



Research Memorandum

ETS RM–19-10

Development of the Redesigned *TOEIC Bridge*® Tests

Philip Everson

Elizabeth Carter Grissom

Trina Duke

Elizabeth Park

Pablo Garcia Gomez

Jonathan Schmidgall

August 2019

ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Senior Research Scientist

Heather Buzick
Senior Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Director

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Consultant

Sooyeon Kim
Principal Psychometrician

Anastassia Loukina
Research Scientist

John Mazzeo
Distinguished Presidential Appointee

Donald Powers
Principal Research Scientist

Gautam Puhan
Principal Psychometrician

John Sabatini
Managing Principal Research Scientist

Elizabeth Stone
Research Scientist

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ariela Katz
Proofreader

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.

Development of the Redesigned *TOEIC Bridge*[®] Tests

Philip Everson, Trina Duke, Pablo Garcia Gomez, Elizabeth Carter Grissom,
Elizabeth Park, and Jonathan Schmidgall
Educational Testing Service, Princeton, New Jersey

August 2019

Corresponding author: P. Everson, E-mail: peverson@ets.org

Suggested citation: Everson, P., Duke, T., Garcia Gomez, P., Carter Grissom, E., Park, E., & Schmidgall, J. (2019). *Development of the redesigned TOEIC Bridge[®] tests* (Research Memorandum No. RM-19-10). Princeton, NJ: Educational Testing Service.

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

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

Action Editor: Donald Powers

Reviewers: Elizabeth Bredlau and Yeonsuk Cho

Copyright © 2019 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, MEASURING THE POWER OF LEARNING., TOEIC, and TOEIC BRIDGE are registered trademarks of Educational Testing Service (ETS). All other trademarks are the property of their respective owners.



Abstract

The suite of *TOEIC*[®] tests measures the English language proficiency of nonnative speakers in the international workplace and everyday life. The TOEIC test focuses on measuring intermediate to advanced workplace English proficiency. The *TOEIC Bridge*[®] test focuses on measuring beginning to intermediate everyday English proficiency. The TOEIC Listening and Reading test has been in use since 1979. In response to market needs for direct measures of speaking and writing skills, TOEIC Writing and Speaking tests were added in 2006 to the suite of TOEIC tests. The TOEIC Bridge test was introduced in 2001 as a measure of the listening and reading abilities of beginning to intermediate English learners. In 2017, in order to bring the TOEIC Bridge test into alignment with the TOEIC tests and to more fully meet test taker and score user needs, the TOEIC program initiated the design of a 4-skills suite of TOEIC Bridge assessments. Because the TOEIC program recognizes the evolving nature of international English communication, the decision was made to not simply add speaking and writing assessments to the existing TOEIC Bridge Listening and Reading tests but also design new assessments for all 4 skills.

Key words: assessment, design, *TOEIC Bridge*[®], four skills, task design, item specifications

Table of Contents

	Page
Task Modeling	3
Task Design	8
Speaking Task Modeling	13
Item Prototyping Usability Study	18
Pilot Testing	18
Results of Pilot Testing	19
Scoring the Constructed Responses From Pilot Testing	22
Field Study	28
Field Test Usability Study	30
Rater Survey	30
Final Adjustment to Item Presentations	32
Proficiency Descriptors	33
Conclusion	34
References	35
Appendix A. Redesigned TOEIC Bridge Speaking and Writing Tests Scoring Guides	36
Appendix B. Redesigned TOEIC Bridge Proficiency Descriptors	46

The test design process for the redesigned *TOEIC Bridge*[®] tests was a collaboration among researchers, content developers, psychometricians, and the business directors of the *TOEIC*[®] program following a process of evidence-centered design (ECD). ECD can be viewed as a methodology that comprises best practices for the creation and ongoing development of an assessment. It clarifies what is being measured by a test and supports inferences made on the basis of evidence derived from the test. ECD systematizes test design by specifying a process with five stages or layers, including domain analysis, domain modeling, construction of an assessment framework, assessment implementation, and assessment delivery (Mislevy & Yin, 2012). As shown in Figure 1, these stages concretize what we want to be able to say about test takers based on observations we make on their performance on the test tasks.

Layer	Role	Key entities or components	Explanation of key entity or component
1. Domain analysis	Gather information about what is to be assessed	Analysis and summary of theory, research, and expert judgment as it pertains to what is to be assessed	Language framework, proficiency guidelines, etc.
2. Domain modeling	Incorporation of information from stage one into three components; sketch of potential variables and substantive relationships	Proficiency paradigm	Substantive construct expressed as claims
		Evidence paradigm	Observations required to support claims
		Task paradigm	Types of situations that provide opportunities for test takers to show evidence of their proficiencies
3. Construction of conceptual assessment design framework	Development of a final blueprint; provide technical detail required for implementation including statistical models, rubrics, specifications, and operational requirements	Student model	Statistical characterization of the abilities to be assessed
		Evidence model	1. Rules for scoring test tasks
			2. Rules for updating variables in the student model
		Task model	Detailed description of assessment tasks
		Presentation model	Specification of how the assessment elements will look during testing
Assembly model	Specification of the mix of tasks on a test for a particular student		
4. Assessment implementation	Operational item writing and form assembly	Task materials, work products, operational data	Rendering protocols for tasks, tasks as displayed, etc.
5. Assessment delivery	Test administration and scoring	Tasks as presented, work products as created, scores as evaluated	Actual rendering of task materials in assessment, score reports, etc.

Figure 1. Layers of the evidence-centered design process. Green indicates evidence-centered design steps addressed in this research memorandum.

This research memorandum is concerned primarily with the development of the ECD steps shaded in green in Figure 1:

- Task models
- Presentation models
- Assembly models

The other steps in the process that precede and follow these three are discussed in Lin et al. (2019) and Schmidgall et al. (2019).

Task modeling begins with the development of prototype tasks. Multiple tasks were developed for each of the four assessments. In many cases, two or more versions of the same prototype task were developed, where the specifications for the versions varied in some important way—for instance, different response times or different levels of specificity in the directions. Prototype tasks were evaluated through small-scale user-acceptance testing and larger scale piloting. Through pilot testing, developers were able to finalize task specifications and, for speaking and writing, finalize the rubrics used to score productive tasks.

To a certain extent, task modeling overlaps with the evidence paradigm and the task paradigm from the domain modeling stage of ECD. The domain definitions for the redesigned TOEIC Bridge tests were based on the proposed construct definition that was a result of the domain analysis stage, described in detail in Schmidgall et al. (2019). The domain definitions include communication goals, and the communication goals are, for the most part, definitions of task paradigms. They state, at an abstract level, the kinds of situations that allow test takers to show evidence of ability. In the case of the listening domain of the redesigned TOEIC Bridge, the domain description includes the communication goals (among others) of “understand short, simple descriptions” and “understand short conversations.” These communication goals define the kinds of tasks that would be appropriate to include in an operational assessment aligned with the domain definitions.

If task models are concerned with representing as fully as possible specific communication goals, or evidence paradigms, as they occur in the real world, presentation models focus on the task types as test items and evaluate the tasks from the point of view of the test taker. Primary questions include the following: Is the task accessible? Do test takers know what they are supposed to do? If the task is timed, do test takers have adequate time to consider

and complete the task? Are all the tools available in the testing platform easy to access and use? These questions are particularly important for an assessment like the redesigned TOEIC Bridge tests because directions and collateral material are in English, and the test takers are beginning to intermediate English learners.

After pilot testing, test developers were able to create draft test blueprints for each of the four assessments. Pilot testing provided evidence for which prototype tasks or versions of tasks produced usable evidence to support the claims derived from the domain model. The draft test blueprints were used to create the forms to be administered in the field test.

Task Modeling

The process of designing task prototypes for the redesigned TOEIC Bridge suite of assessments began with discussions of the program requirements that were necessary to make the final product useful in the marketplace and that affected test design. These program requirements informed the initial domain analysis and construct definitions for the redesigned TOEIC Bridge assessments as described in Schmidgall et al. (2019) but led to additional considerations for task modeling that initiated the process of operationalizing the construct definition.

The following is a partial list of the business requirements that were most relevant to assessment design:

- The redesigned TOEIC Bridge tests will measure all four language skills—listening, reading, speaking, and writing—and provide scores and feedback on each.
- Each of the four assessments will focus on representative communication skills at the A1, A2, and B1 levels of the Common European Framework of Reference (CEFR).
- The tests will be module based so that score users can require and test takers can take different combinations of skills.
- The listening and reading assessments will be administered on paper but designed so that future computer-delivered versions will be possible.
- The speaking and writing assessments will be computer based.
- The listening and reading assessments will be machine scored.
- The speaking and writing assessments will be scored by human raters.

- The combined testing time for the listening and reading assessments should not exceed the testing time of the existing TOEIC Bridge test.
- Accents from the United States, Canada, the United Kingdom, and Australia will be used in the listening and speaking stimulus materials.
- The assessments will include, where possible, contemporary means of communication, such as e-mail and instant messages.
- The assessment design will promote meaningful mapping to the CEFR.
- The assessments will provide meaningful feedback to teachers and learners in the form of proficiency descriptors.

Some of these requirements were motivated by the desire that the redesigned TOEIC Bridge assessments be consistent in important respects with other components of the TOEIC family of assessments—for instance, the inclusion of varied accents in the listening assessment. Others were motivated by considerations of the overall cost structure of the operational assessment, such as the use of human raters to score the speaking and writing tests. They all were taken into account in task design so that the final assessment design was helpful to end users. The requirements that the redesigned TOEIC Bridge assessments be meaningfully mapped to the CEFR and other internationally recognized language standards and that they provide appropriate feedback to teachers and learners made following an ECD process especially important.

