scenario-based contexts and scaffolding techniques for young learners.

Construct and Language Skills

The overall construct of the TOEFL Primary Writing test is defined as young EAL/EFL students’ computer-based English writing abilities to communicate about familiar topics related to their daily lives. The primary communication goals of focus in the TOEFL Primary Writing test include the following:

- describing familiar objects, people, animals, places, activities, and situations
- narrating a story to peers or adults (e.g., teachers and parents) with the aid of scaffolding
- sequencing simple events
- reviewing peer writing and making appropriate edits in order to make the text meaningful and accurate

These communication goals may be realized in both social/interpersonal and school contexts for young learners. As previously mentioned, the communication goals were selected based on a review of widely used English language proficiency standards (e.g., K–12 English language proficiency standards for L2 learners in Australia, Canada, and the United States, and the CEFR) and relevant literature regarding L2 writing for school-aged children (e.g., Bae & Lee, 2012; Bui & Luo, 2021; Hasselgreen, 2013; Lee, 2016; McKay, 2006). To perform these language functions in writing, enabling language knowledge and skills are required. In the TOEFL Primary Writing test, the following enabling, or foundational, writing skills are also measured:

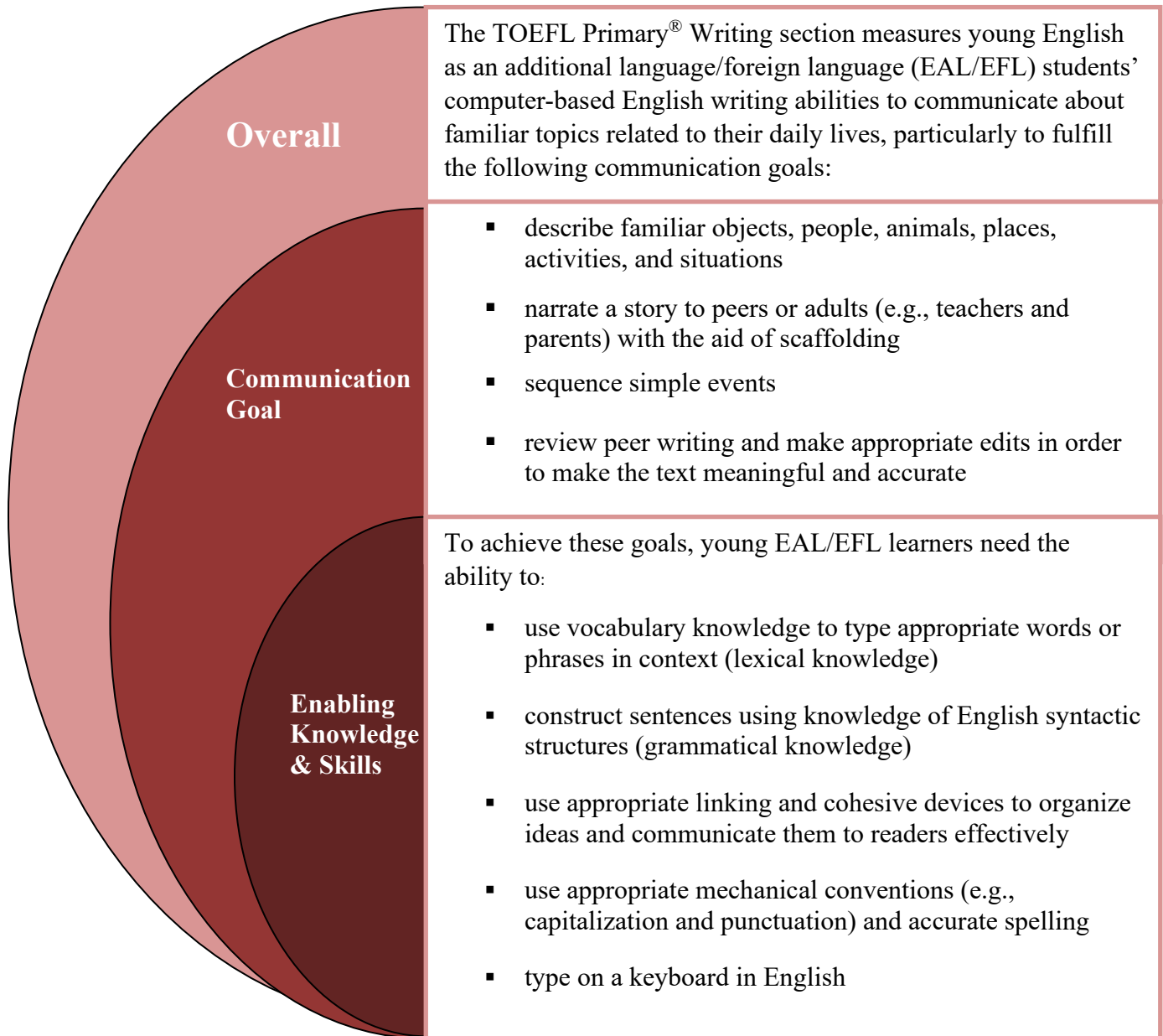
- using vocabulary knowledge to type appropriate words or phrases in context (lexical knowledge)
- constructing sentences using knowledge of English syntactic structures (grammatical knowledge)
- using appropriate linking and cohesive devices to organize ideas and communicate them to readers effectively
- using appropriate mechanical conventions (e.g., capitalization and punctuation) and accurate spelling
- typing on a keyboard in English

