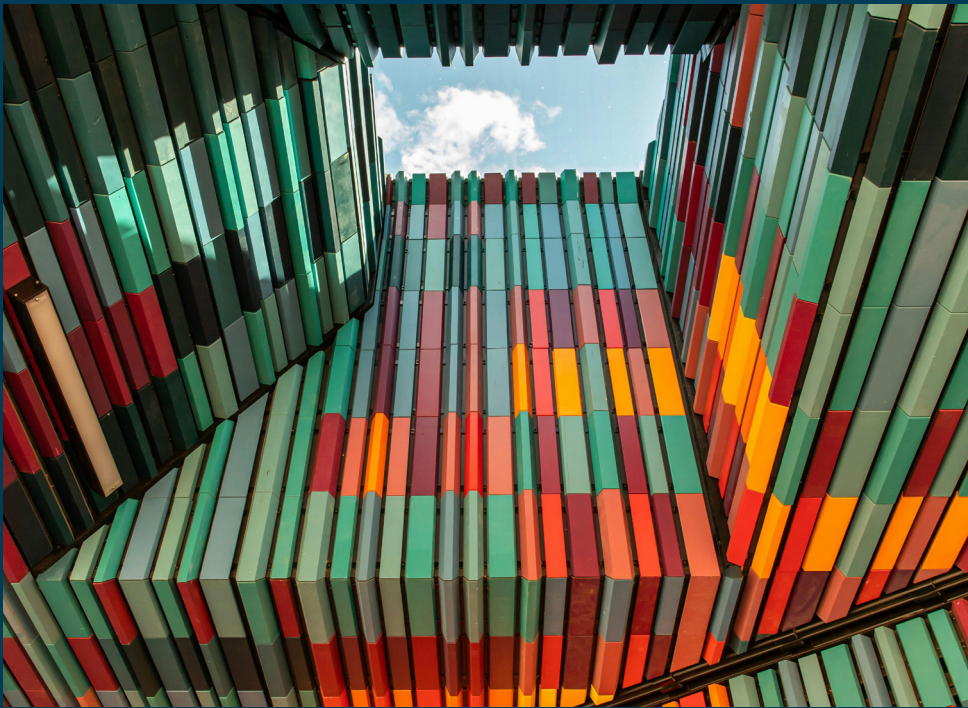


*ets research institute

APRIL
2026
RR-26-04

TOEIC[®] Link[™] Assessments Technical Manual



AUTHORS

Jaime Cid, Jonathan Schmidgall, and Elizabeth Park

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey
Lord Chair in Measurement and Statistics

ASSOCIATE EDITORS

Usama Ali
Senior Measurement Scientist

Beata Beigman Klebanov
Principal Research Scientist, Edusoft

Katherine Castellano
Managing Principal Research Scientist

Larry Davis
Director Research

Jamie Mikeska
Managing Senior Research Scientist

Teresa Ober
Research Scientist

Jonathan Schmidgall
Senior Research Scientist

Jesse Sparks
Managing Senior Research Scientist

Zuwei Wang
Senior Measurement Scientist

Klaus Zechner
Senior Research Scientist

Jiyun Zu
Senior Measurement Scientist

PRODUCTION EDITORS

Ayleen Gontz
Manager Editorial Services

Emily Ramsower
Manager Editorial Services

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 Report 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 Report 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.

TOEIC[®] Link[™] Assessments Technical Manual

Version 1.0

Jaime Cid, Jonathan Schmidgall, and Elizabeth Park
ETS Research Institute, ETS, Princeton, New Jersey, United States

Abstract

This technical manual provides a comprehensive overview of the TOEIC[®] Link[™] assessments, offering detailed insights into their purpose, design, and intended users. The manual begins with an introduction that outlines the assessment objectives and target audience. Subsequent sections delve into the specific constructs and tasks of the four TOEIC Link assessments: Listening, Reading, Speaking, and Writing. The remaining sections examine the design and development processes for both the listening and reading, as well as the speaking and writing assessments, highlighting methodologies used to ensure validity and reliability. Together, these sections present a thorough guide for test users and stakeholders seeking to understand the TOEIC Link assessments and their application in measuring English language proficiency. Designed as a living document, this manual will be updated as the test's design, administration, scoring, and evidence of measurement quality (including reliability, validity, and fairness) evolve, along with its intended uses.

Keywords: English language proficiency, test tasks, test design, TOEIC[®] Link[™], reliability, validity, fairness

Corresponding author: J. Cid, Email: jcid@ets.org

Contents

I. Introduction	1
II. Assessment Constructs And Tasks	2
II-1. TOEIC Link Listening Assessment.....	2
Construct Definition	3
Task Types	4
Photographs.....	4
Question Response.....	5
Conversations.....	6
Talks.....	7
II-2. TOEIC Link Reading Assessment.....	8
Construct Definition	8
Task Types	9
Incomplete Sentences.....	9
Text Completion.....	10
Reading Comprehension Passages.....	11
II-3. TOEIC Link Speaking Assessment.....	12
Construct Definition	12
Foundational Task Types	13
Read a Dialogue Aloud	13
Listen and Repeat.....	14
Communicative Task Type	15
Express an Opinion	15
II-4. TOEIC Link Writing Assessment.....	16
Construct Definition	16
Foundational Task Type.....	17
Build a Sentence	17
Communicative Task Types.....	18
Write a Brief Email	18
Write an Extended Email	19
Write About Your Perspective.....	20
III. Assessment Design and Development	22

III-1. TOEIC Link Listening and Reading Assessment Design	22
Theoretical Considerations: Adaptive Testing Design	22
MSA Design for Listening and Reading Assessments	23
III-2. TOEIC Link Speaking and Writing Assessment Design	24
Speaking Assessment Design	24
IV. Technical Summary	26
IV-1. Listening and Reading Scoring	26
IV-2. Speaking and Writing Scoring	27
The Automated Scoring Approach of Speaking and Writing	27
IV-3. Evaluation of Machine Scores for Speaking and Writing	30
V. Score Reporting	31
V-1. Scale Scores and Ranges	31
V-2. Development of Scoring Guides for Speaking and Writing	31
V-3. The Common European Framework of Reference Languages	31
V-4. Reporting of Descriptors	32
VI. Score Reliability and Standard Error of Measurement	33
VII. Validity and Fairness	35
VII-1. Validity	35
VII-2. Fairness	37
VIII. Assessment Administration and Security	39
VIII-1. Assessment Display Sequence	39
VIII-2. Secure Exam Browser	39
VIII-3. Remote Proctoring Options	40
References	41
Appendix A: Speaking Assessment Human Scoring Rubrics	44
Scoring Guide: Read a Dialogue Aloud	44
Scoring Guide: Listen and Repeat	45
Scoring Guide: Express an Opinion	46
Appendix B: Writing Assessment Human Scoring Rubrics	47
Scoring Guide: Write a Brief Email	47
Scoring Guide: Write an Extended Email	48

I. Introduction

The purpose of the TOEIC® Link™ assessments is to measure the English language proficiency of second or foreign language speakers. The assessments are intended for young adults and adults learning English as a second or foreign language who have diverse nationalities, native languages, educational backgrounds, and various purposes for learning English. The assessments are expected to be used by businesses and institutions to support job screening, selection, promotion, and placement decisions and to monitor language learning progress over time.

The TOEIC Link assessments evaluate English language proficiency from basic to advanced levels. The assessments measure up to four skills (listening, reading, speaking, and writing), offering a quick, convenient, and fully online assessment solution. The assessments can be administered modularly, allowing any combination of the four skills to be assessed together or separately.

Designed to assess foundational and communicative language skills, the TOEIC Link assessments provide an indicator of general English proficiency. The content focuses on everyday adult life and general workplace settings, avoiding specialized knowledge or vocabulary.

This technical manual adheres to the professional guidelines outlined in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014) and the *ETS Standards for Quality and Fairness* (ETS, 2014). These guidelines represent the consensus of measurement professionals, and adherence to them reflects ETS's commitment to maintaining high professional and ethical standards.

Designed as a living document, this manual will be updated as the assessments' design, administration, scoring, and evidence of measurement quality (including reliability, validity, and fairness) evolve, along with its intended uses. Version 1.0 of the manual was produced to accompany the initial launch of the assessments in March 2025.

II. Assessment Constructs and Tasks

The TOEIC Link assessments were developed through a collaborative effort involving researchers, assessment developers, and psychometricians. The design team worked closely with business directors to establish requirements ensuring the assessment meets the needs of score users, English language learners, and other stakeholders.

Key requirements for assessment design included

- assessing a range of CEFR proficiency levels from A1 to C1,
- limiting the total testing time for all four skills to no more than 90 minutes,
- using the same reporting scale for individual section scores and the total score of all four skills,
- ensuring fully digital assessment delivery, from registration to score reporting,
- supporting a mobile phone-based delivery model, and
- completing automated scoring and score report delivery within 48 hours.

With these requirements in mind, the team adopted a principled approach to assessment design, which involved evaluating an extensive catalog of assessment tasks for appropriateness and drafting an initial blueprint. The design team consulted with experts in automated scoring, artificial intelligence (AI), user experience (UI/UX), and information technology (IT) to explore how technology could enhance task scoring, feature design, and user interfaces. After reviewing ETS's extensive catalog of existing test tasks, the team identified those that could provide suitable evidence to support inferences about test takers' English proficiency while meeting the assessment design requirements.

Preliminary analysis by psychometricians determined the number of items per task type needed to ensure reliable test scores. For the Speaking and Writing assessments, rubrics and AI scoring models were evaluated and refined. A summary of the design of each assessment is provided in this section.