A second, and equally important, set of guidelines for prototype task development was the product of the domain analysis ECD step, as described in Schmidgall *et al.* (2019). The first product was the definition of the assessments' overall target language use (TLU) domain. The TLU was defined as “everyday adult life” and included three subdomains: the personal sphere, the public sphere, and the workplace sphere. Building on the overall definition of the TLU domain of everyday adult life, the test designers then created domain definitions for each of the four skills—listening, reading, speaking and writing—with explicit communication goals and underlying competencies that support the successful completion of the communication goals. These domain definitions also incorporated information from a principled review of the language proficiency standards expected to be most relevant to score users, including the CEFR standards, Canadian Language Benchmarks, and American Council on the Teaching of Foreign Language's

proficiency guidelines. The review of language proficiency standards also produced summaries of the language activities, strategies, and competencies relevant to the range of proficiency levels targeted by the test (i.e., CEFR A1 to B1) that informed test development. Figures 2–5 show the four domain definitions that guided task development for each section of the redesigned TOEIC Bridge.

<p>Listening Domain Definition</p> <p>The TOEIC Bridge Listening section measures the ability of beginning to lower-intermediate English language learners to understand short spoken conversations and talks in personal, public, and familiar workplace contexts. This includes the ability to understand high-frequency vocabulary, formulaic phrases, and the main ideas and supporting details of clearly articulated speech across familiar varieties of English on familiar topics. Test takers can comprehend simple greetings, introductions, requests, instructions, and directions; descriptions of people, objects, situations, personal experiences, or routines; and other basic exchanges of information.</p>
<p>Communication Goals</p> <p>In English, test takers can understand commonly occurring spoken texts, demonstrating the ability to</p> <ul style="list-style-type: none"> ▪ understand simple descriptions of people, places, objects, and actions ▪ understand short dialogues or conversations on topics related to everyday life (e.g., making a purchase) ▪ understand short spoken monologues as they occur in everyday life (e.g., an announcement in a public area) when they are spoken slowly and clearly
<p>Linguistic Knowledge and Subcompetencies</p> <p>To achieve these goals, beginning and lower-intermediate English language learners need the ability to</p> <ul style="list-style-type: none"> ▪ understand common vocabulary and formulaic phrases (lexical knowledge) ▪ understand simple sentences and structures (grammatical knowledge) ▪ understand sentence-length speech and some common registers (discourse knowledge) ▪ recognize and distinguish English phonemes and the use of common intonation and stress patterns and pauses to convey meaning in slow and carefully articulated speech across familiar varieties (phonological knowledge) ▪ infer implied meanings, speaker roles, or context in short, simple spoken texts (pragmatic competence) ▪ understand the main idea and stated details in short, spoken texts (listening strategies)

Figure 2. Listening domain definition.

<p>Reading Domain Definition</p> <p>The TOEIC Bridge Reading section measures the ability of beginning and lower-intermediate English language learners to understand short written English texts in personal, public, and familiar workplace contexts and across a range of formats. This includes the ability to understand high-frequency vocabulary, formulaic phrases, and the main ideas and supporting details of short, written texts dealing with familiar topics. Test takers can comprehend simple texts such as signs, lists, menus, schedules, advertisements, narrations, routine correspondence, and short descriptive texts.</p>
<p>Communication Goals</p> <p>In English, test takers can understand commonly occurring written texts, demonstrating the ability to</p> <ul style="list-style-type: none">▪ understand nonlinear written texts (e.g. signs, schedules)▪ understand written instructions and directions▪ understand short, simple correspondence▪ understand short informational, descriptive, and expository written texts about people, places, objects, and actions
<p>Linguistic Knowledge and Subcompetencies</p> <p>To achieve these goals, beginning and lower-intermediate English language learners need the ability to</p> <ul style="list-style-type: none">▪ understand common vocabulary (lexical knowledge)▪ understand simple sentences and structures (grammatical knowledge)▪ understand the organization of short written texts in a variety of formats (discourse knowledge)▪ recognize simple mechanical conventions of written English (orthographic knowledge)▪ infer implied meanings, including context or writer’s purpose in short, simple written texts (pragmatic competence)▪ understand the main idea and stated details in short, written texts; infer the meaning of unknown written words through context clues (reading strategies)

Figure 3. Reading domain definition.

<p>Speaking Domain Definition</p> <p>The TOEIC Bridge Speaking section measures the ability of beginning and lower-intermediate English language learners to carry out spoken communication tasks in personal, public, and familiar workplace contexts. This includes the ability to communicate immediate needs, provide basic information, and interact on topics of personal interest with people who are speaking clearly. Test takers can answer simple questions on familiar topics and use phrases and sentences to describe everyday events. They can provide brief reasons for and explanations of their opinions and plans and narrate simple stories.</p>
<p>Communication Goals</p> <p>In spoken English, perform simple communication tasks, demonstrating the ability to</p> <ul style="list-style-type: none">▪ ask for and provide basic information▪ describe people, objects, places, activities▪ express an opinion or plan and give a reason for it▪ give simple directions▪ make simple requests, offers, and suggestions▪ narrate and sequence simple events
<p>Linguistic Knowledge and Subcompetencies</p> <p>To achieve these goals, beginning and lower-intermediate English language learners need the ability to</p> <ul style="list-style-type: none">▪ use high-frequency vocabulary appropriate to a task (lexical knowledge)▪ use common grammar structures (grammatical knowledge)▪ use simple transitions to connect ideas, e.g., <i>so</i>, <i>but</i>, <i>after</i> (discourse knowledge)▪ pronounce words in a way that is intelligible to native speakers and proficient nonnative speakers of English; use intonation, stress, and pauses to pace speech and contribute to comprehensibility (phonological knowledge)▪ produce speech that is appropriate to the communication goal (pragmatic competence)

Figure 4. Speaking domain definition.

<p>Writing Domain Definition</p> <p>The TOEIC Bridge Writing test measures the ability of beginning and lower-intermediate English language learners to carry out written communication tasks in personal, public, and familiar workplace contexts. This includes the ability to use high-frequency vocabulary and basic grammar structures to produce phrases, sentences, and paragraphs on subjects that are familiar or of personal interest. Test takers can write notes and messages relating to matters of immediate need. They can write simple texts such as personal letters describing experiences and giving simple opinions.</p>
<p>Communication Goals</p> <p>In written English, perform simple communication tasks, demonstrating the ability to</p> <ul style="list-style-type: none"> ▪ ask for and provide basic information ▪ make simple requests, offers, and suggestions; express thanks ▪ express a simple opinion and give a reason for it ▪ describe people, objects, places, activities ▪ narrate and sequence simple events
<p>Linguistic Knowledge and Subcompetencies</p> <p>To achieve these goals, beginning and lower-intermediate English language learners need the ability to</p> <ul style="list-style-type: none"> ▪ use high-frequency vocabulary appropriate to a task (lexical knowledge) ▪ write a sentence using simple word order, such as SVO (subject/verb/object); interrogatives; imperatives; use common grammatical structures to contribute to meaning (grammatical knowledge) ▪ arrange ideas using appropriate connectors (e.g., <i>for example, in addition, finally</i>); sequence ideas to facilitate understanding (discourse knowledge) ▪ use mechanical conventions of English (spelling, punctuation, and capitalization) to facilitate comprehensibility of text (orthographic knowledge) ▪ produce text that is appropriate to the communication goal (pragmatic competence)

Figure 5. Writing domain definition.

Task Design

It should be noted that at the beginning of the task model development process, the communication goals to be measured on the assessment are aspirational. That is, test designers may be more or less successful in creating tasks that are authentic and valid representations of the communication goals. Communication tasks that are carried out within the subdomains of everyday adult life (personal, public, and familiar workplace contexts) at beginning to low intermediate levels of proficiency cannot always be exactly replicated in a language test. To the extent possible, real-world tasks are used or approximated in the TOEIC family of tests to maximize the validity of test scores and proficiency descriptors.

For each language skill, the domain definitions outline the claims to be made about test-taker abilities. More tasks were developed for pilot testing than final versions of the tests would contain, and specific questions were posed so that pilot test results would better inform later decisions. These decisions included task choice as well as specific task characteristics such as the presentation of items, preparation and response times, and rubric refinement. It should be noted that the pilot forms were not intended to be draft versions of operational forms but merely the delivery of individual tasks that designers considered to be likely candidates for operational use.

Listening task modeling. In developing prototype tasks for the redesigned TOEIC Bridge Listening assessment, 12 different tasks or task variants were considered before the pilot. Of these, five were chosen for piloting based on the criteria of most efficient representation of the construct, as shown in Table 1. The pilot tasks included the following (tasks are identified here by the shorthand name used during design discussions):

1. **Photographs.** The test taker looks at a photograph, listens to four 1-sentence options, and chooses the option that best describes the content of the photograph.
2. **Four Pictures (Listening).** The test taker listens to a one-sentence description of a person, place, or object and then selects from four graphic options the picture that is consistent with the stimulus.
3. **Question-Response.** The test taker listens to the first half of a conversational exchange and then selects from four options the response appropriate to the exchange.
4. **Conversations.** The test taker listens to a short conversation, and comprehension is assessed by a set of multiple-choice questions. In a variant of this task, the spoken stimulus is supplemented by a short, very simple text graphic, such as a schedule or address, and one or more of the multiple-choice questions requires the test taker to connect information in the audio stimulus to information in the text.
5. **Talks.** The test taker listens to a monologue, and comprehension is assessed by sets of multiple-choice questions. As with conversations, some monologues are supplemented by simple graphical information that must be synthesized with the spoken information to answer one or more questions.

Table 1 shows how the listening tasks were expected to align with the CEFR levels that the redesigned TOEIC Bridge tests (A1-B1) were meant to measure, the communication goals

outlined in the listening domain definition, and the relevant linguistic subskills. All the listening prototype tasks are shown in the table to align with more than one CEFR level because specific items within the task type, depending on content or context or other features, may align at different levels. For instance, in the Question-Response item type, the exchange being tested may be an extremely common formula and align with CEFR level A1. Another exchange may be less common, require more contextual knowledge by the listener, and align with level A2.