Figure 1 illustrates the overlapping relationships among the communication goals and the enabling knowledge and skills measured in the TOEFL Primary Writing test.

In addition to enabling skills, more macrolevel writing skills such as content development, language use, and organization (particularly coherence and cohesion) are assessed in the TOEFL Primary Writing test. These skills are commonly assessed as key characteristics that contribute to writing effective texts for communication (Cumming *et al.*, 2000; Grabe & Kaplan, 1996). The enabling skills and macrolevel skills to achieve the communication goals of describing and narrating within familiar contexts are aligned with CEFR descriptors that are relevant for young learners in levels A1 through B1.

Task Type Design

The ETS research and development team carried out multiple prototyping studies and a large-scale field test to develop suitable task types that measure the intended construct and subskills for young learners (Wolf & Suhan, 2023; Wolf *et al.*, 2023). Based on the study findings and test design principles, four task types were finalized for the TOEFL Primary Writing test (for details on the item development and review processes, see ETS, 2019). The sample questions for the TOEFL Primary Writing test can be found and accessed on the ETS website (<https://www.ets.org/toefl/primary/prepare.html>).

Figure 1. Construct of Writing

The first task, Write a Word, requires students to type a missing word in a sentence to describe a situation in a picture, assessing the ability to write appropriate words or phrases using accurate spelling and form. In the second task, Build a Sentence, students order words to form a sentence about a situation presented in a picture. This task mainly assesses the ability to construct a sentence to describe a situation, using English grammar and vocabulary knowledge.

The third task, Edit a Text, presents students with a scenario in which they must help one of the avatar characters revise a text. To do so, they read a paragraph and select the correct language forms for expressions in the text from one of three options. This task measures the ability to review texts and use knowledge of English lexico-grammar and usage to make the texts meaningful and accurate, mimicking the reviewing and editing process of writing. The final task, Write a Story, asks students to write a simple story based on a sequence of events presented in four pictures, measuring the ability to write a coherent story with appropriate details using knowledge of vocabulary, grammar, and mechanics. Table 1 summarizes information about number of items, response format, and timing for each task type. A summary of the task description is also presented in Appendix A.

Table 1. TOEFL Primary Writing Task Types

Task type	Number of items	Response format	Time allotted
Write a Word	5	Constructed response	5 minutes
Build a Sentence	5	Drag and drop/click a zone	5 minutes
Edit a Text	4	Multiple choice	5 minutes
Write a Story	1	Constructed response	12 minutes

Note. For Build a Sentence, both drag and drop and click a zone are available depending on the devices such as tablets and computers.

Tasks are situated in a scenario related to everyday or school-related situations. Avatar characters are embedded throughout the task types to present an engaging context for young learners. The use of visual elements such as avatar characters and pictures serves not only to increase students engagement but also provide support for simulating meaningful communication.

At the test level, tasks are deliberately sequenced from simpler word/phrase level communication to more extended communication with gradually increased complexity both linguistically and cognitively. In a way, the completion of earlier tasks (Write a Word, Build a Sentence, and Edit a Text) supports performance on the later task (Write a Story). For example,

the Edit a Text task models reviewing/editing, a key metacognitive learning strategy supporting performance on the Write a Story task.