II-1. TOEIC Link Listening Assessment

The design of the TOEIC Link Listening assessment is based on the TOEIC Listening test section and consequently intends to measure the same listening skills using the same test tasks. Thus, the definition of the construct of listening comprehension in the TOEIC Link Listening

assessment and its operationalization in item and test specifications is comparable to the TOEIC Listening test section and draws upon the same conceptual framework.

Construct Definition

The Listening assessment measures the ability of beginning to advanced English language learners to understand spoken English texts in everyday and workplace contexts. The TOEIC Listening test section—the source of the TOEIC Link Listening assessment’s construct definition—was developed and updated using evidence-centered design (see Ashmore et al., 2018; Schedl, 2010), an approach to test design that requires specifying claims about test takers’ abilities.

Test tasks are designed to elicit evidence of test takers’ abilities in the following areas:

- understanding the gist in short spoken text,
- understanding the gist in extended spoken text,
- understanding details in short spoken text,
- understanding details in extended spoken text, and
- understanding purpose or implied meaning/pragmatic understanding.

These abilities are assessed in task types that incorporate a variety of everyday and workplace settings, including corporate development, dining out, entertainment, finance and budgeting, general business, health, housing/corporate property, manufacturing, offices, personnel, purchasing, technical areas, and travel (ETS, 2022, pp. 3–4). These settings provide context for test questions, but test takers are not required to have business experience or to understand specialized vocabulary (Schmidgall et al., 2024).

Task Types

The Listening assessment includes four distinct parts or task types: *Photographs*, *Question Response*, *Conversations*, and *Talks*.

Photographs

Test takers hear four statements about a picture that is displayed on their screen. Test takers must select the one statement that best describes the picture by selecting the appropriate option (A through D; see Figure 1).

This task type includes questions that measure a test taker’s understanding of the gist or central idea of the photograph and understanding of obvious details in the photograph .

Figure 1. Example of *Photographs* Task Type

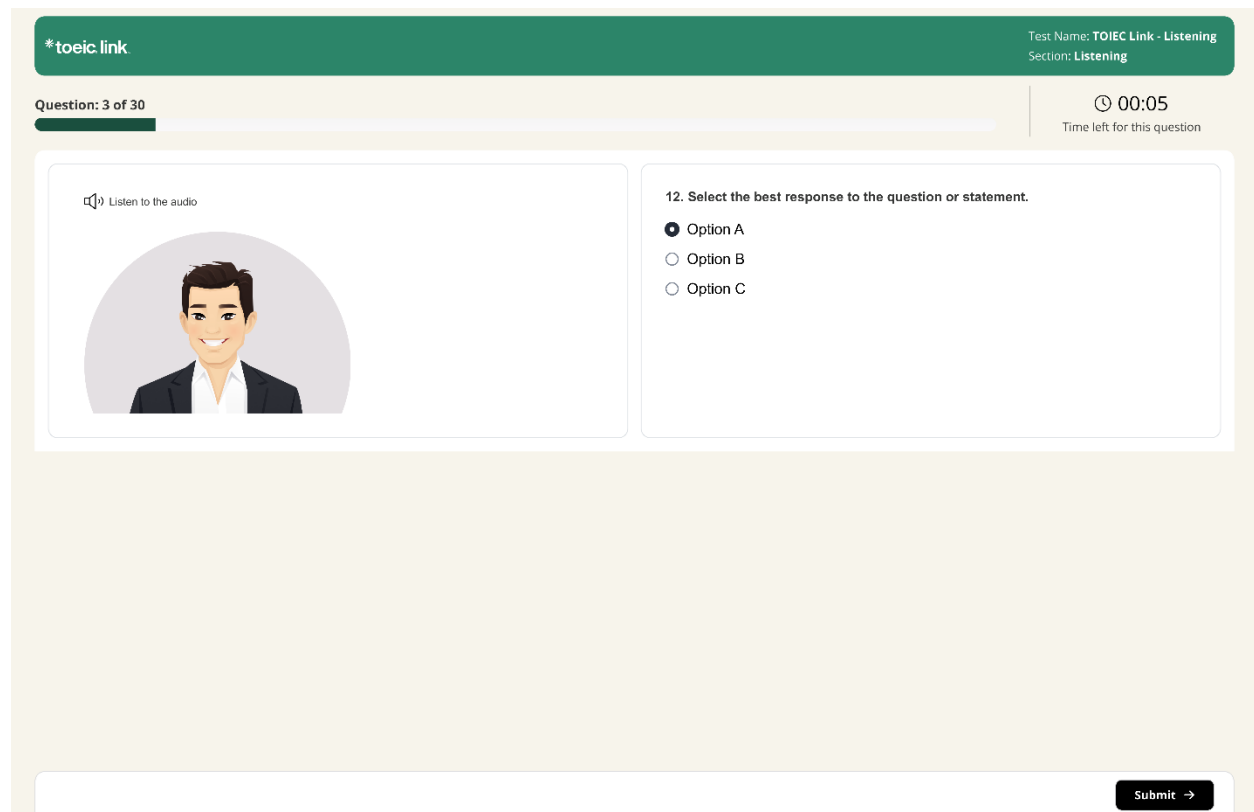
Source: TOEIC® Link™ test, ETS

Question Response

Test takers hear a question or statement followed by three potential responses spoken in English. The question and response involve two speakers. Neither the question/statement nor responses are displayed on the screen. Test takers must select the best response to the question/statement by selecting the appropriate option displayed on the screen (A through C; see Figure 2).

This task type includes questions that measure a test taker’s ability to understand gist, purpose, and basic context in short spoken texts; ability to understand details in short spoken texts; and ability to understand implied meaning in short spoken texts.

Figure 2. Example of *Question Response* Task Type



The screenshot displays a digital assessment interface. At the top, a green header bar contains the text '*toeic link' on the left and 'Test Name: TOEIC Link - Listening' and 'Section: Listening' on the right. Below the header, the interface is divided into two main sections. On the left, there is a 'Listen to the audio' section featuring a speaker icon and a circular portrait of a man with dark hair, wearing a dark suit jacket over a white shirt. On the right, the question text reads '12. Select the best response to the question or statement.' Below this text are three radio button options: 'Option A' (which is selected), 'Option B', and 'Option C'. At the bottom right of the interface, there is a dark button labeled 'Submit →'. In the top right corner of the interface, a clock icon indicates a time limit of '00:05' with the text 'Time left for this question' below it. In the top left corner, it says 'Question: 3 of 30'.

Source: TOEIC® Link™ test, ETS

Conversations

Test takers hear a conversation between two or three speakers. The length of each conversation varies between 70 and 115 words. Some sets also contain a graphic (visual stimulus). After listening to each conversation, test takers answer three questions about what speakers say (see Figure 3). Each question is spoken aloud, and questions and response options are displayed on the screen. Test takers select the best response displayed on the screen (A through D).

The questions associated with this task measure a test taker’s ability to infer gist (main idea, context), understand details, and understand a speaker’s purpose or implied meaning in a phrase or sentence.

Figure 3. Example of *Conversations* Task Type

*toeic.link

Test Name: TOEIC Link - Listening
Section: Listening

Question: 6 - 8 of 30

00:05
Time left for this question

Listen to the audio

6. What problem does the man mention?

- His car is out of fuel.
- His phone battery is empty.
- He is late for an appointment.
- He forgot his wallet.

7. Where are the speakers?

- At a train station
- At an electronics repair shop
- At a furniture store
- At a coffee shop

8. What does the woman suggest the man do?

- Check a Web site
- Call a taxi
- Return at a later time
- Go to the library

Submit →

Source: TOEIC® Link™ test, ETS

Talks

Test takers listen to talks given by a single speaker. The length of the talks varies between 70 and 115 words. Some talks also contain a graphic (visual stimulus). After hearing each talk, test takers answer three questions about what the speaker says (see Figure 4). Each question is spoken aloud and displayed on the screen. Test takers select the best response option displayed on the screen (A through D).

Similar to *Conversations*, questions associated with this task type measure a test taker’s ability to infer gist (main idea, context), understand details, and understand a speaker’s purpose or implied meaning in a phrase or sentence.

Figure 4. Example of *Talks* Task Type

The screenshot displays a TOEIC Link Listening question interface. At the top, it shows the test name 'TOEIC Link - Listening' and the section 'Listening'. The question is identified as 'Question: 12 - 14 of 30' and has a timer of '00:05' left. The main content area is divided into two columns. The left column features a speaker's profile (a man in a suit) and a 'Listen to the audio' button. The right column contains three multiple-choice questions. The first question asks 'What is the talk mainly about?' with four options: 'A mobile phone model' (selected), 'An office security system', 'High-speed Internet service', and 'Business scheduling software'. The second question asks 'Why did the company choose the product?' with four options: 'It makes arranging meetings easy.' (selected), 'It is reasonably priced.', 'It has good security features.', and 'It has received positive reviews.'. The third question asks 'What does the speaker say is offered with the product?' with four options: 'An annual upgrade' (selected), 'A money-back guarantee', 'A mobile phone application', and 'A customer-service help line'. A 'Submit' button is located at the bottom right of the question area.

Source: TOEIC® Link™ test, ETS

II-2. TOEIC Link Reading Assessment

The design of the TOEIC Link Reading assessment is based on the TOEIC Reading test section and consequently measures the same reading skills using the same test tasks. Thus, the definition of the construct and its operationalization in item and test specifications is comparable to the TOEIC Reading test section and draws upon the same conceptual framework.

Construct Definition

The TOEIC Link Reading assessment measures the ability of beginning to advanced English language learners to understand written English texts in everyday and workplace contexts. The TOEIC Reading test section was developed and updated using evidence-centered design (Schedl, 2010; Ashmore et al., 2018), an approach to test design that requires specifying claims about test takers' abilities.