Table 1. Alignment of Listening Prototype Tasks With Domain Definition

Domain definition	Description	Photo	Four Pics	Q/R	Cons	Talks
CEFR level	Corresponds to A1	✓	✓	✓		
	Corresponds to A2	✓	✓	✓	✓	✓
	Corresponds to B1			✓	✓	✓
Communication goals from domain definition	Understand simple descriptions of people, places, objects, and actions	✓	✓			
	Understand short dialogues or conversations on topics related to everyday life (e.g., making a purchase)			✓	✓	
	Understand short spoken monologues as they occur in everyday life (e.g., an announcement in a public area) when they are spoken slowly and clearly					✓
Linguistic knowledge and subcompetencies	Understand common vocabulary and formulaic phrases (lexical knowledge)	✓	✓	✓	✓	✓
	Understand simple sentences and structures (grammar knowledge)	✓	✓	✓	✓	✓
	Understand sentence-length speech and some common registers (discourse knowledge)	✓		✓	✓	✓
	Recognize and distinguish English phonemes and the use of common intonation and stress patterns and pauses to convey meaning in slow and carefully articulated speech across familiar varieties (phonological knowledge)	✓	✓	✓	✓	✓
	Infer implied meanings, speaker roles, or context in short, simple spoken texts (pragmatic competence)				✓	✓
	Understand the main idea and stated details in short, spoken texts (strategic competence)	✓	✓	✓	✓	✓

Note. Photo = Photographs; Four Pics = Four Pictures (Reading); Q/R = Question-Response; Cons = Conversations.

Reading task modeling. Eight different task types were modeled for the redesigned TOEIC Bridge Reading test. Of these, four were chosen for piloting based on considerations of construct coverage and practical issues of timing.

1. Four Pictures (Reading). The test taker reads a phrase or short sentence, then selects from four graphic options the one that best represents the content of the stimulus.
2. Sentence Completion. The test taker completes a cloze item based on a single-sentence assessing vocabulary and relatively simple grammatical structures.
3. Text Completion. The test taker completes a series of cloze items in the context of a multisentence-length paragraph. Sets include items that test vocabulary, appropriate word forms, and discourse knowledge (by selecting a sentence to be inserted into the paragraph).
4. Reading Comprehension. The test taker reads a 30- to 140-word stimulus and shows comprehension by answering two to three multiple-choice questions. The stimulus may be based on a range of genres, including websites and text message chains.

Table 2 shows how these prototype task models aligned with the CEFR levels the redesigned TOEIC Bridge assessments were intended to measure (A1–B1), the communication goals outlined in the reading domain definition, and the reading enabling skills or linguistic subskills. All the reading prototype tasks are shown as aligning with more than one CEFR level because they could be adapted to different levels of reading ability. For instance, the Sentence Completion task can be used to assess vocabulary knowledge. If the word tested is very common, then the task may be successfully completed by readers at the A1 level; if the word tested is less common, the task may align with A2 readers' skills.

Table 2. Alignment of Reading Prototype Tasks With Domain Definition

Domain definition	Description	Four Pics	Sent. Comp.	Text Comp.	Reading Comp.
CEFR level	Corresponds to A1	✓	✓		
	Corresponds to A2	✓	✓	✓	✓
	Corresponds to B1		✓	✓	✓
Communication goals	Understand nonlinear written texts (e.g., signs, schedules)				✓
	Understand written instructions and directions		✓	✓	✓
	Understand short, simple correspondence		✓	✓	✓
	Understand short informational, descriptive, and expository written texts about people, places, objects, and actions		✓	✓	✓
Linguistic knowledge and subcompetencies	Understand common vocabulary (lexical knowledge)	✓	✓	✓	✓
	Understand simple sentences and structures (grammatical knowledge)		✓	✓	✓
	Understand the organization of short written texts in a variety of formats (discourse knowledge)			✓	✓
	Recognize simple mechanical conventions of written English (orthographic knowledge)	✓	✓	✓	✓
	Infer implied meanings, including context or writer's purpose, in short, simple written texts (pragmatic competence)				✓
	Understand the main idea and stated details in short, written texts; infer the meaning of unknown written words through context clues (strategic competence)				✓

Note. Four Pics = Four Pictures (reading); Sent. Comp. = Sentence Completion; Text Comp. = Text Completion; Reading Comp. = Reading Comprehension

Speaking Task Modeling.

Approximately 12 tasks were modeled, of which 10 were selected for piloting the TOEIC Bridge Speaking assessment. The prototype items ranged from a direct measure of an enabling skill—pronunciation—to several different task models intended to capture evidence of the ability to complete communication goals at varying levels of complexity. Because speaking task response time is much shorter than writing response time, more speaking prototype tasks and their variants can be piloted within practical administration time than writing tasks. The following items were administered in the speaking pilot:

1. **Read Aloud.** The test taker is given a short text to read aloud. The task assesses the subcompetencies of pronunciation and intonation.
2. **Describe a Picture.** The test taker is instructed to describe a photograph. The task assesses description of people, places, and objects.
3. **Tell a Story.** The test taker is presented with a series of four pictures that graphically convey a simple narrative. The test taker is instructed to tell the story out loud. The task assesses narration.
4. **Respond to Questions.** The test taker is asked two related questions about personal experiences. (What time do you get up? What do you eat for breakfast?) The task assesses asking for and providing straightforward information.
5. **Respond to Questions With Information Provided.** The test taker is given a brief text, such as an advertisement or schedule, with information in telegraphic form. The test taker then responds to three specific questions that can be answered with information from the text.
6. **Give Two Reasons.** The test taker is asked for a preference on a relatively concrete and immediate topic and to give two reasons for the preference. The task is intended to assess giving and supporting an opinion.
7. **Express an Opinion.** The test taker is presented with a prompt on a relatively abstract topic that requires the construction of an argument for support.
8. **Listen-Speak.** The test taker listens to a short (40–60 word) informative stimulus—for example, an announcement. The test taker is then required to tell a third person

- the important information in the stimulus. The task is designed to assess the communication goals of giving directions and narrating.
9. **Ask/Invite/Request.** The test taker is given a short text, such as a ticket to a sporting event or a receipt from a purchase, and is instructed to role play a specific communication goal, such as inviting someone to do something or asking for help with a problem. The task is intended to assess making simple requests, offers, or suggestions.
 10. **Suggest a Solution.** The test taker is given an audio stimulus of a telephone message in which the caller presents a problem and asks the test taker to respond with a solution. The task is intended to assess the communication goal of offering a suggestion.

Table 3 shows how the speaking prototype task models aligned with the CEFR levels the redesigned TOEIC Bridge assessments were intended to measure (A1–B1), the communication goals outlined in the speaking domain definition, and the relevant enabling skills or linguistic subskills. The Read Aloud task type did not align with any of the communication goals in the domain definition. It was intended to provide relevant information about a subskill, pronunciation and intonation, especially for test takers at the A1 level with very limited ability to communicate through speaking. Several of the communication goals in the table are aligned with multiple prototype tasks. At this point in the development process, it was not clear which task type would be most useful in an operational assessment, and the test designers expected to use data from pilot testing to make further decisions about the prototype tasks.

Table 3. Alignment of Speaking Prototype Tasks With Domain Definition

Domain definition	Description	RA	Desc Pic	Story	Resp	Resp With Info	Give 2	Op	L-S	Ask/Inv/Req	Sol
CEFR level	Corresponds to A1	✓	✓								
	Corresponds to A2				✓	✓	✓		✓	✓	✓
	Corresponds to B1			✓			✓	✓			
Communication goals	Ask for and provide basic information				✓	✓			✓	✓	
	Describe people, objects, places, activities		✓	✓							
	Narrate and sequence simple events			✓							
	Give simple directions								✓		
	Make simple requests, offers, and suggestions									✓	✓
	Express an opinion or plan and give a reason for it						✓	✓			
Linguistic knowledge and subcompetencies	Use high-frequency vocabulary		✓	✓	✓	✓	✓	✓	✓	✓	✓
	Use common grammatical structures		✓	✓	✓	✓	✓	✓	✓	✓	✓
	Use simple transitions to connect ideas			✓			✓	✓	✓	✓	✓
	Pronounce words in a way that is intelligible to native speakers and proficient nonnative speakers of English	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Produce speech that is appropriate to the communication goal		✓	✓	✓	✓	✓	✓	✓	✓	✓

Note. RA = Read Aloud; Desc Pic = Describe a Picture; Story = Tell a Story; Resp = Respond to Questions; Resp With Info = Respond to Questions With Information Provided; Give 2 = Give Two Reasons; Op = Opinion; L-S = Listen-Speak; Ask/Inv/Req = Ask/Invite/Request; Sol = Propose a Solution.

Writing task modeling. The test designers created drafts of a large number of writing task models and variants of the models, of which six were selected for piloting. For some task types, the pilot forms included some variations of directions and response times. The following six task types were included in the pilot:

1. **Scrambled Sentence.** The test taker is presented with four to six segments of a sentence in random order and must put them in appropriate order. This task was designed to assess the linguistic subskill of using common grammatical structures.
2. **Write a Sentence Based on a Picture.** The test taker is presented with a picture and two words and must write a sentence using both words that is consistent with the picture.
3. **Respond to an E-mail.** The test taker reads an e-mail message and then writes a response. The prompt explicitly gives the test taker two functions to be included in the response (i.e., “ask two questions,” or “give two pieces of information”).
4. **Respond to an Instant Message.** The test taker reads an instant message that requires a short narration in past, present, or future tenses in response.
5. **Blog Post.** The test taker is instructed to write a short narrative about a specific personal topic (for example, describe a time when you helped a friend).
6. **Opinion.** The test taker is instructed to write a 100–150-word text giving an opinion with support on a specific topic.

Table 4 shows how the writing prototype task models aligned with the CEFR levels the redesigned TOEIC Bridge assessments were intended to measure (A1–B1), the communication goals outlined in the writing domain definition, and the relevant enabling skills or linguistic subskills. One task type, Scrambled Sentences, did not align with any of the communication goals in the domain definition but only with a linguistic subskill. It was intended to provide relevant information about test takers who may be at a very basic level of writing development. Several of the communication goals in the table are aligned with multiple prototype tasks. At this point in the development process, it was not clear which task type would be most useful in an operational assessment, and the test designers expected to use data from pilot testing to make further decisions about the prototype tasks.