For the Write a Story task, which is intended to be the most complex task, a variety of scaffolding is provided. First, key elements for a narrative text such as the names of characters and a brief description of the setting of the story is embedded in the directions along with a four-picture sequence of events. Second, students are guided through a prewriting stage with step-by-step scaffolding questions that correspond to each picture. For the first scaffolding question, an example response is provided as a modeling strategy. Third, a vocabulary word critical to the sequence of events is provided with each scaffolding question to help students respond to each question. After students respond to each scaffolding question, students are prompted to write a story based on the picture sequence. While writing their final stories, students are able to view their own scaffolding responses, presented as an outline in a bullet list, for support. Only the final story response is scored, and the previous prewriting scaffolding responses are not scored.

Scoring Scheme and Score Report

Scores for the TOEFL Primary writing test are reported on a scale of 0–17 in 1-point increments. The score is the number of correct responses to the Write a Word, Build a Sentence, and Edit a Text items plus the rating for the Write a Story task based on a 4-point scoring rubric. All items on the test are scored automatically. Table 2 provides an overview of a scoring scheme.

Table 2. Score Points for Each Task Type

Task type	Number of items	Points per item	Percentage of score
Write a Word	5	0 or 1	29%
Build a Sentence	5	0 or 1	29%
Edit a Text	4	0 or 1	24%
Write a Story	1	0 to 3	18%

Automated Scoring of the Write a Story Task

The holistic scoring rubric for the Write a Story task highlights key writing subskills such as content, cohesion, and language use (see Appendix B for the rubric descriptor). The rubric design was also informed by the CEFR overall scale for written production and relevant subscales for writing ability (Council of Europe, 2001, 2018a, 2018b). Using this rubric, the Write a Story task is scored by ETS’s automated writing evaluation (AWE) engine trained on human ratings of task responses.

To develop the AWE engine model, students’ responses ($N = 2,261$) collected from the field test were first scored by ETS’s trained human raters. During the field test, students’ responses were collected from a representative sample of the TOEFL Primary target test takers from Asia, Latin America, Europe, and the Middle East. The human raters underwent training and calibration sessions with benchmark and practice samples, applying the scoring rubric. The interrater reliability of human ratings was examined, and any disagreements were resolved through discussions to reach consensus. (Interrater reliability statistics are provided in Table 4.) These human ratings were used to train the AWE engine.

The engine utilizes a support vector regression model with a nonlinear kernel to predict a single holistic score on a scale of 0–3 based on an array of linguistic features related to the descriptors from the rubric. Table 3 summarizes the main feature categories included in the scoring engine. Note that individual features for each category are not detailed in this table. These feature categories demonstrate the conceptual connection between the AWE engine and the descriptors of the rubric for this task. For example, regarding the content dimension, the rubric states, “The response is complete with appropriate details . . . all of the (required) words are used” as presented in Appendix A. To evaluate this aspect, categories of features such as the use of required words and detailing features were included in the TOEFL Junior® AWE engine.

Table 3. Main Feature Categories Included in the Scoring Engine

Writing subskills	Feature categories
Content	Use of required words from the prompt Proportion of content words Detailing features
Cohesion and coherence	Use of linking expressions Use of pronouns Lexical cohesion
Language use	Range of syntactic structures Range of vocabulary Lexico-grammatical accuracy Use of collocations Mechanics

A series of statistical analyses were performed to determine the final AWE engine model and evaluate its performance (Li & Futagi, 2023). Table 4 exhibits a summary of reliability statistics between AWE-human and human-human ratings, providing evidence for a higher degree of reliability obtained from the AWE engine.

Table 4. Reliability Statistics

Reliability indices	Human–human	AWE–human
Exact agreement	0.78	0.83
Pearson correlation	0.85	0.91
Quadratic weighted kappa	0.85	0.90

Note. AWE = automated writing evaluation.