Test tasks are designed to elicit evidence of test takers' abilities in the following areas:

- making inferences based on information that is explicitly stated in texts (within text inferences and across text gist),
- understanding specific (factual) information in tables and passages,
- connecting information across multiple sentences in single texts and across two or three texts,
- understanding vocabulary, and
- understanding grammar.

These abilities are assessed in task types that incorporate a variety of everyday and workplace settings, including corporate development, dining out, entertainment, finance and budgeting, general business, health, housing/corporate property, manufacturing, offices, personnel, purchasing, technical areas, and travel (ETS, 2022, pp. 3–4). These settings provide context for test questions, but test takers are not required to have business experience or to understand specialized vocabulary (Schmidgall et al., 2024).

Task Types

The Reading assessment includes three distinct parts or task types: *Incomplete Sentences*, *Text Completion*, and *Reading Comprehension Passages*.

Incomplete Sentences

Test takers are presented with a sentence that has a missing word or phrase (see Figure 5). Test takers must then select from among four options the word or phrase that best completes the sentence.

Test questions associated with this task type measure a test taker’s understanding of grammar or vocabulary at the sentence level.

Figure 5. Example of *Incomplete Sentences* Task Type

The screenshot shows a digital assessment interface. At the top, there is a green header with the text '*toeic link' on the left and 'Test Name: TOEIC Link - Reading' and 'Section: Reading' on the right. Below the header, the question number 'Question: 01 of 14' is displayed. A row of 14 numbered buttons (01 to 14) is shown, with '01' highlighted. To the right of the question number, a timer shows '14:59' and the text 'Time left for this Unit'. The main question area contains the text: 'Successful candidates need to be licensed to operate _____ cranes and backhoes.' Below this text are four radio button options: 'once', 'there', 'after', and 'both'. At the bottom of the interface, there are two buttons: '← Previous' and 'Next →'.

Source: TOEIC® Link™ test, ETS

Text Completion

Test takers read short texts in a variety of formats. Each short text is missing four elements, such as words, phrases, or key sentences. Test takers must correctly identify each missing element by selecting the appropriate word, phrase, or sentence from among four options (see Figure 6).

These test questions measure a test taker’s understanding of grammar, vocabulary, and ability to connect information within a short text.

Figure 6. Example of *Text Completion* Task Type

Source: TOEIC® Link™ test, ETS

Reading Comprehension Passages

Test takers must read everyday texts (e.g., notices, letters, forms, advertisements) and answer two to five questions about each text (see Figure 7). Passages can be single (based on one passage), double (based on two passages), or triple (based on three passages).

The questions accompanying each text measure a test taker’s understanding of vocabulary, ability to identify the main idea, ability to identify stated details, ability to infer implied meanings such as the context or the writer’s purpose, or ability to connect information within or across texts.

Figure 7. Example of *Reading Comprehension Passages* Task Type (Single Passage)

Source: TOEIC® Link™ test, ETS

II-3. TOEIC Link Speaking Assessment

The TOEIC Link Speaking assessment was developed utilizing a hybrid approach to test design that balances the measurement of foundational and communicative skills, drawing upon speaking tasks and an approach to construct definition originally developed for the TOEFL Essentials® test (Davis et al., 2023; Papageorgiou et al., 2021). In this approach, some test tasks are designed to measure foundational language skills (e.g., fluency, pronunciation, accurate reproduction of language) that have been shown to be strong predictors of speaking proficiency. These test tasks are complemented by tasks designed to provide a more direct measurement of communicative skills in an authentic or real-world setting. This hybrid approach to test design is intended to achieve efficiencies through targeted assessment of foundational skills while retaining measurement of communicative skills.

Construct Definition

The TOEIC Link Speaking assessment measures the general speaking proficiency of beginning to advanced English language learners in everyday and workplace contexts.

Measurement of general speaking proficiency is achieved using test tasks that elicit evidence of (a) foundational skills that have been shown to be predictive of overall speaking proficiency and (b) communicative skills that are typical of those needed in everyday and workplace contexts.

Foundational Task Types

Read a Dialogue Aloud

Test takers read aloud one part of a dialogue that takes place in a daily life or workplace context (see Figure 8). Test takers are first provided with information on who the interlocutor is (e.g., friend), and then they hear the interlocutor make a statement or ask a question. Test takers then respond by reading aloud their part of the conversation, which is written on the screen. Each test taker’s “turn” in the conversation consists of one to three sentences. The topic of the conversation includes everyday and workplace contexts. Test takers are given 25 seconds to prepare and 40 seconds to complete each item.

This task measures a test taker’s ability to accurately and fluently process written text into speech in a way that is intelligible to proficient speakers of English. This ability includes foundational speaking skills such as fluency, pronunciation, pacing, and prosodic features such as stress and intonation.

Figure 8. Example of Read a Dialogue Aloud Task Type

The screenshot displays the TOEIC Link™ assessment interface for a 'Read a Dialogue Aloud' task. At the top, a green header contains the TOEIC Link logo and the text '*toeic link'. On the right side of the header, it indicates 'Test Name: TOEIC Link - Speaking' and 'Section: Speaking'. Below the header, a progress bar shows 'Question: 1 of 11'. The main content area is divided into two panels. The left panel features a speaker icon with the text 'Listen to the audio' and 'Read the words aloud only once.' Below this is an illustration of a man with a beard and curly hair, wearing a yellow shirt. Underneath the illustration, the text reads: 'It is rather nice not having to commute far for work, and I love not having to share with a roommate. But I just found out that I'll be sent to the office in Italy for a few months next spring. I wasn't sure if I'd be selected, but it looks like my assignment just came through.' The right panel contains a 'Preparation Time' box with a timer set to '00:00:25' and a 'Response time: 40 seconds' label. At the bottom right of the interface, there is a 'Submit →' button.

Source: TOEIC® Link™ test, ETS

Listen and Repeat

Test takers repeat a series of sentences within a scenario in an everyday or workplace context (see Figure 9). Test takers are first presented with the scenario which provides a communicative purpose for listening and repeating the sentences. The topic of the series of sentences includes everyday and workplace contexts. Each sentence test takers listen to is associated with a visual/graphic that corresponds to the content of what is being heard. After each sentence, there is a pause, and test takers repeat exactly what was said. Sentences get progressively longer and more complex as test takers progress through the scenario.

This task measures a test taker’s ability to process linguistic information and reproduce it accurately and intelligibly, foundational skills that are essential elements of spoken language production.

Figure 9. Example of *Listen and Repeat* Task Type

Source: TOEIC® Link™ test, ETS

Communicative Task Type

Express an Opinion

Test takers are presented with a topic and are asked to give their opinion about an everyday or workplace topic (see Figure 10). Test takers are given 45 seconds to prepare and 60 seconds to respond.

This communicative task measures a test taker’s ability to create connected, sustained, intelligible discourse appropriate to the typical workplace or casual social situation. The ability to state and support an opinion, as measured by this task, also includes foundational language skills related to delivery, intelligibility, effective and accurate use of grammar and vocabulary, and topic development and coherence.

Figure 10. Example of *Express an Opinion* Task Type

The screenshot displays the TOEIC Link assessment interface for an 'Express an Opinion' task. At the top, the header shows '*toeic link' on the left and 'Test Name: TOEIC Link - Speaking' and 'Section: Speaking' on the right. Below the header, a progress bar indicates 'Question: 11 of 11'. The main content area is divided into two sections. The left section contains a speaker icon and the text 'Listen to the audio'. Below this, it asks 'Do you agree or disagree with the following statement?' and provides the statement: 'The best way for university students to spend a long university break is by gaining work experience through internships at a company.' It then prompts the test taker to 'Give reasons or examples to support your opinion.' The right section features a 'Preparation Time' timer set to '00:00:45' and a 'Response time: 60 seconds' indicator. At the bottom right, there is a 'Submit' button with a right-pointing arrow.

Source: TOEIC® Link™ test, ETS

II-4. TOEIC Link Writing Assessment

The TOEIC Link Writing assessment was developed using the same hybrid approach to test design described previously for the TOEIC Link Speaking assessment, drawing upon writing tasks developed for the TOEFL Essentials test (see also Davis et al., 2023; Papageorgiou et al., 2021).

Construct Definition

The TOEIC Link Writing assessment measures the general writing proficiency of beginning to advanced English language learners in everyday and workplace contexts.

Measurement of general writing proficiency is achieved using test tasks that elicit evidence of (a) foundational skills that have been shown to be predictive of overall writing proficiency and (b) communicative skills that are typical of those needed in everyday and workplace contexts.