Table 4. Alignment of Writing Prototype Tasks With Domain Definition

Domain definition	Description	Scramb. Sent	Pic-Sent	E-mail	IM	Blog Post	Op
CEFR level	Corresponds to A1	✓	✓				
	Corresponds to A2		✓	✓	✓		
	Corresponds to B1			✓		✓	✓
Communication goals	Ask for and provide basic information			✓			
	Describe people, places, objects, activities		✓		✓	✓	
	Narrate and sequence simple events				✓	✓	
	Make simple requests, offers, and suggestions			✓			
	Express a simple opinion and give a reason for it			✓			✓
Linguistic knowledge and subcompetencies	Use high-frequency vocabulary appropriate to a task (lexical knowledge and use)		✓	✓	✓	✓	✓
	Write a sentence using simple word order, such as SVO, interrogatives, imperatives; use common grammatical structures to contribute to meaning (grammatical knowledge and use)	✓	✓	✓	✓	✓	✓
	Arrange ideas using appropriate connectors (e.g., <i>for example, in addition, finally</i>); sequence ideas to facilitate understanding (discourse knowledge and use)		✓	✓	✓	✓	✓
	Use mechanical conventions of English (spelling, punctuation, and capitalization) to facilitate comprehensibility of text (orthographic knowledge and use)		✓	✓	✓	✓	✓
	Produce text that is appropriate to the communication goal (pragmatic competence)		✓	✓	✓	✓	✓

Note. Scramb Sent = Scrambled Sentence; Pic-Sent = Write a Sentence Based on a Picture; E-mail = Respond to an E-mail; IM = Instant Messaging; Blog Post = Blog Post; Op = Write an Opinion.

Item Prototyping Usability Study

After the prototype pilot tasks had been selected, a small-scale usability study was undertaken. The usability study focused on speaking and writing tasks because they required navigating a computer interface and included task types that were likely to be unfamiliar to the testing population. The aims of the study were threefold: (a) to understand how beginning to intermediate learners of English would react to the new item types; (b) to evaluate the clarity of item directions; and (c) to understand potential challenges that learners of English, particularly beginning to intermediate learners, may have in navigating a computer-based assessment. By conducting a small-scale usability study, test designers were able to refine the new item types for speaking and writing administration to a larger number of pilot participants in the fall.

The usability study was conducted at the Educational Testing Service (ETS) headquarters in New Jersey in August 2017. Four English language learners of beginning to intermediate proficiency each completed approximately 20 speaking and writing tasks. Immediately after completing the tasks, the participants were interviewed. For all of the structured interviews, an interpreter was available to provide real-time translation for the participant and interviewer.

The feedback to the new item types was generally positive. Participants reported that they understood item directions. However, some issues were raised regarding navigating from screen to screen and the difficulty of some of the audio stimulus components of two item types, Listen and Retell and Respond to Questions with Information Provided.

Based on this feedback, variants of these two speaking tasks were included in the pilot forms: One version included a transcript of the audio on screen so that participants could read along with the audio, and the other version did not.

Pilot Testing

In September 2017, pilot tests were administered to 464 participants from Brazil ($n = 84$), Japan ($n = 257$), Korea ($n = 57$), and Taiwan ($n = 66$). The pilot administration included the assessment of all four skills (i.e., listening, reading, speaking, and writing). For the skills of listening and reading, one pilot form was created consisting of 50 multiple-choice listening items followed by 50 multiple-choice reading items. For the skills of speaking and writing, two pilot forms were created. One pilot form consisted of 10 speaking tasks followed by nine writing items, and another pilot form consisted of nine speaking tasks followed by seven writing items. The pilot forms were intended not to be draft operational forms but rather to produce information

about the performance of tasks and to be as similar as possible to one another in overall administration time.

The listening and reading pilot form was paper based, and the speaking and writing pilot forms were computer based. All 464 participants took the listening and reading pilot form, and 436 participants were randomly assigned to one of the two of the speaking and writing pilot forms.

Results of Pilot Testing

The goals of pilot testing for listening and reading and for speaking and writing were somewhat different. Because there were existing *TOEIC Bridge* Listening and Reading assessments, the *TOEIC* program and the test designers were concerned that the overall difficulty of redesigned *TOEIC Bridge* Listening and Reading tests be approximately the same as the existing assessments. That is, they did not want the new assessments to be so difficult that they would be discouraging for the current test-taking population nor so easy as to not give that population meaningful information. To that end, the listening and reading pilot forms included a subset of items from the existing assessments so that meaningful comparisons could be made between the old item types and the new. Because there were no existing *TOEIC Bridge* Speaking and Writing assessments, the comparative difficulty of the tasks was not a concern, and the purpose of the pilot was to collect information on whether or not beginning to intermediate learners understood the tasks and produced responses that could be meaningfully and reliably scored by human raters and whether the draft rubrics were as useful as possible.

After finishing the speaking and writing pilot test, test takers in Japan ($n = 30$) and Brazil ($n = 5$) completed surveys administered in their local language that asked them to provide feedback on the usability of the test (e.g., clarity of directions, adequacy of preparation and response time), various perceptions of the test (e.g., authenticity, difficulty), and task-specific questions (e.g., usefulness of a visual stimulus). Although a majority of participants indicated that the English directions were not difficult to understand for all of the tasks, a relatively high proportion (>40%) indicated that the directions for several of the speaking test tasks were difficult to understand (Ask, Invite, Request; Respond to Questions With Information Provided). In almost all instances, a majority of participants believed the preparation and response time provided for pilot test tasks was “OK,” although a larger proportion (>40%) believed preparation and response times were too short for several speaking and writing tasks (Ask, Invite, Request;

Respond to Questions With Information Provided; Give a Reason). Participants' perceptions of task difficulty and authenticity largely aligned with expectations, as tasks designed to target higher levels of proficiency were viewed as more difficult. Finally, task-specific questions helped identify features that could be refined for the next phase of testing.

Pilot results for listening. Overall, the results of the pilot for the listening prototype tasks were positive. All of the prototype items performed within the range of acceptable reliability and difficulty. The following items were of particular note:

- The Four-Picture items were comparable in difficulty to the easiest items on the existing TOEIC Bridge Listening measure and differentiated among students at the beginning level.
- The Question-Response items were piloted in two versions. In one, similar to an existing TOEIC Bridge Listening item type, the options were audio only; in the other, the options were presented as audio and as text in the test book. The two versions of the task performed similarly, and it was decided to include the text and audio version in field study.
- The pilot Short Conversation two-item sets were not obviously more difficult than Short Conversation items from the existing TOEIC Bridge Listening test with one item per stimulus.
- The pilot Short Talk (monologue) two-item sets were not obviously more difficult than Short Talks from the existing TOEIC Bridge Listening test with one item per stimulus.

Based on the results of the pilot and in consultation with the program's psychometricians, Table 5 shows the form blueprint that was developed for the listening test.

Table 5. Redesigned TOEIC Bridge Listening Field Test Blueprint

Test part	Question type
Part 1	Four Pictures 6 questions
Part 2	Question-Response 20 questions
Part 3	Conversations 5 sets, 2 items per set, 10 items
Part 4	Talks 6 sets, 2-3 items per set, 14 items

Pilot results for reading. As with listening, the analysis of item difficulty and discrimination in the reading pilot was encouraging. The most relevant findings included the following:

- The Four-Picture reading items were of similar difficulty to the Four-Picture listening items.
- The pilot Sentence Completion items covered a wide range of difficulty.
- The Text Completion items were more difficult on average than the Sentence Completion items.
- The new genres of stimuli for reading sets—text messages, FAQs, website material with reader comments—all performed well, were in a similar range of difficulty as the legacy TOEIC Bridge Reading sets, and represented a more up-to-date range of real-world texts.

Based on the results of the pilot and in consultation with the relevant psychometricians, Table 6 shows the form blueprint that was developed for reading.

Table 6. Redesigned TOEIC Bridge Reading Field Test Blueprint

Test part	Question type
Part 1	Sentence Completion 15 items
Part 2	Text Completion 5 sets, 3 items per set, 15 items
Part 3	Reading Comprehension 8 sets, 2-3 items per set, 20 items

Scoring the Constructed Responses From Pilot Testing

For the speaking and writing task types, rubric refinement was an integral component of test design. The test design team developed rubrics for evaluating spoken and written responses in tandem with designing the new task types. Each piloted task type was accompanied by its own holistic rubric, with 0 to 3 or 0 to 4 score points per rubric. Once the constructed responses from the pilot tests were available, the test design team analyzed the spoken and written responses to hone the rubrics that had been created during the task design phase.

The rubrics were further refined during the rangefinding process by a group of senior-level test developers. The three or more members of the rangefinding team each applied the rubrics to score between 5% and 10% of responses for each piloted item. The team members independently assigned scores to pilot test responses. Disagreements about scores were resolved by discussing team members' rationales for scoring until consensus among the team members was reached. Based on these discussions, the rangefinding team revised rubrics for clarity so as to minimize potential confusion that might lead to low agreement among raters. Samples for rater training were selected through the rangefinding process, and these trainings samples were annotated with scoring rationales.

Two main considerations guided the revision of rubrics during pilot scoring: to make the rubrics as responsive as possible to actual observed differences in performance among the pilot population and to make the rubrics easy to use for raters. The team decided that each task type required its own rubric with different specific details of task completion included. However, because individual operational raters would almost certainly be rating different task types in one scoring session, the team felt that the rubrics would be easiest to use if they were as parallel in construction and wording as possible. To that end, the team decided that each rubric should be structured with a general and repeated statement of task completion for each score point and the details of task-specific completion in bullets below the heading. Figures 6 and 7 show how this was put in practice for a speaking task (Describe a Picture) and a writing task (Blog Post).

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Describe a Picture

Score	Response description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> • The response describes the important parts of the picture. • Delivery is generally intelligible but may require some listener effort. • Choice of vocabulary and use of structures are appropriate, though minor errors that do not affect meaning may be present.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> • The response is connected to the picture, but the meaning is obscured in places. • Delivery is sometimes unintelligible and requires listener effort. • Choice of vocabulary and use of structures are limited, and errors interfere with comprehensibility.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> • The response is not connected to the picture. • Delivery is mostly unintelligible. • Severely limited choice of vocabulary and/or use of structures obscure meaning. • The response may consist of isolated words or phrases.
0	No response OR no English in the response.