Score Report

In line with other score reports from the TOEFL Primary tests (i.e., Listening, Reading, and Speaking tests), the TOEFL Primary Writing test score report offers the following information: the numeric score (0–17), the CEFR level aligned to the score, a number of ribbons corresponding to the CEFR level (a band score), performance descriptors for the CEFR/ribbon level (presented as "can-do" statements), and a description of the next steps for each level. The ribbons were designed for inclusion in certificates for students taking the TOEFL Primary Writing test. The number of ribbons displayed on the certificate varies depending on the student's CEFR level. The performance descriptors were adapted from the CEFR descriptors by

the assessment development team to align with the skills and abilities evaluated by the TOEFL Primary Writing test. This score report information is intended to provide guidance for teaching and learning for teachers, school administrators, parents, and students. Based on the score, a can-do statement addressing overall writing competence is presented as a headline above a set of can-do statements related to writing subskills. The next steps are presented adjacently and include suggestions for improving writing skills at the given level. Appendix C provides the descriptors of each proficiency level and associated next steps.

To classify students' performances according to the targeted CEFR levels (i.e., A1–B1), a standard-setting study was conducted with a panel of experienced assessment specialists. The details about the study can be found in a separate report by Suhan et al. (in press). Based on the study findings, minimum scores (cut scores) for each CEFR level were established. Table 5 displays the score range for each CEFR level along with the corresponding number of ribbons indicated in the score report.

Table 5. TOEFL Primary Writing Test CEFR Levels and Number of Ribbons

CEFR level	Ribbons ^a	Score range
B1		16 – 17
A2		11 – 15
A1		6 – 10
Below A1		0 – 5

Note. CEFR = Common European Framework of Reference for Languages.

^a Graphics show the number of ribbons: B1 = 4; A2 = 3; A1 = 2; Below A1 = 1.

Directions for Validity Research

As mentioned earlier, the primary purpose of the TOEFL Primary tests is to provide useful guidance for teaching and learning. The TOEFL Primary Writing test results can be used as one piece of evidence of students' computer-based English writing proficiency in relation to the CEFR levels and students' progress in attaining English writing skills. Although the TOEFL Primary Writing test is designed for low stakes uses, validation research remains important to ensure the appropriate interpretations and uses of the test scores as intended. As described throughout this document, the development of the TOEFL Primary Writing test was informed by prior literature and empirical data from the prototyping and field-test studies. The design

document and study results supply essential evidence that supports the intended interpretations and uses of the TOEFL Primary Writing test to a degree. Beyond the evidence gathered during the test development phase, there are significant validation areas that involve data from the test users.

In this section, we outline the key research areas to accumulate empirical evidence for evaluating the validity of the interpretations and uses of the TOEFL Primary Writing test results. In organizing the research areas, we adopt an argument-based approach to validation (Bachman & Palmer, 2010; Kane, 1992, 2013, 2016). In this approach, a variety of evidence should be collected to uphold each claim and its associated warrants about the score interpretations and use of the scores. Chapelle (2008) offered a useful validation framework for the TOEFL iBT® test, leveraging the argument-based approach and organizing the diverse types of inferences linked to the claims and warrants. Drawing from this framework, we present in Table 6 key research areas related to the types of inferences and warrants for the TOEFL Primary Writing test.

Table 6. Key Validation Research Areas by the Types of Inferences and Warrants for the TOEFL Primary Writing Test

Inference	Warrant	Validation research area
Domain description	Observations of performance on the TOEFL Primary Writing test reveal relevant knowledge, skills, and abilities in situations representative of those in the target domain of language use, particularly for students in primary grades and lower secondary grades.	Target language use domain analysis for young learners' English writing activities and purposes, including various standards, curricula, and instructional materials
Evaluation	Observations of performance on the TOEFL Primary Writing test tasks are evaluated to produce scores reflective of targeted language abilities in relation to CEFR levels (i.e., foundational and communicative writing abilities in English aligned to the CEFR levels of A1 through B1).	The technical qualities of items and tasks (e.g., difficulty and discrimination) for target test takers with various backgrounds Content and linguistic analysis of students' responses to examine the correspondence to the characteristics of the targeted CEFR levels