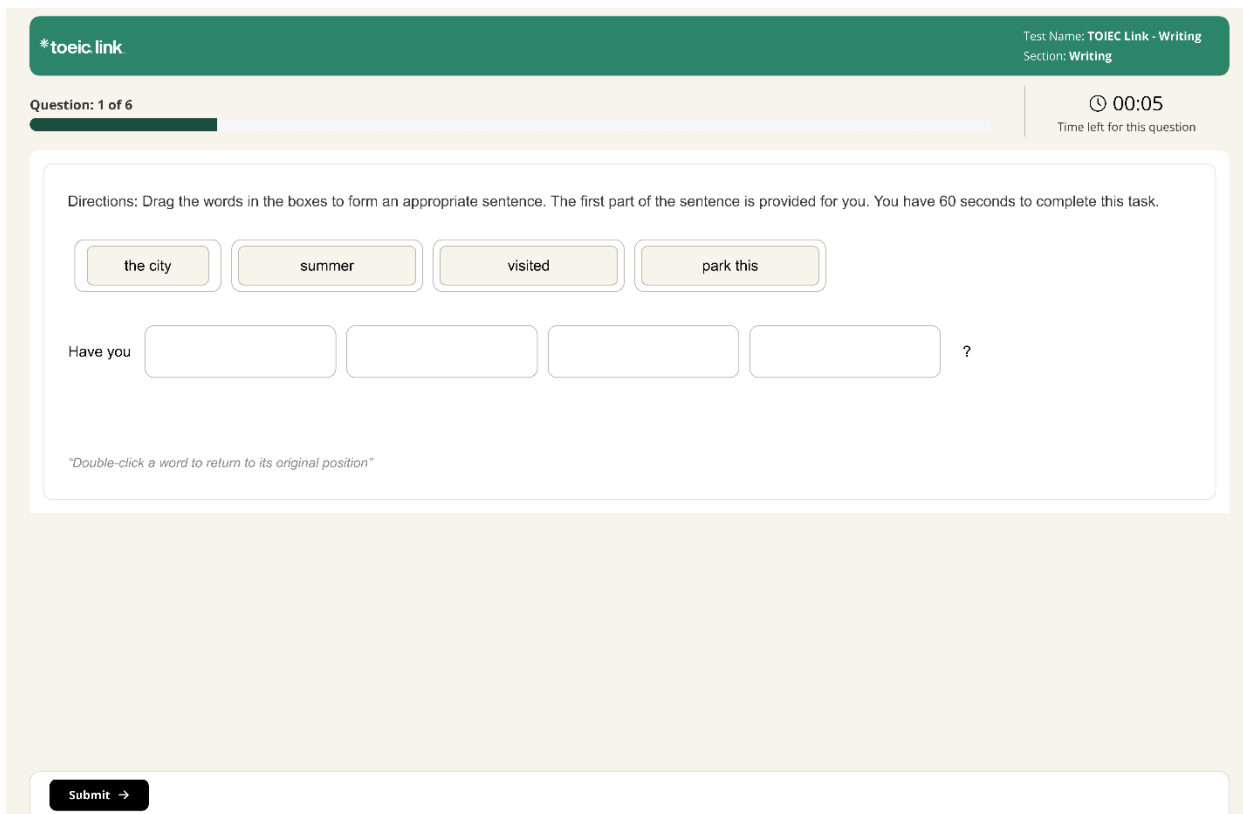
Foundational Task Type

Build a Sentence

Test takers read a sentence or question with words or phrases in the wrong order, and they drag and drop words or phrases to form a grammatical sentence or question (see Figure 11). The maximum length of each sentence or question is 15 words, and the topics include everyday personal, public, and familiar workplace topics. Test takers have 1 minute to complete each item.

This task measures a test taker’s ability to employ word-order rules in English to produce grammatically correct sentences, an important foundational skill for writing proficiency.

Figure 11. Example of *Build a Sentence* Task Type



Source: TOEIC® Link™ test, ETS

Communicative Task Types

Write a Brief Email

Test takers are presented with a scenario in either a workplace or social setting and a table with information used to provide the context (see Figure 12). They are asked to write an email and are provided with two specific tasks to accomplish in the email. Scenarios are drawn from contexts such as dining, health, purchasing, transportation, and work. The table presented to the test takers includes contents such as information about a product or service. The two tasks test takers need to accomplish in the email include communicative goals such as providing information, making a recommendation, asking for help, and responding to a request. Test takers are given 7 minutes to respond and are informed to write as much as they can without a word count recommendation.

This communicative task measures a test taker’s ability to write a brief message that achieves a designated communication goal (e.g., share information, extend an invitation, make a recommendation) and provide some elaboration. Achieving the communicative goal involves foundational skills that include grammatical accuracy and range, vocabulary accuracy and range, elaboration, and pragmatic appropriateness.

Figure 12. Example of *Write a Brief Email* Task Type

The screenshot displays a digital writing task interface. At the top, a green header bar contains the text '*toeic link' on the left and 'Test Name: TOEIC Link - Writing' and 'Section: Writing' on the right. Below the header, the question is identified as 'Question: 4 of 6'. A progress bar is visible. On the right side, a timer shows '00:05' with the text 'Time left for this question' below it. The main content area is divided into two sections. The left section contains the task prompt: 'Your friend asked you to give her advice about which of two business courses to take. Use the information from the table and write an e-mail to your friend Padma.' It lists two bullet points: 'Recommend one of the courses in the table' and 'Explain why you think it is the better choice'. Below this is a table with three columns: 'Course', 'Advantages', and 'Disadvantages'. The table contains two rows of data. The right section is titled 'Your Response:' and includes a text input area with a 'To: Padma' and 'Subject: Business course' header. Below the input area are buttons for 'Cut', 'Paste', 'Undo', and 'Redo', and a 'Hide word count: 0' option. At the bottom of the interface is a 'Submit →' button.

Course	Advantages	Disadvantages
Advertising	Popular professor	Early morning class
Business Law	Meets once a week	Lots of homework

Source: TOEIC® Link™ test, ETS

Write an Extended Email

Test takers are presented with a scenario in either a workplace or social setting). They are tasked to write an email and are provided three specific tasks to accomplish in the email (see Figure 13). The context of the scenario includes various work and life situations such as dining, health, purchasing, transportation, and work. The three tasks test takers need to accomplish in the email may involve providing information, making a recommendation, describing a problem, expressing complaints, and refusing plans. The relationship among between the sender (test taker) and the recipient of the email includes friend–friend, employee–customer, and customer–business. Test takers are given 7 minutes to respond and are informed to write as much as they can without a word count recommendation.

This communicative task measures a test taker’s ability to write a message that achieves a designated communication goal (e.g., describe problem, plans, situations, request or suggest a solution, make a complaint) and provide some elaboration. The ability to write a text for a communication goal also involves foundational skills that include grammatical accuracy and range, vocabulary accuracy and range, elaboration, and pragmatic appropriateness.

Figure 13. Example of *Write an Extended Email* Task Type

The screenshot displays the TOEIC Link assessment interface for a 'Write an Extended Email' task. At the top, the header shows the 'toeic link' logo and the test name 'TOEIC Link - Writing'. Below the header, the question number 'Question: 5 of 6' is shown on the left, and a timer indicates '00:05' time left. The task scenario is presented in a text box: 'Your friend Clara is moving to a new home tomorrow. She has asked you to help her with the move. You want to help; however, you hurt your knee yesterday, and your doctor said you should avoid carrying anything heavy for a week.' Below the scenario, the instruction reads: 'Write an e-mail to your friend Clara. In your e-mail, do the following.' This is followed by three bullet points: 'Express a desire to help.', 'Explain why you cannot help.', and 'Suggest an alternative.' The response area is titled 'Your Response:' and includes a header with 'To: Clara' and 'Subject: Your move'. Below this is a rich text editor toolbar with 'Cut', 'Paste', 'Undo', and 'Redo' buttons, and a 'Hide word count: 0' indicator. At the bottom of the interface, there is a 'Submit' button with a right-pointing arrow.

Source: TOEIC® Link™ test, ETS

Write About Your Perspective

Test takers are presented with an online discussion forum that includes a moderator's post (see Figure 14). The moderator's post is the initial post in the discussion and introduces the topic of the discussion and two posts from two different participants. The posts provide ideas and information to stimulate test takers' responses (see Figure 14). Test takers are tasked to read the posts by the moderator and two participants and write a post responding to the moderator's question. They are told to express and support their opinion and make a contribution to the discussion. The topics of the forum are related to everyday life and workplace and are accessible to a general audience. Test takers are given 10 minutes to complete the task and are recommended to write at least 100 words.

This task measures a test taker's ability to produce a written text that clearly elaborates an argument for a position, responding to arguments and/or using information provided in short texts. This includes foundational skills such as grammatical accuracy and range, vocabulary accuracy and range, elaboration, and content relevance.

Figure 14. Example of *Write About Your Perspective* Task Type

*toeic link
Test Name: TOEIC Link - Writing
Section: Writing

Question: 6 of 6
🕒 00:05
Time left for this question

Your are participating in an online discussion. Write a post responding to the moderator's question.

In your response you should:

- express and support your opinion
- make a contribution to the discussion

An effective reponse will contain at least 100 words.

I'm wondering what people think about the impact that technological advances in transportation have had on society. I've been thinking about the kinds of transportation that have had the greatest impact on people's lives. For this online discussion, please choose one of the following types of vehicles and explain why you think it had the greatest influence on society.

- **The automobile**
- **The bicycle**
- **The airplane**

Automobiles have definitely had the greatest impact. I just read that most modern cities are designed with an urban area in the center which is surrounded by industrial and suburban areas. This layout of the modern city is the direct result of an extensive highway system, and our highway system is what it is today because of the automobile.

I believe bicycles led to the most change. I recently watched a documentary in which historians argued that without bicycles, automobiles wouldn't have been invented. Even though they're the older technology, bicycles remain a popular vehicle to this day. Transportation as we know it would be vastly different if we hadn't invented the bicycle.

Cut Paste Undo Redo
🔍 Hide word count: 0

Submit →

Source: TOEIC® Link™ test, ETS

III. Assessment Design and Development

III-1. TOEIC Link Listening and Reading Assessment Design

Theoretical Considerations: Adaptive Testing Design

Computer adaptive testing offers significant efficiency advantages over traditional linear testing. In linear testing, all test takers receive the same set of items, regardless of their ability levels. This approach requires many items to cover a wide range of difficulties, making the test lengthy and time-consuming.

In contrast, adaptive testing adjusts the difficulty of test items to match the test taker's ability, requiring fewer items and less time. If a test taker answers incorrectly, the computer presents an easier item; if correctly, a more difficult item follows. This method ensures that items are appropriately challenging, providing meaningful information about the test taker's true ability. Unmatched items, on the other hand, offer little insight into a test taker's proficiency.