Figure 6. Sample TOEIC Bridge Speaking rubric.

Redesigned TOEIC Bridge Writing Test

Scoring Guide for Blog Post

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response fully addresses the topic and task. ▪ Overall meaning is clear, though minor grammatical errors that do not obscure meaning may be present. ▪ The story is told in a logical sequence, and any connectors are used appropriately. ▪ The choice of vocabulary is appropriate to the topic of the prompt.
2	<p>The response is partially effective at addressing the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response partially addresses the topic or partially completes the task. ▪ Use of language structures contributes to meaning, though grammatical errors may occasionally obscure meaning. ▪ The logical sequence of the story is mostly clear. ▪ The choice of vocabulary is sometimes limited or inappropriate to the topic.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is unsuccessful at addressing the topic or task, though it may contain some related words. ▪ The response is off topic or seriously underdeveloped. ▪ Frequent and serious grammatical errors interfere with the comprehensibility of most of the response. ▪ The choice of vocabulary is limited (use of isolated words), inaccurate, and/or it relies on repetition of the prompt in most of the response.
0	No response OR no English in the response.

Figure 7. Sample TOEIC Bridge Writing rubric.

After rangefinding and rubric refinement, a larger group of 17 test developers with extensive experience in scoring constructed responses for assessments of English language proficiency was trained using the samples selected during rangefinding. All pilot test responses were scored by at least two raters. Exact agreement rates by item ranged between 64% and 93%, with most items (58%) ranging between 70% and 79% exact agreement. When scores were discrepant—that is, when raters disagreed by more than one score point—a member of the design team reviewed the response and provided the final, adjudicated score. Fewer than 1% of pilot test responses caused discrepant scores.

Pilot results for speaking. After scoring the pilot speaking responses, test developers evaluated the effectiveness of each of the pilot task types. Three task types were considered for operational use with very few modifications:

- Two different versions of the Read a Short Text Aloud task type were piloted, one with a 20- to 25-word stimulus and one with a 30- to 45-word stimulus. Pilot results indicated that the longer stimulus provided more differentiation among responses.
- Directions for the Describe a Picture item type were elaborated to give test takers more explicit help in creating a response. Also, detailed specifications for the picture content were created to ensure that the task was approachable for beginning to intermediate learners.
- Directions for Tell a Story were elaborated to give test takers more support in creating full responses.

Three task types were revised post pilot:

- The Listen and Retell task type was revised to be less challenging than the pilot version by reducing the number of important points the test taker should include in the response.
- The Answer Questions With Information Provided item type was revised to allow for question formation as well as suggestions.
- Aspects of the Opinion and Give Two Reasons task types were combined in a post-pilot version of the task intended to elicit opinion. In this task, the test taker is given a table or list with short, telegraphic information about two options, alternatives, or choices that can be compared and contrasted and uses this information to express an opinion, make a recommendation, give advice, or state a preference and support it.

In addition to the task modifications described above, the test developers gave the speaking task types the names that would be used publicly when describing the test contents. The goal was to give names to items that were connected to the communication goal being assessed and to be as transparent as possible. Table 7 presents the form blueprint that was created for the speaking field test forms.

Table 7. Redesigned TOEIC Bridge Speaking Field Test Blueprint

Question type	Description of questions	Timing
Read a Short Text Aloud (two questions)	The test taker reads a short paragraph aloud.	Prepare 25 sec. Respond 30 sec.
Describe a Photograph (two questions)	The test taker describes a photograph presented on the computer screen.	Prepare 30 sec. Respond 30 sec.
Listen and Retell	The test taker listens to a talk on an everyday social or workplace-related situation and communicates the main ideas and important details of the talk to a third person.	Prepare 10 sec. Respond 30 sec.
Short Interaction	The test taker reads a brief text and asks for information and/ or makes requests, offers, suggestions, and invitations.	Prepare 20 sec. Respond 30 sec.
Tell a Story	The test taker narrates a story based on a picture sequence.	Prepare 45 sec. Respond 60 sec.
Make and Support a Recommendation	The test taker reads a short, simple piece of information showing options, alternatives, or choices that can be compared and contrasted and uses this information to express an opinion, make a recommendation, give advice, or state a preference and support it.	Prepare 45 sec. Respond 60 sec.

Note. Total test time, with directions and audio stimuli, 12 min. 45 sec.

Pilot results for writing. For both speaking and writing, the test designers were also members of the rating team, and thus for each item type, they had rated many, if not all, the responses. Based on their own experience rating and on the group’s ratings for each pilot task type item, designers evaluated the pilot tasks. In general, they focused on these questions:

- Did a significant number of pilot participants misunderstand the task and produce responses unrelated to the targeted communication goal?
- Were a reasonable number of responses successful according to the a priori criteria represented in the draft rubric (i.e., was the task too difficult for most of the pilot population)?
- Did the responses allow differentiation between pilot participants? Did the rubric produce a range of ratings?

- Did raters find the rubrics easy to use, and could they score efficiently?
- Did two or more tasks focus on the same communication goal? Did one produce better evidence for the relevant claim(s)?

Test designers made no changes to two of the pilot writing tasks: Write a Sentence Based on a Picture and Blog Post.

After evaluating the pilot responses, test designers made the following changes to task designs:

- In the pilot version of the Scrambled Sentence task, test takers were required to type out the unscrambled sentence. The pilot responses made clear that many participants' limited keyboarding skills were interfering with their ability to respond. The task was redesigned so that test takers can drag and drop the sentence elements into the appropriate order.
- Elements of different pilot tasks were combined to create a task that assessed the communication goals of providing basic information, asking for basic information, and making simple requests, offers, suggestions, or invitations. A short message (20–45 words) prompts the response. The stimulus includes two questions that require two or three details in response. The stimulus is preceded by a lead-in enumerating the communication goals test takers should produce (i.e., ask one question and make one suggestion in response to the message).
- Features of the E-mail pilot item and the Opinion pilot item were combined in the Respond to an Extended Message item type. The stimulus is presented as an e-mail inquiry and asks two separate questions that are thematically related. The first question requires straightforward information as a response. The second question requires a brief opinion.

In addition to the task modifications described above, test developers also gave the task types the names that would be used publicly in referring to the tests. The goal was to use task names that were as transparent as possible. After analysis of the pilot responses, a form blueprint for the writing field test was developed (see Table 8).

Table 8. Redesigned TOEIC Bridge Writing Field Test Blueprint

Question type	Description of task	Timing
Build a Sentence (three questions)	Arrange a set of words or phrases in the appropriate order to form a grammatically correct sentence or question.	Respond 3 min. for three questions
Write a Sentence (three questions)	The test taker sees a photograph with two key words or phrases below it. Using both of the key words or phrases, the test taker writes one grammatically correct sentence that describes the picture. The test taker can change the forms of the words and can use them in any order.	Respond 3 min. for three questions
Respond to a Brief Message	The test taker reads a brief message (e.g., an instant message) from an acquaintance and composes a response that completes two communication goals (e.g., providing basic information; asking for basic information; or making simple requests, offers, suggestions, and invitations).	Respond 8 min.
Write a Narrative	The test taker reads a prompt on the screen that specifies a category of past experience and writes a narrative based the prompt.	Respond .8 min.
Respond to an Extended Message	The test taker reads an extended message (e.g., an e-mail) from a person or entity, and writes a response.	Respond.10 min.

Note. Total test time, with directions: 33 minutes.

Field Study

After task types were selected for use in the field test, task-level specifications could be finalized. Written specifications for each of the tasks had been evolving since the initial modeling phase before piloting. In preparation for the field test, task-level specifications were needed to generate relatively large (more than 100 items) usable pools of listening and reading items from which forms could be constructed. Fewer items were needed for the speaking and writing field tests than for the listening and reading field tests. However, because speaking and writing forms are not equated, it was important that the items representing task types across forms be as similar in construction as possible. The most effective way to create similar items is to create and use very explicit task-level specifications. Task-level specifications, in the form of item shells, included the claim about the test taker for which the task provides evidence, a description of the task, the scoring criteria, the fixed elements of the task (those aspects of the task that are the same in every instance), and the variable elements (the things that change to make two items representing the same task different from one another). Figure 8 presents an item shell for the redesigned TOEIC Bridge Write a Sentence task type.

Claims	Task	Fixed elements	Variable elements	List of variants
<p>The test taker can write sentences to describe people, objects, places, and activities.</p>	<p>Test takers will see a photograph with two key words or phrases below it. Using both of the key words or phrases, test takers will write one grammatically correct sentence that describes the picture. The form of the key words or phrases can be manipulated.</p> <p>Scoring:</p> <ul style="list-style-type: none"> ▪ 0–3 holistic scale ▪ Criteria: use of key words, consistency with picture, correctness of grammar. 	<p>Directions: Write ONE sentence based on the picture. Use the TWO words or phrases under the picture. You can change the forms of the words, and you can use them in any order. You have 90 seconds to write.</p> <p>Stimuli:</p> <p>1. One color photograph. The picture should not rely on the text and must have a clear focus. A description of the photograph should not require specialized vocabulary.</p> <p>2. Two key words or phrases. Key words or phrases are located below the photograph, and key words require no or limited form transformation and are presented in canonical word order (SVO).</p> <p>3. Response:</p> <ul style="list-style-type: none"> ▪ Response length: a sentence ▪ Response time: 90 seconds per item 	<p>1. Photograph contexts</p> <p>2. Key words/phrases</p> <p>3. Actions</p>	<p>1. Examples of contexts: activities, dining out, entertainment, family and friends, business, health, housing, offices, news, school, shopping, travel</p> <p>2. Parts of speech: adjective adverb coordinating conjunction noun preposition verb</p> <p>3. Examples of actions: Activities (participating in hobbies, playing sports) Dining (eating, drinking, ordering) Entertainment (playing music, visiting museums) Health (visiting doctor, attending a class) Household tasks (cleaning, repairing, moving, cooking) Shopping (buying groceries, selecting clothes) Travel (taking trains, waiting in airports, buying tickets, checking schedules, looking at maps or documents)</p>

Figure 8. Redesigned TOEIC Bridge Write a Sentence item shell.