		The quality and adequacy of the automated scoring model to predict human scores
Generalization	Observed scores are estimates of expected scores over the relevant parallel versions of tasks and test forms.	<p>Comparability of the relevant parallel versions of tasks in terms of the expected linguistic and cognitive characteristics (e.g., prompt/topic effects)</p> <p>Reliability of the test</p>
Explanation	Expected scores are attributed to the relevant construct of computer-based English writing abilities in daily life and school contexts.	<p>Factor analysis to examine the internal structure of the test, as conformed to the test design</p> <p>Relationship among different task types within the TOEFL Primary Writing test</p> <p>Relationship with scores of other TOEFL Primary tests (reading, listening, speaking)</p> <p>Relationship with students' English writing development</p> <p>Students' test-taking processes and strategies</p>
Extrapolation	The construct of computer-based English writing abilities as assessed by the TOEFL Primary Writing test accounts for the linguistic performance in English-medium instruction settings at primary and lower secondary grades.	<p>Relationships with other measures of a similar construct (e.g., teacher ratings, other writing tests, self-evaluation, course placement)</p> <p>Longitudinal analysis of the TOEFL Primary writing scores and constructed responses for repeated test takers</p>
Utilization	<p>Scores and information provided from the TOEFL Primary Writing test are useful for guiding the teaching and learning of English writing skills for young learners.</p> <p>TOEFL Primary Writing scores are also useful for tracking progress in</p>	<p>Washback/impact studies to examine positive and/or unintended negative consequences</p> <ul style="list-style-type: none"> • teachers/parents/students' increased understanding about students' writing skills

students' computer-based English writing skills.

- teachers' use of the test results for instructional planning
 - students' increase in positive attitudes and motivation toward English writing skills
 - students' development of English writing skills over time
-

Note. CEFR = Common European Framework of Reference for Languages.

It is important to note that the warrants and validation research areas included in Table 6 are not exhaustive but represent prioritized considerations. As stated in the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), “validation is the joint responsibility of test developer and the test user” (p. 13). To support the intended purposes of the TOEFL Primary Writing test, ETS is committed to monitoring the technical quality of the test and conducting ongoing investigations to furnish validity evidence. Regularly examining the item statistics across different test-taker groups (e.g., by region, countries, grade, age) for different forms is a standard procedure that ETS performs as a test provider. Furthermore, the ongoing evaluation of the adequacy of the AWE engine and scoring model is crucial for ETS research and development because of the rapid advancements in technology, artificial intelligence capabilities, and the diverse backgrounds of young learners. Validity evidence can also be examined by other external researchers. Collectively, these endeavors can contribute to the continuous improvements of the TOEFL Primary Writing test, ultimately benefitting young EAL/EFL learners and their educators.

References

- American Educational Research Association, the American Psychological Association, & the National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Bae, J., & Lee, Y.-S. (2012). Evaluating the development of children's writing ability in an EFL context. *Language Assessment Quarterly*, 9(4), 348–374.
<https://doi.org/10.1080/15434303.2012.721424>
- Bailey, A. L. (2008). Assessing the language of young learners. In N. Hornberger (Ed.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (2nd ed., pp. 2509–2528). Springer. https://doi.org/10.1007/978-0-387-30424-3_188
- Bui, G., & Luo, X. (2021). Topic familiarity and story continuation in young English as a foreign language learners' writing tasks. *Studies in Second Language Learning and Teaching*, 11(3), 377–400. <https://doi.org/10.14746/ssl1t.2021.11.3.4>
- Butler, Y. G. (2015). English language education among young learners in East Asia: A review of current research (2004–2014). *Language Teaching*, 48(3), 303–342.
<https://doi.org/10.1017/S0261444815000105>
- Butler, Y. G. (2019). Assessment of young English learners in instructional settings. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 477–496). Springer.
https://link.springer.com/referenceworkentry/10.1007/978-3-030-02899-2_24
- Chapelle, C. A. (2008). The TOEFL® validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319–352). Routledge. <https://doi.org/10.4324/9780203937891-9>
- Choi, I., Wolf, M. K., Pooler, E., Sova, L., & Faulkner-Bond, M. (2019). Investigating the benefits of scaffolding in assessments of young English learners: A case for scaffolded retell tasks. *Language Assessment Quarterly*, 16(2), 161–179.
<https://doi.org/10.1080/15434303.2019.1619180>

- Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal*, 100(S1), 64–80.
<https://doi.org/10.1111/modl.12302>
- Copland, F., Garton, S., & Burns, A. (2014). Challenges in teaching English to young learners: Global perspectives and local realities. *TESOL Quarterly*, 48(4), 738–762.
<https://doi.org/10.1002/tesq.148>
- Council of Chief State School Officers & National Governors Association. (2010). *Common Core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. <https://learning.ccsso.org/wp-content/uploads/2022/11/ADA-Compliant-ELA-Standards.pdf>
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
<https://rm.coe.int/1680459f97>
- Council of Europe. (2018a). *Collated representative samples of descriptors of language competences developed for young learners. Resource for educators* (Vol. 1: Ages 7–10).
<https://rm.coe.int/16808b1688>
- Council of Europe. (2018b). *Collated representative samples of descriptors of language competences developed for young learners. Resource for educators* (Vol. 2: Ages 11–15).
<https://rm.coe.int/collated-representative-samples-descriptors-young-learners-volume-2-ag/16808b1689>
- Coyle, Y., Guirao, J. C., & de Larios, J. R. (2018, December). Identifying the trajectories of young EFL learners across multi-stage writing and feedback processing tasks with model texts. *Journal of Second Language Writing*, 42, 25–43.
<https://doi.org/10.1016/j.jslw.2018.09.002>
- Cumming, A. (Ed.). (2012). *Adolescent literacies in a multicultural context*. Routledge.
<https://doi.org/10.4324/9780203120033>
- Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph No. MS-18). ETS.
<http://www.ets.org/Media/Research/pdf/RM-00-05.pdf>

- Cushing Weigle, S. (2002). *Assessing writing*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511732997>
- Echevarria, J., Vogt, M. E., & Short, D. J. (2004). *Making content comprehensible for English language learners: The SIOP model* (2nd ed.). Allyn & Bacon/Pearson.
- ETS. (2019). *TOEFL Primary® framework and test development* (TOEFL® Research Insight, Vol. 8).
<https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v8.pdf>
- ETS. (2023). *TOEFL Primary® test taker handbook*. <https://www.ets.org/pdfs/toefl/toefl-primary-handbook.pdf>
- Geng, F., Yu, S., Liu, C., & Liu, Z. (2022). Teaching and learning writing in English as a foreign language (EFL) school education contexts: A thematic review. *Scandinavian Journal of Educational Research*, 66(3), 491–504.
<https://doi.org/10.1080/00313831.2021.1897872>
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom*. Heinemann.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. Routledge. <https://doi.org/10.4324/9781315835853>
- Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337–354. <http://doi.org/10.1191/0265532205lt312oa>
- Hasselgreen, A. (2013). Adapting the CEFR for the classroom assessment of young learners' writing. *Canadian Modern Language Journal*, 69(4), 415–435.
<https://doi.org/10.3138/cmlr.1705.415>
- Hasselgreen, A. (2017). Assessing young learners. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 93–105). Routledge.
<https://doi.org/10.4324/9780203181287-14>
- Inbar-Lourie, O., & Shohamy, E. (2009). Assessing young language learners: What is the construct? In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 83–96). De Gruyter. <https://doi.org/10.1515/9783110218282.83>
- Jang, E. E., Vincett, M., van der Boom, E. H., Lau, C., & Yang, Y. (2017). Considering young learners' characteristics in developing a diagnostic assessment intervention. In M. K.

- Wolf, & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 193–213). Routledge. <https://doi.org/10.4324/9781315674391-11>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Lee, I. (2016). EFL writing in schools. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 113–140). De Gruyter Mouton. <https://doi.org/10.1515/9781614511335-008>
- Lee, I., Yu, S., & Liu, Y. (2018). Hong Kong secondary students' motivation in EFL writing: A survey study. *TESOL Quarterly*, *52*(1), 176–187. <https://doi.org/10.1002/tesq.364>
- Lee, I., & Yuan, R. (2021). Understanding L2 writing teacher expertise. *Journal of Second Language Writing*, *52*, Article 100755. <https://doi.org/10.1016/j.jslw.2020.100755>
- Li, F., & Futagi, Y. (2023, October 5). *Evaluating the automated scoring of the TOEFL Junior® and TOEFL Primary® writing tests* [Paper presentation]. TOEFL® Research Symposium, Princeton, NJ, United States.
- Manchón, R. (Ed.). (2009). *Writing in foreign language contexts: Learning, teaching, and research*. Multilingual Matters.
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733093>
- Nikolov, M. (2016). Trends, issues, and challenges in assessing young language learners. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 1–17). Springer. https://doi.org/10.1007/978-3-319-22422-0_1
- Papp, S. (2018). Assessment of young English language learners. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 389–408). Routledge. <https://doi.org/10.4324/9781315623672-25>