Adaptive testing involves two basic steps: item selection and ability estimation. First, the most appropriate items are selected based on the test taker's current estimated ability. Then, the test taker's ability is re-estimated based on their responses. This iterative process continues until a target criterion, such as a specific number of items, is met. A large pool of test items with accurately pre-estimated psychometric characteristics, typically based on item response theory (IRT), is essential for this process.

Adaptive tests can be designed to adapt after each individual item (item level) or after a set of items (module level). Module-level adaptive testing, known as multistage adaptive (MSA), begins with an identical set of items for all test takers and directs them to different levels in subsequent stages based on their performance. MSA balances practicality, flexibility, measurement accuracy, and control over test content coverage (Zenisky et al., 2010).

When stringent psychometric requirements are met, MSA offers practical benefits over item-level adaptive testing, such as better management of item pool usage, more control over test content, greater flexibility in test assembly, and the ability for test takers to navigate within modules. MSA is suitable for tests with both discrete items and sets of items associated with the same passage, whereas item-level adaptive testing is more suited to tests with discrete items only.

MSA Design for Listening and Reading Assessments

The TOEIC Link Listening and Reading assessments utilize an MSA design with two testing stages. Each assessment includes a routing module in Stage 1 and two modules (easy and hard) in Stage 2, following an MSA 1-2 design.

In Stage 1, all test takers complete a pre-assembled Routing Module with 17 items of medium difficulty. Based on their performance, scores are estimated and routing decisions are made. Test takers who meet or exceed a predetermined proficiency criterion are routed to the hard module in Stage 2, which contains 13 more challenging items. Those who score below the criterion are routed to the easy module, also with 13 items, but at a lower difficulty level.

Overall, there are two paths:

- Hard path: 17 items from the Routing Module + 13 items from the hard module
- Easy path: 17 items from the Routing Module + 13 items from the easy module

The two-stage MSA concludes after the second-stage module, and the final test score is calculated based on all items from both the routing module and the second-stage module taken by the test taker. That is, the final score is based not only on how many items were answered correctly but also on the difficulty of the second-stage module. Figure 15 illustrates the TOEIC Link MSA design for the Listening and Reading assessments, and Table 1 details the content design for both assessments.

Figure 15. TOEIC Link 30-Item Multistage Adaptive 1-2 Design for Each of the Listening and Reading Assessments

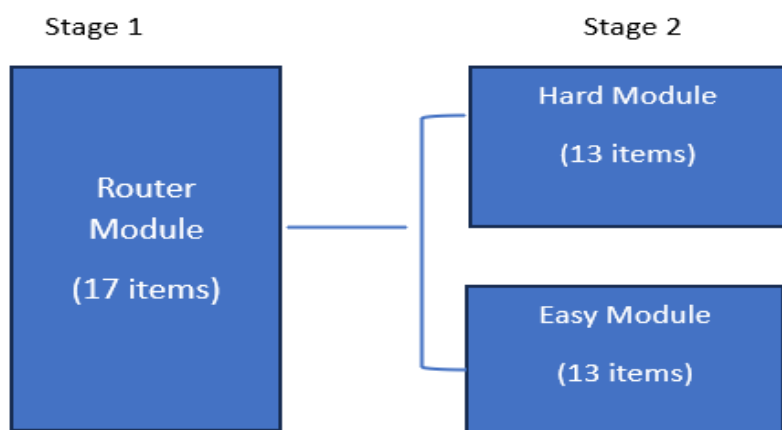


Table 1. TOEIC Link Multistage Adaptive Content Design for Listening and Reading Assessments

Assessment	Task types	Number of items in modules			Number of items in paths	
		Stage 1	Stage 2 - easy	Stage 2 - hard	Easy path	Hard path
Listening	<i>Photographs</i>	2	0	0	2	2
	<i>Question Response</i>	3	7	7	10	10
	<i>Conversations</i>	6	6	0	12	6
	<i>Talks</i>	6	0	6	6	12
	Total	17	13	13	30	30
Reading	<i>Incomplete Sentences</i>	4	5	5	9	9
	<i>Text Completion</i>	4	0	0	4	4
	<i>Reading Comprehension Passages</i>					
	Single passages	4	3	3	7	7
	Double passages	5	5	0	10	5
	Triple passages	0	0	5	0	5
	Total	17	13	13	30	30

III-2. TOEIC Link Speaking and Writing Assessment Design

The TOEIC Link Speaking and Writing assessments use a linear test design, where all test takers receive the same set of test items. Tasks in both sections are designed to be accessible across a range of proficiency levels with many opportunities for test takers to demonstrate their writing and speaking skills. A range of difficulty combined with multiple measurement opportunities makes it possible to cover the full range of language proficiency without the need for separate stages. Scores for the Writing and Speaking sections are based on overall performance on all tasks.

Speaking Assessment Design

The Speaking assessment consists of three task types:

- Task 1: *Read a Dialogue Aloud*
- Task 2: *Listen and Repeat*
- Task 3: *Express an Opinion*

The *Read a Dialogue Aloud* task requires three spoken responses. *Listen and Repeat* contains seven questions. *Express an Opinion* has only one question. Each response is scored

separately, on a raw score scale of either 0–4 (*Read a Dialogue Aloud*) or 0–5 (*Listen and Repeat/Express an Opinion*). Refer to Table 2 for the task types, the number of questions, and the score ranges in the Speaking assessment. Refer to Section II-3 for the details of each task type.

The Writing assessment consists of four task types:

- Task 1: *Build a Sentence*
- Task 2: *Write a Brief Email*
- Task 3: *Write an Extended Email*
- Task 4: *Write About Your Perspective*

Build a Sentence contains three sentences. *Write a Brief Email*, *Write an Extended Email*, and *Write About Your Perspective* each require one written response. Each response is scored separately on a raw score scale of 0–5 except for *Build a Sentence*, which is scored on a raw score scale of 0–2. Refer to Table 2 for the task types, the number of questions, and the score ranges in the Writing assessment. Refer to Section II-4 for the details of each task type.

Table 2. TOEIC Link Content Design for Speaking and Writing Assessments

Assessment	Task types	Number of tasks	Raw score range
Speaking	<i>Read a Dialogue Aloud</i>	3	0–4
	<i>Listen and Repeat</i>	7	0–5
	<i>Express an Opinion</i>	1	0–5
	Total	11	0–52
Writing	<i>Build a Sentence</i>	3	0–2
	<i>Write a Brief Email</i>	1	0–5
	<i>Write an Extended Email</i>	1	0–5
	<i>Write About Your Perspective</i>	1	0–5
	Total	6	0–21

IV. Technical Summary

IV-1. Listening and Reading Scoring

As noted previously, the TOEIC Link Listening and Reading assessments follow an MSA design with two stages in each measure. Every question presented to the test taker in both assessments contributes to the final score. In both the Listening and Reading assessments, the responses to questions are evaluated as either correct or incorrect, with 1 score point awarded for each correct answer.

The total score points that a test taker earns in each assessment (i.e., Listening and Reading), known as the raw score, are converted to a reporting scale score through a statistical process called equating. IRT methodology is used for score equating in the TOEIC Link Listening and Reading assessments. All items are calibrated, and the IRT parameters (difficulty and discrimination) are placed on the same scale based on previously collected data for these item types. For Listening and Reading, test forms are assembled by selecting items from the item pool according to the content and statistical specifications of the TOEIC Link MSA design. Since the IRT parameters for difficulty and discrimination of all items have already been estimated and placed on a common scale, IRT true score equating is used to produce the raw-to-scale conversion tables for each assembled test form.

In summary, the raw-to-scale conversions are first established for a base form, which is the form chosen to set up the common scale for reporting scale scores. The base form is used as a reference form to ensure that all reported scores from different forms are comparable. For each new test form, an IRT true score equating is conducted to convert the raw scores on the new form to equated raw scores, which correspond to the raw scores on the base form. These equated raw scores are then converted to scaled scores by applying the raw-to-scale conversion of the base form.

This equating process helps to ensure fairness and score comparability in two important ways. First, the equated score for both Listening and Reading adjusts for any differences in difficulty caused by the MSA. Second, it accounts for slight variations in difficulty across different test forms. Consequently, for both Listening and Reading assessments, a reported score reflects the same level of language ability, regardless of which second stage was taken, or which test form was administered.

IV-2. Speaking and Writing Scoring

Both the Speaking and Writing assessments contain tasks that require test takers to produce responses, which are scored using ETS's automated scoring engines (except for the *Build a Sentence* task in the Writing assessment, which is scored correct or incorrect). A correct response gets a score of 2, a wrong response gets a score of 1, and a non-response gets a score of 0. These automated scoring models are trained on human ratings. The total score points from all questions in the Speaking or Writing assessment are transformed into scaled scores. This section focuses on the use of speaking and writing scoring engines that leverage advanced natural language processing techniques to evaluate responses.

The Automated Scoring Approach of Speaking and Writing

The ETS proprietary Writing and Speaking automated scoring engines for TOEIC Link integrate advanced natural language processing techniques, combining research with extensive operational expertise for enhanced performance. ETS builds automated scoring engines through an iterative process of response data modelling and rigorous evaluation of system performance. These models are regularly refined to maintain a secure, precise, and up-to-date scoring system.

Scoring models are built in the same fashion for both the speaking and the writing items. The scoring model is trained using supervised learning, where it learns to map these features to human-assigned scores. This training allows the model to make consistent, accurate assessments of writing quality by recognizing patterns in the features that correspond to all predefined scores. The model is then rigorously tested against established standards, using a variety of cutting-edge analytical methods to assess overall performance, with particular attention to fairness and accuracy for all test takers and subgroups. If a model does not meet the required standards, it undergoes refinement, retraining, and further optimization to enhance its precision.