Field Test Usability Study

Prior to the field test, researchers again conducted cognitive interviews to evaluate the usability of the refined computer-based speaking and writing field test forms for low-proficiency adult learners of English ($n = 9$). Again, the study identified general usability issues (e.g., some participants began speaking before recording started, timing directions for writing tasks were sometimes misunderstood) and item-specific concerns (e.g., some participants did not notice the two words under the picture for the Picture Sentence task in the writing test) and provided suggestions for remediating these issues and concerns. This phase of development corresponds to the presentation model identified in the ECD framework.

Field test results—listening and reading. In May–June of 2017, two parallel forms of the redesigned TOEIC Bridge Listening and Reading assessments were administered to a total of 2,484 test taker in six countries (Japan, Korea, Taiwan, Brazil, Mexico, and Colombia). Each participant took either Form 1 ($N = 1,220$) or Form 2 ($N = 1,264$). The results of the field test confirmed the listening and reading test designs. Item types and individual items were within the expected range of difficulty for the TOEIC Bridge population, and the assessments reliably distinguished four levels of performance for both listening and reading, which served as the basis for the development of level descriptors (see below). Further detailed discussion of the field test results for the listening and reading assessments are available in Lin (2019).

Field test results—speaking and writing. The May–June field test also included administration of two parallel computer-based forms of speaking and writing assessments. The same six countries participated (Form 1, $N = 1,228$; Form 2, $N = 1,174$). Assessment developers and experienced TOEIC Speaking and Writing raters used the rubrics finalized as part of the post-pilot item specification process to identify benchmark responses, training responses, and calibration sets for each of the speaking and writing task types. TOEIC raters, who are expected to be the operational redesigned TOEIC Bridge raters, were trained and scored the bulk of the field test responses. All field test responses were double scored. A detailed discussion of the speaking and writing field test results is available in Lin (2019).

Rater Survey

Raters who scored the speaking field test and writing field test were invited to participate in an online survey. Raters of the speaking test ($n = 156$) and writing test ($n = 41$) who responded were asked to indicate (using 5-point Likert-type scales) the extent to which they agreed or

disagreed with the statements “It was easy to form judgments” and “I felt confident in my scores” for each of the scoring rubrics they used. Overall, a high percentage of raters agreed (i.e., agree or strongly agree) with the statements for each of the scoring rubrics. The percentage of raters who agreed with these statements ranged from 64% to 87% across scoring rubrics for the speaking test and 71% to 97% for the writing test. Raters were also asked to estimate the approximate percentage of test takers who (a) did not seem to have adequate time to provide responses and (b) did not seem to understand task directions. A large percentage of raters (> 75%) indicated that most test takers had adequate time to provide a response and seemed to understand directions, with several exceptions. Results of the survey suggested that at least 35% of raters believed that test takers could use more time to provide responses for the Tell a Story task on the speaking test and for the Blog and E-mail tasks on the writing test. At least 35% of raters also believed that test takers seemed to have some confusion about the directions for the Give Reasons task on the speaking test and for the Blog task on the writing test.

Test taker survey—speaking and writing. Test takers who completed the speaking and writing field test were invited to complete a follow-up survey in their local language, and responses were obtained from participants in Brazil ($n = 268$), Colombia ($n = 18$), Japan ($n = 1251$), Korea ($n = 323$), Mexico ($n = 48$), and Taiwan ($n = 333$). In the surveys, participants again provided feedback on the usability of the test (e.g., clarity of directions, adequacy of preparation and response time) and their various perceptions of the test (e.g., authenticity, difficulty) and answered task-specific questions (e.g., usefulness of a visual stimulus). A majority of participants indicated that the English directions were not difficult to understand for most tasks (ranging from 62% to 87% across tasks and forms for speaking and 74% to 87% for writing) with the exception of the Short Interaction task on the speaking test (48%). For most tasks, a majority of participants indicated that the preparation and/or response times provided were good (ranging from 61% to 74%) with the exception of several speaking test tasks where at least 30% of participants believed preparation times were insufficient (Short Interaction, Tell a Story, Listen and Retell) and/or response times were insufficient (Tell a Story). Participants’ perceptions of task difficulty and authenticity were largely aligned with expectations, as tasks designed to target higher levels of proficiency were viewed as more difficult. Finally, participant responses to task-specific questions indicated that features of the field test tasks were largely functioning as intended.

Final Adjustment to Item Presentations

Based on feedback from the field test usability study, the rater survey, the survey of speaking and writing test takers, and the field test results, final adjustments were made to speaking and writing task types.

The order of the speaking items was changed to reflect increasing difficulty of tasks as observed in the field test results (see Figure 9).

Field test		Operational test	
1	Read a Short Text Aloud	1	Read a Short Text Aloud
2	Read a Short Text Aloud	2	Read a Short Text Aloud
3	Describe a Photograph	3	Describe a Photograph
4	Describe a Photograph	4	Describe a Photograph
5	Short Interaction	5	Listen and Retell
6	Tell a Story	6	Short Interaction
7	Listen and Retell	7	Tell a Story
8	Make and Support a Recommendation	8	Make and Support a Recommendation

Figure 9. Redesigned TOEIC Bridge Speaking field test and operational test item order.

Adjustments were also made to the preparation and response times of several speaking and writing task types (see Tables 9 and 10). With these changes, the task design and test blueprint design processes were complete.

Table 9. Redesigned TOEIC Bridge Speaking Field Test and Operational Test Item Timing

Speaking	Field test prep (seconds)	Field test response (seconds)	Operational prep (seconds)	Operational response (seconds)
1 Read a Short Text Aloud	20	30	25	30
2 Read a Short Text Aloud	20	30	25	30
3 Describe a Photograph	30	30	30	30
4 Describe a Photograph	30	30	30	30
5 Listen and Retell	10	30	15	30
6 Short Interaction	20	30	30	30
7 Tell a Story	45	45	45	60
8 Make and Support a Recommendation	45	60	60	60

Note. Figures shaded in gray indicate a change from field test to operational test timing.

Table 10. Redesigned TOEIC Bridge Writing Field Test and Operational Test Item Timing

	Writing	Field test response (minutes)	Operational response (minutes)
1	Build a Sentence	1	1
2	Build a Sentence	1	1
3	Build a Sentence	1	1
4	Write a Sentence	1	1.5
5	Write a Sentence	1	1.5
6	Write a Sentence	1	1.5
7	Respond to a Brief Message	8	8
8	Write a Narrative	8	10
9	Respond to an Extended Message	10	10

Note. Figures shaded in gray indicate a change from field test to operational test timing.

Proficiency Descriptors

The last assessment design task for test developers in test design prior to moving to support the operational test was the creation of proficiency descriptors for each of the domains of listening, reading, speaking, and writing. One of the business requirements for the redesigned TOEIC Bridge tests was to provide meaningful feedback to teachers and learners in the form of proficiency descriptors. With this end in mind, the test development team created descriptions of what test takers can do using English. The team used the following to inform the construction of the proficiency descriptors: results of the initial ECD test design process, recommendations from psychometric analysis of the field test, findings from a mapping study, results from a survey of field test participants, and for speaking and writing tasks, review of the field test responses.

The test development team began this phase of the project by revisiting the findings from the initial domain analysis and modeling and also reviewing the domain definitions and the TLU. The purpose of returning to the initial domain analysis and modeling was to ensure the resulting descriptors would be aligned with the TLU/domain definition, in accordance with ECD. Next, the test development team revisited the results from the task modeling phase of the project, reviewing the task specifications, the claims for each task type derived from the domain model, and the rubrics for the speaking and writing tasks, again in keeping with ECD to ensure alignment with the domain model.

Psychometric analysis of field test scores indicated test takers could be grouped into four distinct score ranges for each skill assessed (Lin et al., 2019). For each of these score ranges, the most common patterns in field test participants' performances were identified and examined by the test development team. For listening and reading, the average percent of items answered

correctly by task type was used to identify patterns, and for speaking and writing, the average item score by task type was used. To draft the descriptors, the test development team linked these patterns back to the task claims from the domain model, and for speaking and writing, the patterns were linked back to the rubrics.

To validate the proficiency descriptors, the test development team compared the drafts to the results of a study mapping the TOEIC Bridge field test scores onto international standards of language proficiency (Schmidgall et al., 2019). The test development team also compared the proficiency descriptors to the results of a can-do survey conducted with the field test participants (Schmidgall et al., 2019). For speaking and writing, responses from the field test were reviewed and compared to the drafted proficiency descriptors. Finally, the validated drafts were reviewed by subject matter experts, researchers, product managers, marketing, and ETS partners prior to the finalization of the proficiency descriptors.

Conclusion

This paper described the process of developing the task types, presentation models, and assembly models of the four parts of the redesigned TOEIC Bridge tests: listening, reading, speaking, and writing. Design for each part began with consideration of the business requirements for the assessment program and well-defined domain definitions. Task models, or prototype tasks, were developed, tried out in cognitive labs, and piloted. The results of the pilot informed modifications of the prototype tasks and tentative selection of task types for the field test. Further cognitive labs preceded the field test. Field test data, supplemented by surveys of the raters of constructed response tasks, were used to set the final specifications for all task types and the operational assembly models.

ECD is often presented as a systematic approach to test development that emphasizes how a test may be used to elicit evidence of the ability to be assessed from test-taker performance. Another benefit of such a systematic approach is the collection of documentation throughout the test development process to justify design decisions by test developers. With this in mind, we described the various sources of data we obtained throughout the test development process (e.g., cognitive labs, surveys, item performance) and how each influenced item and test design decisions.