- Patekar, J. (2021). A look into the practices and challenges of assessing young EFL learners' writing in Croatia. *Language Testing*, 38(3), 456–479.
<https://doi.org/10.1177/0265532221990657>
- Poehner, M. E. (2013). Dynamic assessment in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–8). Blackwell Publishing Ltd.
<https://doi.org/10.1002/9781405198431.wbeal0345>
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100 (S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Rea-Dickins, P. (2000). Assessment in early years language learning contexts. *Language Testing*, 17(2), 115–122. <https://doi.org/10.1177/026553220001700201>
- Rixon, S., & Prošić-Santovac, D. (2019). Introduction: Assessment and early language learning. In D. Prošić-Santovac & S. Rixon (Eds.), *Integrating assessment into early language learning and teaching* (pp. 29–56). Multilingual Matters. <https://doi.org/10.21832/PROSIC4818>
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, 20(1), 1–23.
<https://doi.org/10.1080/15305058.2018.1551224>
- Shin, J. K., & Crandall, J. (2019). Teaching reading and writing to young learners. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 188–202). Routledge. <https://doi.org/10.4324/9781315623672-13>
- Suhan, M., Papageorgiou, S., & Wolf, M. K. (in press). *Mapping the scores of the TOEFL Primary® Writing test to the Common European Framework of Reference (CEFR) levels* (Research Memorandum No. RM-24-03). ETS.
- Thorn, S. L., & May, S. (Eds.). (2017). *Language, education, and technology* (3rd ed.). Springer.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Harvard University Press.
- Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21(4), 321–331. <https://doi.org/10.1016/j.jslw.2012.09.007>

- Wolf, M. K., & Butler, Y. G. (2017). An overview of English language proficiency assessments for young learners. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 3–21). Routledge.
<https://doi.org/10.4324/9781315674391-1>
- Wolf, M. K., Deane, P., Chen, L., Choi, I., & Suhan, M. (2023, June 5–9). *Young EFL students' writing processes during a computer-based assessment: An examination of keystroke logs* [Paper presentation]. Language Testing Research Colloquium (LTRC) Conference, New York, New York, United States.
- Wolf, M. K., Guzman-Orth, D., Lopez, A., Castellano, K., Himelfarb, I., & Tsutagawa, F. (2016). Integrating scaffolding strategies into technology-enhanced assessments of English learners: Task types and measurement models. *Educational Assessment*, 21(3), 151–175. <https://doi.org/10.1080/10627197.2016.1202107>
- Wolf, M. K., & Suhan, M. (2023, March 21–24). *Providing scaffolding in a writing assessment for young EFL students* [Paper presentation]. TESOL Conference, Portland, Oregon, United States.

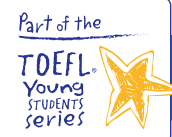
Appendix A: Task Type Description

Task type	Description and subskills measured	Response format & score point	# of items
Write a Word	Students type a missing word in a sentence in order to describe a situation presented in a picture The ability to write words using the accurate form to describe a situation in a simple sentence	Constructed Response 0 or 1	5
Build a Sentence	Students construct a sentence to describe a situation presented in a picture The ability to construct a sentence to describe a situation using knowledge of English syntax and vocabulary	Drag & Drop, Clicking a Zone 0 or 1	5
Edit a Text	Students read a paragraph and select correct language forms The ability to review written texts and use knowledge of English grammar and usage in order to make the texts meaningful and accurate	Multiple Choice 0 or 1	4
Write a Story	Students write a story based on a sequence of events presented in four pictures The ability to write a coherent story with appropriate details using knowledge of English vocabulary, grammar, and mechanics	Constructed Response 0 to 3	1