The Speaking scoring engine ensures a secure and up-to-date environment for automated scoring. The system is based on a scoring engine with distinct scoring models for each task type. The scoring models are trained on previously collected item responses that received at least one human score (used as criterion score). The automated scoring engine for speaking tasks evaluates responses by analyzing key speech features that indicate speech fluency (e.g., words spoken per minute), intelligibility (e.g., pronunciation accuracy), grammatical accuracy (e.g., correct phrases or sentences), and coherence (e.g., discourse transition cues). Tables 3 and 4 provide an overview of the main construct areas (based on the scoring rubrics) for the *Read a Dialogue Aloud/Listen*

and Repeat item types and the *Express an Opinion* item type, respectively. For each construct area, a few examples of speech features or feature categories that are used by the scoring engine are provided.

Table 3. *Read a Dialogue Aloud and Listen and Repeat* Task Types

Scoring dimension	Feature examples
Fluency	<ul style="list-style-type: none"> • speaking rate • length of uninterrupted “runs” (word sequences without pauses) • number of pauses • number of hesitations
Intelligibility	<ul style="list-style-type: none"> • correctness of pronunciation • naturalness of speech rhythm • naturalness of prosody (e.g., syllable stress)
Accuracy	<ul style="list-style-type: none"> • correctly read/repeated words • similarity to prompt

Table 4. *Express an Opinion* Task Type

Scoring dimension	Feature examples
Fluency	<ul style="list-style-type: none"> • speaking rate • length of uninterrupted “runs” (word sequences without pauses) • number of pauses • number of hesitations
Intelligibility	<ul style="list-style-type: none"> • correctness of pronunciation • naturalness of speech rhythm • naturalness of prosody (e.g., syllable stress)
Language use (vocabulary and grammar)	<ul style="list-style-type: none"> • vocabulary diversity • vocabulary richness • grammaticality • grammatical accuracy (few grammar errors)
Organization	<ul style="list-style-type: none"> • discourse coherence • use of discourse connectives

The Writing scoring engine is designed to handle various item types through task-specific models tailored to assess different dimensions of writing, ensuring a comprehensive evaluation. Similar to the scoring of spoken responses, the scoring process for written responses is underpinned by a detailed mapping of writing features computed by the scoring engine. Tables 5 and 6 provide feature examples for construct areas of the different Writing task types.

Table 5. *Write a Brief Email and Write an Extended Email Task Types*

Scoring dimension	Feature examples
Content	<ul style="list-style-type: none"> • number of sentences • discourse coherence • similarity to item prompt
Syntactic/lexical variety	<ul style="list-style-type: none"> • sentence variety • word frequency • correctness of collocations
Social conventions	<ul style="list-style-type: none"> • use of politeness indicators (e.g., modals, hedge words)
Accuracy/errors	<ul style="list-style-type: none"> • grammaticality • grammatical errors • word or usage errors • mechanical errors (e.g., spelling or interpunctuation errors)

Table 6. *Write About Your Perspective Task Type*

Scoring dimension	Feature examples
Content	<ul style="list-style-type: none"> • number of sentences • discourse coherence • similarity to item prompt
Syntactic/lexical variety	<ul style="list-style-type: none"> • sentence variety • word frequency • correctness of collocations
Accuracy/errors	<ul style="list-style-type: none"> • grammaticality • grammatical errors • word or usage errors • mechanical errors (e.g., spelling or interpunctuation errors) • collocation errors

IV-3. Evaluation of Machine Scores for Speaking and Writing

Ensuring the prediction accuracy of automated machine scoring in Speaking and Writing assessments is crucial for maintaining the reliability of the scores. To evaluate this accuracy, a random sample of responses to each Speaking and Writing task was scored by two human raters. This allows for the evaluation of the engines' performance in relation to the human raters' scores. Table 7 shows the correlation (Pearson r) between the automated score and the first human rating (i.e., machine–human) and between human ratings (i.e., human–human) for single responses for Speaking and Writing assessments. The human–machine correlations for Writing and Speaking range from 0.88–0.89, suggesting a strong agreement between human and machine scores.

Table 7. Correlation of Speaking and Writing Assessments by Scoring Method

Assessment	Human–machine correlation	Human–human correlation
Speaking	0.89	0.96
Writing	0.88	0.90

Note. The *Build a Sentence* writing task does not involve human or machine scoring.

V. Score Reporting

V-1. Scale Scores and Ranges

Each of the four TOEIC Link assessments has a reporting scale from 0 to 25 with an increment of 1. The overall test score is derived by averaging the individual assessment scores and also ranges from 0 to 25, increasing in increments of 1.

V-2. Development of Scoring Guides for Speaking and Writing

Separate scoring guides—known as rubrics—were developed for each task type to reflect the fact that each task makes specific demands on the test taker and elicits different evidence of language ability. Scoring rubrics for Speaking and Writing are provided in Appendices A and B, respectively.

V-3. The Common European Framework of Reference Languages

Each of the four TOEIC Link assessments was designed to measure test takers' English proficiency from A1 to C1 levels on the Common European Framework of References (CEFR; Council of Europe, 2001). The TOEIC Link assessments report scale scores and CEFR levels for any single assessment and any combinations of the four assessments. The scale scores and CEFR levels are on the same scale regardless of which test forms are taken.

A standard setting study was conducted to map Listening, Reading, Speaking, and Writing scale scores on the TOEIC Link to the A2, B1, B2, and C1 levels of the CEFR (Hille et al., in press). The study utilized a panel of 14 experts located in five countries (France, Italy, Türkiye, Mexico, and the United States) and took place online over a span of three full days. The first day, the panelists set cut scores for Listening and Reading using the Bookmark method (Reckase, 2023). The Bookmark method was selected in part because it is based on IRT, which is also used for the calibration, equating, and scoring of the TOEIC Link Listening and Reading assessments. On the other two days, the panelists set cut scores for Speaking and Writing using the Expected Task Score method (Plake & Cizek, 2012), which is a variation of the Angoff method that is suitable for tests with constructed response items that are polytomously scored using a rubric.

The study produced a total of 16 cut scores: minimum scale scores for CEFR levels A2, B1, B2, and C1, for each of the four assessments (Listening, Reading, Speaking, and Writing).

The standard error of judgment for the cut scores—in scale score units—ranged from 0.22 to 0.38 for Listening, 0.13 to 0.43 for Reading, 0.27 to 0.44 for Speaking, and 0.23 to 0.30 for Writing, indicating high levels of agreement among panelists. Table 8 presents the final recommended cut scores from the standard setting study mapping the TOEIC Link scores to CEFR levels. These cut scores may need to be revisited when operational data for the TOEIC Link assessments is available.

Table 8. Mapping TOEIC Link Scores to CEFR Levels

Assessment	Minimum scale score for A2	Minimum scale score for B1	Minimum scale score for B2	Minimum scale score for C1
Listening	8	13	18	22
Speaking	8	13	18	22
Reading	8	13	18	22
Writing	8	13	18	22

V-4. Reporting of Descriptors

Descriptors were developed to help stakeholders better understand the meaning of test scores. The research team that conducted the CEFR mapping study utilized information from the study (i.e., descriptors from relevant CEFR scales and descriptors of “just qualified candidates” created by panelists) to summarize the knowledge, skills, and abilities associated with test takers in scale score ranges that corresponded to each CEFR level. The relevant set of descriptors associated with each test score is provided on a test taker’s score report, and the full set of descriptors is summarized in a separate document available to the public at <https://www.etsglobal.org>.

VI. Score Reliability and Standard Error of Measurement

A critical aspect of any assessment's quality is the reliability of its scores. Reliability indicates the replicability of the test scores across different conditions of administration and/or administration of alternate forms of a test.

In the real world, there is no such thing as a perfectly reliable test score. Test results are always influenced to some degree by factors that have nothing to do with the targeted proficiency construct. For example, a test taker unusually tired or distracted on testing day may perform below their true level of language proficiency. Or, a test taker may have greater familiarity with the content of questions on one test form than on another test form or simply guess more questions on one occasion than on another. Such irrelevant factors contribute to what is called measurement error, which in turn determines how reliable test scores are. The more reliable scores are, the smaller the amount of measurement error is.

In essence, “the concern of reliability is to quantify the precision of test scores and other measurements” (Haertel, 2006, p. 65). Because tests are imperfect, a person's “real” or “true” language proficiency can never be perfectly measured on a test. The observed test score is instead a composite of a true score component and a measurement error component. A well-developed test is expected to yield scores that reflect the test takers' real proficiency as much as possible and minimize measurement error.

Since a person's true score is never obtainable, the best we can do is to estimate from the observed score using statistical methods. One way that the precision of test scores can be expressed is with a statistical index called a reliability coefficient, where values can range from 0 (*not at all reliable*) to 1 (*perfectly reliable*). Reliability coefficients are estimated in different ways depending on their intended use and the underlying theoretical framework of the assessment. High reliability is considered a prerequisite for drawing useful inferences from test scores.

Another statistical index used to express the precision of test scores is the standard error of measurement (SEM). SEM is an estimate of the average difference between true scores and observed test scores. To illustrate SEM, suppose that a test taker could take a large number of repeated tests designed to the exact same specifications. This test taker would receive many “observed” test scores, but because these observed test scores always contain some measurement error, none of them would be their true score. This is the case for any reported test score: We can

never be certain of a given test taker’s true language proficiency score. However, using an observed score together with SEM, it is possible to estimate a range above and below the observed score and the chance (typically 68% or 95%) that the true score may fall within this range. Generally speaking, 1 SEM indicates a 68% chance and 2 SEMs indicate a 95% chance. Two SEMs are most often used in practice.