References

- Lin, P., Cid, J., & Zhang, J. (2019). *Field study statistical analysis for the redesigned TOEIC Bridge® tests* (Research Memorandum No. RM-19-09). Princeton, NJ: Educational Testing Service.
- Mislevy, R. J., & Yin, C. (2012). Evidence-centered design in language testing. In G. Fulcher and F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 208–222). New York, NY: Routledge.
- Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). *Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests—Framework paper* (Research Report No. RR-19-30). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12267>

Appendix A. Redesigned TOEIC Bridge Speaking and Writing Tests Scoring Guides

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Read a Short Text Aloud

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The entire text is read aloud AND other-language influence does not affect overall intelligibility. ▪ At the word level, pronunciation is mostly intelligible, but there may be some minor lapses. ▪ At the phrase and sentence level, intonation and stress are mostly appropriate, though the response may include some lapses and/or some other language influence.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ A section of the text is NOT read aloud. ▪ At the word level, pronunciation is sometimes unintelligible and requires some listener effort. ▪ At the phrase and sentence level, intonation and stress are somewhat appropriate, but lapses and/or other language influence are present.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ Most of the text is NOT read aloud. ▪ The response is off topic. ▪ Speech is mostly unintelligible and requires significant listener effort to understand.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Describe a Photograph

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response describes the important parts of the picture. ▪ Delivery is generally intelligible but may require some listener effort. ▪ Choice of vocabulary and use of structures are appropriate, though minor errors that do not affect meaning may be present.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is connected to the picture, but the meaning is obscured in places. ▪ Delivery is sometimes unintelligible and requires listener effort. ▪ Choice of vocabulary and use of structures are limited, and errors interfere with comprehensibility.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is not connected to the picture. ▪ Delivery is mostly unintelligible. ▪ Severely limited choice of vocabulary and/or use of structures obscure meaning. ▪ The response may consist of isolated words or phrases.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Listen and Retell

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response adequately communicates two main ideas of the talk. ▪ Delivery is generally intelligible but may require some listener effort. ▪ The choice of vocabulary and use of structures fulfills the demands of the task.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response communicates at least one of the main ideas of the talk, but the response is incomplete or one of the main ideas is inaccurate. ▪ Delivery is sometimes unintelligible and/or sometimes requires listener effort. ▪ The choice of vocabulary and use of structures are limited and interfere with overall comprehensibility.
1	<p>The response is not effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response may consist of isolated words or phrases or is off topic. ▪ Delivery is mostly unintelligible and/or requires significant listener effort.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Short Interaction

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The required task (request, offer, suggestion, invitation) and details from the prompt are successfully communicated. ▪ Delivery is generally intelligible but may require some listener effort. ▪ Choice of vocabulary and use of structures fulfill the demands of the prompt. ▪ Minor errors do not obscure overall meaning.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The required task (request, offer, suggestion, invitation) is attempted and relevant details are included, but the response is incomplete. ▪ Delivery is sometimes unintelligible and requires listener effort. ▪ Choice of vocabulary and use of structures are limited and sometimes affect meaning.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The required task (request, offer, suggestion, invitation) is not attempted. ▪ The response is off topic. ▪ Delivery is mostly unintelligible. ▪ The choice of vocabulary and use of structures are severely limited (use of isolated words). Meaning is obscured.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Tell a Story

Score	Response Description
4	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The presentation of a cohesive story is based on the main features of the picture sequence. ▪ The delivery is generally intelligible and does not interfere with meaning. ▪ The choice of vocabulary and use of structures and connecting language fulfill the demands of the task.
3	<p>The response is mostly effective at addressing the prompt. The response consists of a mostly cohesive story based on the picture sequence, although part of the story may be incomplete or unclear because</p> <ul style="list-style-type: none"> ▪ delivery is occasionally unintelligible or requires listener effort, and/or ▪ choice of vocabulary and use of structures and connecting language occasionally interfere with overall comprehensibility.
2	<p>The response is partially effective at addressing the prompt. Parts of the picture sequence may be conveyed, but the story is mostly incomplete or unclear because</p> <ul style="list-style-type: none"> ▪ parts of the narrative sequence are missing, and/or ▪ unintelligible delivery interferes with parts of the narrative sequence, and/or ▪ choice of vocabulary and use of structures and connecting language are limited and interfere with overall comprehensibility.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response may be only tangentially related to the picture sequence. ▪ The response is off topic. ▪ Delivery is mostly unintelligible and/or requires significant listener effort throughout. ▪ Choice of vocabulary and use of structures are severely limited (use of isolated words) and may significantly interfere with comprehensibility.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Speaking Test

Scoring Guide for Make and Support a Recommendation

Score	Response Description
4	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ All of the information from the stimulus is clearly and appropriately communicated. ▪ A recommendation is made and is adequately supported. ▪ Delivery is intelligible but may require some listener effort. ▪ Choice of vocabulary and use of structures fulfill the demands of the task.
3	<p>The response is mostly effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ Most of the information provided in the stimulus is appropriately communicated. ▪ A recommendation is made, but support is limited. ▪ Delivery is mostly intelligible, though listener effort is required at times. ▪ Choice of vocabulary and use of structures are fairly effective, though they interfere with comprehensibility at times.
2	<p>The response is partially effective at addressing the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ Information from the stimulus is presented, but only limited original language is used. ▪ A recommendation may be made, but support is missing. ▪ Delivery is sometimes unintelligible and may require listener effort. ▪ Choice of vocabulary and use of structures are limited and often obscure meaning.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ Information from the stimulus is missing. ▪ The response is off topic. ▪ The response is limited to reading aloud the prompt, the directions, or the information in the stimulus without adding original language. ▪ Delivery may be mostly unintelligible and require listener effort. ▪ The response contains errors that obscure meaning most of the time.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Writing Test

Scoring Guide for Write a Sentence

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response is consistent with the picture. ▪ Forms of both key words are used appropriately in one sentence. ▪ No grammatical errors are present.
2	<p>The response is partially effective at addressing the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is consistent with the picture. ▪ Forms of both key words are present, though they may be in different sentences, or the form of the word(s) may not be accurate. ▪ Minor grammatical errors are present but do not obscure meaning.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is not consistent with the picture. ▪ The response omits one or both key words. ▪ Significant grammatical errors are present that obscure meaning.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Writing Test

Scoring Guide for Respond to a Brief Message

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response is clear and fully responsive to the stimulus message. ▪ The choice of vocabulary is effective, with allowance for slight inaccuracies that do not obscure meaning. ▪ The use of grammatical structures fulfills the demands of the task. A few minor errors may be present but do not obscure meaning.
2	<p>The response is partially effective at addressing the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response attempts both tasks, though one or both tasks may not be successful. ▪ The response is somewhat clear. ▪ Errors in use of grammar and choice of vocabulary appear throughout the response and may occasionally obscure meaning.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response attempts only one of the required tasks, or the response does not attempt any of the required tasks. ▪ The response contains very little original language and may contain words or phrases related to or copied from the stimulus. ▪ The response is mostly incoherent. ▪ Errors in grammar and usage frequently obscure meaning.
0	<p>No response OR no English in the response. There may be keystroke characters that convey no meaning.</p>

Redesigned TOEIC Bridge Writing Test

Scoring Guide for Write a Narrative

Score	Response Description
3	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The response fully addresses the topic and task. ▪ Overall meaning is clear, though minor grammatical errors that do not obscure meaning may be present. ▪ The story is told in a logical sequence, and any connectors are used appropriately. ▪ The choice of vocabulary is appropriate to the topic of the prompt.
2	<p>The response is partially effective at addressing the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response partially addresses the topic or partially completes the task. ▪ Use of language structures contributes to meaning, though grammatical errors may occasionally obscure meaning. ▪ The logical sequence of the story is mostly clear. ▪ The choice of vocabulary is sometimes limited or inappropriate to the topic.
1	<p>The response does not effectively address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The response is unsuccessful at addressing the topic or task, though it may contain some related words. ▪ The response is off topic or seriously underdeveloped. ▪ Frequent and serious grammatical errors interfere with the comprehensibility of most of the response. ▪ The choice of vocabulary is limited (use of isolated words), inaccurate, and/or it relies on repetition of the prompt in most of the response.
0	No response OR no English in the response.

Redesigned TOEIC Bridge Writing Test**Scoring Guide for Respond to an Extended Message**

Score	Response Description
4	<p>The response successfully addresses the prompt and is marked by all of the following.</p> <ul style="list-style-type: none"> ▪ The requested information, opinion, and support for that opinion are present and elaborated clearly. ▪ The response is well organized, well developed, and coherent. ▪ Tone and register are appropriate for the intended audience. ▪ Command of the language demonstrates appropriate use of structures, syntactic variety, and vocabulary, though there may be minor lexical or grammatical errors.
3	<p>The response is mostly effective at addressing the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ The requested information, opinion, and support for that opinion are present, though they may not be clear or relevant. ▪ Organization, development, and/or coherence are generally appropriate for the task. ▪ Tone and register are not fully appropriate for the intended audience. ▪ Command of the language demonstrates mostly appropriate use of structures, syntactic variety, and vocabulary, though some lexical and/or grammatical errors occasionally obscure meaning.
2	<p>The response unsuccessfully addresses the prompt and is marked by one or more of the following.</p> <ul style="list-style-type: none"> ▪ Parts of the requested information, opinion, or support for that opinion are missing or inappropriate/incoherent. ▪ Organization, development, and/or coherence is generally inappropriate for the task. ▪ Tone and register are not appropriate for the intended audience. ▪ Command of the language is limited. Use of structures, syntactic variety, and/or vocabulary obscure meaning.
1	<p>The response fails to address the prompt and exhibits one or more of the following.</p> <ul style="list-style-type: none"> ▪ The requested information, opinion, and support for that opinion are not present. ▪ The response is off topic. ▪ Organization, development, and coherence are inadequate. ▪ Serious and frequent errors in structure and vocabulary (use of isolated words) severely limit comprehensibility.
0	No response OR no English in the response.

Appendix B. Redesigned TOEIC Bridge Proficiency Descriptors

Redesigned TOEIC Bridge Listening Test Proficiency Descriptors

The performance descriptors outline the types of general skills and abilities in understanding spoken English that are typical of test takers who have achieved similar scores. The descriptor associated with the score will help test takers understand the strengths and weaknesses of their listening ability in English. Each test taker will receive a description of listening proficiency in English on his or her score report.