Appendix B: Scoring Rubric for the Write a Story Task

Score	Development and language use descriptors
3	<p>The test taker achieves the communication goal. A typical response at this level is characterized by the following:</p> <ul style="list-style-type: none"> • The response is complete with appropriate details. For items with a required word list, all of the words are used. • The response maintains coherence with the support of cohesive devices (e.g., pronouns, transition words). • The language demonstrates accuracy and/or variety in word choice, grammar, and mechanics (e.g., capitalization, punctuation, spelling), though a few errors may be present.
2	<p>The test taker partially achieves the communication goal. A typical response at this level is characterized by the following:</p> <ul style="list-style-type: none"> • The response is partially complete, with some appropriate details. For items with a required word list, some of the words are used. • Parts of the response are coherent. Limitations or inaccuracies in the use of cohesive devices weaken the overall coherence. • The language demonstrates a lack of variety or control of sentence structures and may include multiple errors in word choice, grammar, and mechanics (e.g., missing punctuation or inaccurate spelling).
1	<p>The test taker attempts to achieve the communication goal. A typical response at this level is characterized by the following:</p> <ul style="list-style-type: none"> • The response is incomplete, perhaps addressing only one picture beyond the given sentence or one aspect of the descriptive prompt. Appropriate details may be expressed in single words, short phrases, or even a single sentence. For items with a required word list, few, if any, of the words are used. • The response is mostly incoherent. • The word choice is basic and/or repetitive, and the grammar and mechanics are mostly inaccurate. Major errors are present throughout the response, or the response is too short to evaluate language use.
0	<p>A typical response at this level may be:</p> <ul style="list-style-type: none"> • Off-topic (e.g., a memorized response to a different question) • Entirely in another language • Random strings of letters • No response (i.e., blank) • A copy of the prompt or provided scaffolding language (with no attempt to modify or create new language) • Contains only “I don’t know”

Appendix C: TOEFL Primary® Test—Writing Score Level Descriptors



Ribbons	Can Do	Next Steps
4	<p>Students at this score range are typically able to produce short, coherent texts with details and mostly accurate language use. They typically can:</p> <ul style="list-style-type: none"> • Use their vocabulary knowledge to consistently name and describe a wide range of everyday actions and objects • Produce short narrative texts with details that describe everyday events • Use transition words to maintain coherence throughout a text (examples: <i>and, then, but, first, next, finally, and after</i>) • Construct simple and complex sentences with mostly correct syntax, word choice, and grammatical form, and with adequate capitalization and punctuation 	<p>To improve their writing ability, students should practice:</p> <ul style="list-style-type: none"> • Producing longer narrative texts about both everyday events and unfamiliar situations • Writing well-organized paragraphs for personal and academic purposes, such as longer messages to friends, opinions, and summaries of academic topics • Using a wide variety of vocabulary, sentence structures, and grammatical forms <p>Students may also consider taking the TOEFL Junior® Writing test to learn more about their writing ability</p>
3	<p>Students at this score range are typically able to describe familiar situations and begin to connect ideas in narratives. They typically can:</p> <ul style="list-style-type: none"> • Use their vocabulary knowledge to name and describe some everyday actions and objects, such as daily routines, clothes, body parts, animals, and school supplies • Produce short narrative texts with some details about everyday events • Use transition words to give coherence to parts of a text (examples: <i>and, then, but, first, next, and finally</i>) • Construct simple sentences and questions with some correct syntax, word choice, and grammatical forms 	<p>To improve their writing ability, students should practice:</p> <ul style="list-style-type: none"> • Producing short narrative texts in which all the details are connected so that the entire story is coherent • Writing complex sentences by using a wider vocabulary and a variety of grammatical forms and structures
2	<p>Students at this score range are typically able to describe some objects, places, and routines, and they begin to write narrative texts. They typically can:</p> <ul style="list-style-type: none"> • Write common words describing familiar objects and activities at school and at home • Attempt to write short narrative texts with a few short sentences • Construct short, simple sentences 	<p>To improve their writing ability, students should practice:</p> <ul style="list-style-type: none"> • Producing short descriptive and narrative texts about everyday events • Adding details to sentences • Connecting details with transition words (examples: <i>and, then, but, first, next, and finally</i>) • Writing simple and complex sentences
1	<p>Students at this score range attempt to describe familiar situations using words and phrases. They typically can:</p> <ul style="list-style-type: none"> • Use phonetic knowledge in attempts to write basic words and phrases • Use a basic noun-verb syntactic structure in attempts to write sentences 	<p>To improve their writing ability, students should practice:</p> <ul style="list-style-type: none"> • Writing words and phrases that name everyday objects, activities at school and at home, and places they visit • Producing simple sentences to describe familiar topics and situations