To illustrate, if the test has a score range of 0 to 25 and 1 SEM equals 2 score points, and if a test taker receives a score of 16 on the test, we now know with 95% certainty that this person’s true score lies somewhere between 12 and 20 (16 plus or minus 2 SEMs [4 score points]). Similarly, if 1 SEM equals 1 score point, the range would be narrower—we could say with 95% certainty that the test taker’s true score lies between 14 and 18 (16 plus or minus 2 SEMs [2 score points]). The smaller the value of SEM, the higher the quality of measurement and the more precise the test scores will be.

Table 9 presents the assessment and total-score reliability estimates and SEM based on data from a typical TOEIC Link test form. Reliability estimation for the Reading and Listening sections of the TOEIC Link assessment is carried out using a method based on IRT (Kolen et al., 1996). For the Speaking assessment, reliability was estimated using an index known as coefficient alpha (Cronbach, 1951). Cronbach’s alpha is a way of assessing internal consistency reliability, which indicates the consistency of test takers’ responses across the questions, whether the questions are measuring the same trait that is intended to measure. For the Writing assessment, reliability was estimated using stratified coefficient alpha (Rajaratnam et al., 1965), which is also a measure of internal consistency reliability but is more accurate than regular coefficient alpha when subsets of items measure distinct content categories. In calculating the stratified coefficient alpha for the Writing assessment, the *Build a Sentence* item type is treated as one content category; the remaining item types are grouped into a second content category.

Table 9. Reliability Estimates and Standard Error of Measurement

Assessment	Scale	Reliability estimate	SEM
Listening	0–25	0.83	1.98
Speaking	0–25	0.88	0.93
Reading	0–25	0.83	1.82
Writing	0–25	0.85	1.43

Note. SEM = standard error of measurement.

VII. Validity and Fairness

VII-1. Validity

Validity refers to the extent to which theory and evidence support the intended interpretation and uses of test scores (AERA et al., 2014; ETS, 2014). The construct definition of a test establishes what the test intends to measure. Typically, validity is supported through the collection of evidence from the test development process and subsequent research that shows (a) how test content aligns with the construct (content validity), (b) whether items function as expected within the construct framework (internal structure), (c) how test scores correlate with related outcomes or external measures (criterion-related validity), (d) whether test takers' cognitive processes reflect engagement of the targeted skill (response processes), and (e) the extent to which the outcomes of test use are beneficial (consequential validity).

The TOEIC Link Listening and Reading assessments were developed based on the TOEIC Listening and Reading test and share the same underlying construct framework. To the extent that the testing populations of the TOEIC and TOEIC Link assessments are similar in characteristics that influence test performance, such as proficiency level or first language, existing validity evidence that supports claims about the meaningfulness of TOEIC Listening and Reading test scores may be relevant to support claims about scores on the TOEIC Link Listening and Reading assessments. For example, a criterion-related validity study using Japanese and Korean test-taker self-assessments showed that higher performing TOEIC Listening and Reading test takers were much more likely than lower scoring examinees to report they could perform a variety of practical English language tasks (Powers et al., 2010). To the extent that the characteristics test-taker self-assessments in the TOEIC Link population are comparable, this existing validity evidence for TOEIC Listening and Reading test scores may be relevant. Other studies have examined the internal structure of the TOEIC Listening and Reading test using factor analytic methods, concluding that the test sections (Listening and Reading) measure distinct but related skills (In'nami & Koizumi, 2012; Yoo & Manna, 2015).

The TOEIC Link Speaking and Writing assessments were developed by selecting tasks from ETS's library of existing test tasks. Consequently, existing evidence regarding the validity of these tasks may provide some initial validity evidence that the tasks measure the constructs they were designed to measure and/or predict (e.g., Davis & Norris, 2021; Davis et al., 2023; Papageorgiou et al., 2021). In addition, Tables 10 and 11 show how dimensions of speaking and

writing proficiency have been operationalized in TOEIC Link automated scoring models and human scoring rubrics (for human scoring rubrics, see Appendices A and B). This operationalization supports an expectation that automated scoring models should evaluate similar dimensions as human raters.

Table 10. Dimensions of Speaking Proficiency as Operationalized in TOEIC Link Speaking Assessment Automated and Human Scoring

Dimension	Automated scoring feature examples	Human scoring rubric examples
Fluency	<ul style="list-style-type: none"> speaking rate length of uninterrupted “runs” (word sequences without pauses) number of pauses number of hesitations 	<p>“...fluid, with little hesitation or self-correction, and the rate of speech is appropriate” (<i>Read a Dialogue Aloud</i>, rubric score 4)</p> <p>“...may be long pauses and frequent hesitations” (<i>Express an Opinion</i>, rubric score 2)</p>
Intelligibility	<ul style="list-style-type: none"> correctness of pronunciation naturalness of speech rhythm naturalness of prosody (e.g., syllable stress) 	<p>“...intelligibility is repeatedly affected by inaccuracies in pronunciation or intonation” (<i>Read a Dialogue Aloud</i>, rubric score 2)</p> <p>“...words may be ambiguous because of imprecise pronunciation” (<i>Listen & Repeat</i>, rubric score 4)</p>
Content accuracy	<ul style="list-style-type: none"> correctly read/repeated words similarity to prompt 	<p>“...text is fully and accurately reproduced” (<i>Read a Dialogue Aloud</i>, rubric score 4)</p> <p>“...response contains a majority of the content words or ideas in the prompt” (<i>Listen & Repeat</i>, rubric score 3)</p>
Language use	<ul style="list-style-type: none"> vocabulary diversity vocabulary richness grammaticality grammatical accuracy (few grammar errors) 	<p>“...demonstrates automated and effective use of grammar...use of vocabulary is effective” (<i>Express an Opinion</i>, rubric score 5)</p>
Organization	<ul style="list-style-type: none"> discourse coherence use of discourse connectives 	<p>“...response is somewhat developed; relationships between ideas are mostly clear” (<i>Express an Opinion</i>, rubric score 4)</p>

Table 11. Dimensions of Writing Proficiency as Operationalized in Automated and Human Scoring

Dimension	Automated scoring feature examples	Human scoring rubric examples
Content	<ul style="list-style-type: none"> • number of sentences • discourse coherence • similarity to item prompt 	<p>“Adequate elaboration to support the communicative purpose” (<i>Extended Email</i>, rubric score 4)</p> <p>“Ideas that may be poorly elaborated or only partially relevant” (<i>Write About Your Perspective</i>, rubric score 2)</p>
Syntactic/lexical variety	<ul style="list-style-type: none"> • sentence variety • word frequency • correctness of collocations 	<p>“A limited range of grammar and vocabulary is used” (<i>Brief Email</i>, rubric score 2)</p> <p>“Effective syntactic variety and precise, idiomatic word choice” (<i>Extended Email</i>, rubric score 5)</p>
Grammatical/lexical accuracy	<ul style="list-style-type: none"> • grammaticality • grammatical errors • word or usage errors • mechanical errors (e.g., spelling or punctuation errors) • collocation errors 	<p>“Limitations in...accuracy, or mechanics may affect the readability or clarity of part or all of the response” (<i>Brief Email</i>, rubric score 3)</p>
Social conventions	<ul style="list-style-type: none"> • use of politeness indicators (e.g., modals, hedge words) 	<p>“Appropriate social conventions (e.g., politeness, organization of information, and formulation of actions such as requests, etc.) are consistency used” (<i>Brief Email</i>, rubric score 5)</p>

As the TOEIC Link Assessment is administered under operational conditions, new validity evidence will be collected to support a validity argument.

VII-2. Fairness

Fairness is a central component of all ETS products and services. All materials undergo rigorous reviews for fairness by trained staff who apply ETS standards for quality and fairness (ETS, 2014), in compliance with *Standards for Educational and Psychological Testing* (AERA et al., 2014).

Throughout all stages of design, development, and delivery, the TOEIC Link testing program implements quality control measures to ensure the test and test scores are fair, or, in other words, equally valid for all test takers, regardless of nationality, age, or gender.

Given the points made in the section previously about the comparability between the TOEIC and TOEIC Link Listening and Reading assessments, research evidence pertaining to fairness for the TOEIC Listening and Reading test may be relevant. For example, Yoo and Manna (2015) found that the underlying structure of the construct measured by the TOEIC Listening and Reading test was the same across subgroups, including gender, age, employment status, time spent studying English, and having lived in a country where English is the main language. In another research study, Yoo et al. (2018) found that the comparability and meaning of test scores was similar across these five subgroups. More recently, Schmidgall et al. (2024) found that the TOEIC Listening and Reading test did not unfairly advantage full-time employees or full-time students.

As the TOEIC Link assessment is administered under operational conditions, new evidence regarding fairness will accumulate and will be needed to investigate assumptions and support claims in a validity argument.

VIII. Assessment Administration and Security

VIII-1. Assessment Display Sequence

Test takers taking the four-skill module will receive the Listening assessment first, followed by the Speaking assessment, the Reading assessment, and the Writing assessment. The order keeps skills with audio input together. The same order is used regardless of which skills are tested. Table 12 provides a summary of the assessment display sequence for the four-skill module assessment.

Table 12. An Overview of a TOEIC Link Four-Skill Module Assessment

Assessment	Display sequence	Number of stages	Task types	Number of items
Listening	1	2	<i>Photos; Question Response; Conversations; Talks</i>	30
Speaking	2	1	<i>Read a Dialogue Aloud; Listen and Repeat; Express an Opinion</i>	11
Reading	3	2	<i>Incomplete Sentence; Text Completions; Reading Passages</i>	30
Writing	4	1	<i>Build a Sentence; Write a Brief E-mail; Write an Extended E-mail, Write About Your Perspective</i>	6
Total				77

TOEIC Link assessment administrations may be in person or home-based. In-person test sessions may be administered at the client’s location or at the local ETS Preferred Network office. The assessments may be proctored by ETS Preferred Network staff or trained client or third-party proctors. Live proctor and automated remote proctoring options are available for assessments delivered through the test taker’s personal computer at home. Both in-person and home-based TOEIC Link assessments are delivered through a secure exam browser (SEB) that includes security functions to minimize the opportunity to disclose assessment content and prevent unauthorized assistance.