Listening Score: Scaled Score 15

Test takers in this score range can understand a few words, very simple phrases, and some short sentences that are spoken clearly and very slowly. Some test takers can recognize individual words such as numbers or days of the week. Some test takers can understand highly predictable questions and statements when they are spoken one phrase at a time. Some test takers may be able to understand a limited range of sentences consisting of very simple grammatical structures and very common vocabulary on very familiar topics.

Listening Score Range: Scaled Score 16 to 25

Test takers in this score range typically can understand some slowly spoken words, simple phrases, and short sentences on familiar topics. Test takers in this score range can understand short pieces of speech that are spoken clearly and very slowly. Generally they can understand short formulaic phrases, simple sentences, and simple grammatical structures when common vocabulary is used. They can understand short, predictable messages and instructions on familiar topics. They can understand unconnected speech, one sentence at a time.

Listening Score Range: Scaled Score 26 to 38

Test takers in this score range typically can understand short spoken sentences and a limited range of grammatical structures; they can understand short conversational exchanges on familiar topics. Typically test takers in this score range can understand speech that is clear and slow. They can usually understand key words, formulaic phrases and expressions, and relatively short, sentence-length speech. Generally test takers in this score range can understand spoken language on familiar topics and routines. They understand simple descriptions and information about people, family, shopping, location, and employment. Most of

the time, test takers can understand simple sentences and simple grammatical structures, and they may inconsistently understand some complex sentences and structures. Test takers in this score range can occasionally understand implied meanings.

Listening Score Range: Scaled Score 39 to 50

Test takers in this score range typically can understand short spoken conversations and monologues made up of connected sentences and some complex structures. They can understand some implied meaning and some abstract ideas. Typically test takers in this score range can usually understand a range of common vocabulary and some complex sentences and grammatical structures. Generally test takers in this score range can understand concrete topics and some abstract ideas related to work and other familiar contexts. In this score range, test takers understand some implied meanings and can connect facts in conversations and short spoken monologues with information in a short written text. They can comprehend formal and informal spoken language if the topics are relevant and familiar.

Redesigned TOEIC Bridge Reading Test Proficiency Descriptors

The performance descriptors outline the types of general skills and abilities in understanding written English that are typical of test takers who have achieved similar scores. The descriptor associated with the score will help test takers understand the strengths and weaknesses of their reading ability in English. Each test taker will receive a description of reading proficiency in English on his or her score certificate.

Reading Score Range: Scaled Score 15 to 18

In this score range, test takers may succeed in identifying a limited number of words and phrases related to very familiar needs. The words that test takers are likely to identify are very common words and/or phrases that are strongly supported by context. Some test takers may understand simple instructions such as “Stop,” “No exit,” especially with illustrations to help understanding. Occasionally they may be able to understand material longer than a single phrase. Some test takers may only be able to recognize letters of the alphabet.

Reading Score Range: Scaled Score 19 to 33

Test takers in this score range can typically identify familiar words and phrases in very short texts written with very common vocabulary and basic grammatical structures. They can understand some simple language on familiar topics. Typically test takers in this score range can understand very short texts on familiar topics. They can understand some simple phrases and sentences, especially those supported by visual cues and common formats. For example, they can identify some details of written language on signs and in schedules. They may understand the overall meaning of simple texts by recognizing common words and phrases.

Reading Score Range: Scaled Score 34 to 44

Test takers in this score range can typically understand short texts written with common vocabulary and basic grammatical structures. They can understand simple language used to describe familiar topics. Typically test takers in this score range understand writing that is short and simple. They can understand the overall meaning of written language in a variety of formats such as e-mails, letters, and web pages. They are developing familiarity with the basic organization of texts in English and can sometimes use this knowledge to support their understanding. They can usually understand both the overall meaning and the purpose of written communication on familiar topics such as family, shopping, and employment. Most of the time, test takers can understand simple sentences and simple grammatical structures, and they may occasionally understand a limited range of complex sentences.

Reading Score Range: Scaled Score 45 to 50

Test takers in this score range can understand short written texts in personal, public, and familiar workplace contexts and across a range of formats. Typically test takers in this score range can understand a variety of common texts such as web pages, letters, and articles written in formal and informal styles. They are familiar with the basic organization of short texts in English and can use this knowledge to support their understanding. They can understand vocabulary related to concrete topics as well as some abstract topics related to everyday life. They are familiar with a variety of grammatical structures and are developing the ability to understand complex sentences and structures. They can connect information across sentences. They can understand overall meaning, purpose, and many details. They can sometimes understand meaning that is implied rather than directly stated.

Redesigned TOEIC Bridge Speaking Test Proficiency Descriptors

Speaking Score Range: Scaled Score 15 to 22

Test takers in this score range are developing the ability to produce words and short phrases. Test takers in this score range can occasionally use simple words or phrases to identify people, objects, places, and activities that are highly familiar. They are developing the ability to read short texts aloud.

Speaking Score Range: Scaled Score 23 to 36

Test takers in this score range can typically use spoken English to perform very familiar and routine social interactions. They can use common and some high-frequency words and simple phrases, and they have limited control of simple structures. Listener effort is typically needed to understand the test taker's meaning due to issues with pronunciation, intonation, word stress, choice of vocabulary, and use of grammatical structures.

- Test takers in this score range can occasionally ask for and provide basic information.
- Test takers in this score range are developing the ability to describe people, objects, places, and activities.
- Test takers in this score range can sometimes express basic preferences, likes, and dislikes about very familiar topics.
- Test takers in this score range can occasionally give a basic description of simple and very familiar events.

Speaking Score Range: Scaled Score 37 to 42

Test takers in this score range can typically use spoken English to perform simple communication tasks involving familiar everyday activities, experiences, wants, and needs. They can use phrases, short sentences, and some longer sentences. They have some control over simple grammatical structures and vocabulary. At times, listener effort may be needed to understand the test taker's meaning due to occasional issues with pronunciation, intonation, word stress, choice of vocabulary, and use of grammatical structures.

- Test takers in this score range can typically ask for and provide simple and direct information.

- Test takers in this score range can usually give basic descriptions of people, objects, places, and activities, though meaning may be obscured at times.
- Test takers in this score range can typically sequence simple events to tell a story, but part of the story may be unclear. The test taker can use simple, linear connectors such as *and* or *then*.
- Test takers in this score range can typically ask and answer questions and make simple requests, offers, and suggestions, but attempts may be incomplete or unclear at times.
- Test takers in this score range can sometimes express a simple opinion or recommendation, but they may only be able to provide limited support for the recommendation.

Speaking Score Range: Scaled Score 43 to 50

Test takers in this score range can typically use spoken English to perform a variety of communicative tasks relevant to everyday life and the speaker’s areas of interest. When needed, they can combine sentences to produce connected discourse. Their use of common vocabulary is appropriate. They have good control of simple sentence structures and some control of more complicated sentence structures. Some errors may occur that do not affect meaning. Pronunciation, intonation, and word stress are generally intelligible but may require some listener effort.

- Test takers in this score range can ask for and provide basic information.
- Test takers in this score range can describe objects and people performing activities.
- Test takers in this score range can express thanks and make simple requests, offers, and suggestions.
- Test takers in this score range can narrate and sequence simple events.
- Test takers in this score range can express a simple opinion and give a reason for it.

Redesigned TOEIC Bridge Writing Test Proficiency Descriptors

Writing Score Range: Scaled Score 15 to 19

Test takers in this score range are developing the ability to write simple words and phrases in order to provide basic personal information such as name, address, age, etc. They typically know the alphabet and can copy words.

- Some test takers in this score range can communicate very simple information about themselves.
- Some test takers in this score range can use simple words or phrases to identify people, objects, places, and activities.

Writing Score Range: Scaled Score 20 to 31

Test takers in this score range can typically write phrases and simple sentences and make use of a limited range of very common vocabulary about very familiar subjects. They can use writing to meet some limited, basic, and practical communication needs, though their writing is sometimes unclear. They have limited control of simple grammatical structures and may have difficulty with word order and word forms.

- Test takers in this score range can communicate very basic information about themselves.
- Test takers in this score range can sometimes give a basic description of people, objects, places, and activities.
- Test takers in this score range are developing the ability to narrate events relating to daily life. They can include some relevant details. They can sequence words and phrases with basic connectors such as *and* or *then*.

Writing Score Range: Scaled Score 32 to 42

Test takers in this score range can typically write phrases and sentences about familiar topics, such as family, people, places, and work. They generally have adequate control of simple grammatical structures and an adequate range of common vocabulary that allow them to meet basic communication needs. Typically there are minor errors in their writing, and some errors may obscure meaning at times.

- Test takers in this score range can typically ask for and provide basic information. However, some important details may be missing or otherwise inappropriate for the task.
- Test takers in this score range can sometimes make simple requests, offers, and suggestions in familiar, everyday situations.
- Test takers in this score range can express basic preferences, likes, and dislikes about familiar topics. However, they may be unable to give a clear reason for their preference.
- Test takers in this score range can usually describe people, objects, places, and activities, though errors may obscure meaning at times.
- Test takers in this score range can describe a simple series of events using a logical sequence. However, the story may be incomplete or underdeveloped. Errors may obscure meaning at times.

Writing Score Range: Scaled Score 43 to 50

Test takers in this score range can typically write sentences, paragraphs, and short essays about familiar topics that contain both abstract and concrete ideas. They generally have good control of common grammatical structures and a good range of common vocabulary that allow them to communicate moderately complex messages. They can connect sentences to form paragraphs that are organized and coherent. Typically there are some minor errors in their writing when expressing complex thoughts or unfamiliar topics.

- Test takers in this score range can ask for and provide basic information.
- Test takers in this score range can describe objects and locations as well as people performing activities.
- Test takers in this score range can express thanks and make simple requests, offers, and suggestions.
- Test takers in this score range can narrate and sequence simple events and routines.
- Test takers in this score range can express a simple opinion and give a reason for it.