VIII-2. Secure Exam Browser

At the beginning of each assessment administration, SEB checks for applications that are not part of the TOEIC Link assessment administration and ensures that the screen is not being

shared remotely. If an unauthorized application is running or the screen is being shared, the SEB will display a notification to inform the test taker of corrective steps that they must take in order to proceed to the assessment. During the assessment, the SEB locks down the device to prevent test takers from switching to other applications. It also prevents test takers from using short cut keys to cut, copy, and paste text in the Writing response areas, and it prevents them from copying assessment content and transferring it to another application.

VIII-3. Remote Proctoring Options

Two remote proctoring options are available for home-based TOEIC Link assessments: AI proctoring and live remote proctoring. The AI proctoring option provides automated proctoring during the assessment. AI-driven technology is used to flag test takers when an irregularity is detected during the assessment session. Each test taker receives a proctoring score that can be used to identify cases that may require additional review or score cancellation.

The live remote proctoring option enables a proctor to log into the remote proctoring software and monitor test takers in real time. The remote proctoring software contains functions such as an ID check, room scan, and flagging capability to assist proctors with test-taker check-in and monitoring during the assessment administration.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Ashmore, E., Duke, T., & Sakano, J. (2018). Background and goals of the TOEIC® Listening and Reading update project. In D. Powers & J. Schmidgall (Eds.), *The research foundation for the TOEIC® tests: A compendium of studies* (Vol. 3). ETS.
<https://www.ets.org/Media/Research/pdf/TC-10-02.pdf>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>
- Davis, L., & Norris, J. (2021). *Developing an innovative elicited imitation task for efficient English proficiency assessment* (TOEFL Research Report No. RR-96). ETS.
<https://doi.org/10.1002/ets2.12338>
- Davis, L., Norris, J., Papageorgiou, S., & Sasayama, S. (2023). Balancing construct coverage and efficiency: Test design, security, and validation considerations for a remotely-proctored online language test. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 49–63). Routledge.
- ETS. (2014). *ETS standards for quality and fairness*. Author.
<https://www.ets.org/pdfs/about/standards-quality-fairness.pdf>
- ETS. (2022). *TOEIC® Listening and Reading test: Examinee handbook*. Author.
<https://www.ets.org/pdfs/toeic/toeic-listening-reading-test-examinee-handbook.pdf>
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Westport, CT: Praeger Publishers.
- Hille, K., Schmidgall, J., Oh, S., & Ohta, R. (in press). *Mapping TOEIC® Link™ scores to the Common European Framework of Reference for Languages (CEFR)* (Research Memorandum). ETS.
- In'nami, Y., & Koizumi, R. (2012). Factor structure of the revised TOEIC® test: A multi-sample analysis. *Language Testing*, 29, 131–152. <https://doi.org/10.1177/0265532211413444>

- Kolen, M. J., Zeng, L., & Hanson, B. A. (1996). Conditional standard errors of measurement for scale scores using IRT. *Journal of Educational Measurement*, 33, 129–140.
<https://doi.org/10.1111/j.1745-3984.1996.tb00485.x>
- Papageorgiou, S., Davis, L., Norris, J. M., Garcia Gomez, P., Manna, V. F., & Monfils, L. (2021). *Design framework for the TOEFL Essentials® test 2021* (Research Memorandum No. RM-21-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-21-03.pdf>
- Plake, B. S., & Cizek, G. J. (2012). Variations on a theme: The Modified Angoff, Extended Angoff, and Yes/No standard setting methods. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (pp. 181–100). Routledge.
- Powers, D. E., Kim, H., & Weng, V. Z. (2010). The redesigned TOEIC® Listening and Reading test: Relations to test taker perceptions of proficiency in English. In D. Powers (Ed.), *The research foundation for the TOEIC® tests: A compendium of studies*. ETS.
<https://doi.org/10.1002/j.2333-8504.2008.tb02142.x>
- Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika*, 30, 39–56.
- Reckase, M. D. (2023). *The psychometrics of standard setting: Connecting policy and test scores*. CRC Press.
- Schedl, M. (2010). Background and goals of the TOEIC® Listening and Reading test redesign project. In D. Powers (Ed.), *The research foundation for the TOEIC® tests: A compendium of studies*. ETS. <http://www.ets.org/Media/Research/pdf/TC-10-02.pdf>
- Schmidgall, J., Huo, Y., Cid, J., & Wei, Y. (2024). *Investigating fairness claims for a general-purposes assessment of English proficiency for the international workplace: Do full-time employees have an unfair advantage over full-time students?* (Research Report No. RR-24-06). ETS. <https://doi.org/10.1002/ets2.12380>
- Yoo, H., & Manna, V. F. (2015). Measuring English language workplace proficiency across subgroups: Using CFA models to validate test score interpretation. *Language Testing*, 1–26. <https://doi.org/10.1177/0265532215618987>
- Yoo, H., Manna, V. F., Monfils, L. F., & Oh, H.-J. (2018). Measuring English language proficiency across subgroups: Using score equity assessment to evaluate test fairness. *Language Testing*, 36(2), 289–309. <https://doi.org/10.1177/0265532218776040>

Zenisky, A. L., Hambleton, R. K., & Luecht, R. M. (2010). Multistage testing: Issues, designs, and research. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 355–372). Springer. https://doi.org/10.1007/978-0-387-85461-8_18

Appendix A: Speaking Assessment Human Scoring Rubrics

Scoring Guide: Read a Dialogue Aloud	
4	<p>The passage is read with ease; speech is fluid and intelligible; units of meaning are clearly marked.</p> <p>A typical response exhibits all of the following.</p> <ul style="list-style-type: none"> • Reading is fluid, with little hesitation or self-correction, and the rate of speech is appropriate. • Intonation and pausing are used to group words in meaningful phrases. • Speech is highly intelligible; minor mispronunciations or other-language influences may be present that do not impair intelligibility. • The text is fully and accurately reproduced.
3	<p>The passage is read with little difficulty; it may require minor listener effort.</p> <p>A typical response exhibits most of the following.</p> <ul style="list-style-type: none"> • Pacing is mostly even but with minor hesitation/choppiness. • Intonation may be monotone at times, but words are mostly grouped in meaningful phrases. • Intelligibility is generally sustained, although there may be occasional minor lapses. • Occasional mispronunciations may require minor listener effort. • The text is fully reproduced, with no more than minor changes.
2	<p>The passage is read with some difficulty; the content is partially clear.</p> <p>A typical response exhibits one or more of the following.</p> <ul style="list-style-type: none"> • Pacing is uneven or slow; frequent or inappropriate hesitations cause noticeable choppiness; sentence-level meanings may be hard to follow. • Intelligibility is repeatedly affected by inaccuracies in pronunciation or intonation. • Intonation is monotone or inappropriate for the meaning; there may be partially effective grouping of words into meaningful phrases. • The text is mostly reproduced; variations (substitutions, omissions) may be present.
1	<p>The passage is read with noticeable difficulty; the content may be mostly unintelligible.</p> <p>A typical response exhibits one or more of the following.</p> <ul style="list-style-type: none"> • The speaking rate may be very slow, or the text is mostly read in short chunks, or multiple or extended pauses result in speech that is largely staccato. • The response is mostly unintelligible; only isolated words or phrases from the text are understandable. • The text is substantially incomplete, or the response consists of isolated phrases; variations (substitutions, omissions) may be common.
0	<p>No response OR the response is entirely unintelligible OR there is no English in the response OR the test taker does not read the text provided.</p>

Scoring Guide: Listen and Repeat	
5	<p>The response exactly repeats the prompt.</p> <p>A typical response exhibits the following.</p> <ul style="list-style-type: none"> • The response is fully intelligible and is an exact repetition of the prompt.
4	<p>The response captures the meaning expressed in the prompt, but it is not an exact repetition.</p> <p>A typical response exhibits the following.</p> <ul style="list-style-type: none"> • Minor changes in words or grammar are present that do not substantially change the meaning of the prompt. For example: <ul style="list-style-type: none"> ○ one or two function words may be missing or changed, ○ a content word may be missing (in longer stimuli) or replaced with a related word, ○ markers of tense/aspect/number may be missing or incorrect, or ○ two words may be transposed. • One or two content words may be ambiguous because of imprecise pronunciation. The speaker may self-correct but successfully completes the response.
3	<p>The response is essentially full, but it does not accurately capture the original meaning.</p> <p>A typical response exhibits the following.</p> <ul style="list-style-type: none"> • The response contains a majority of the content words or ideas in the prompt. <ul style="list-style-type: none"> ○ Multiple function words may be changed or missing; one or more content words may be missing or substantively changed. • The response is a full sentence. • In some cases, intelligibility issues cause occasional difficulty in understanding meaning. The speaker may struggle over a word or phrase or run words together, reducing intelligibility.
2	<p>The response is missing a significant part of the prompt and/or is highly inaccurate.</p> <p>A typical response exhibits one or more of the following.</p> <ul style="list-style-type: none"> • A large portion of the prompt is missing, and important original meaning is left out. <ul style="list-style-type: none"> ○ The speaker may repeat the first part of the sentence. Then the speaker may stop or fill with inaccurate content and/or include the last few words. • The response is not a self-standing sentence; meaning is fragmentary. • Intelligibility is low; the response would be difficult to understand for a listener unfamiliar with the prompt.
1	<p>The response captures very little of the prompt or is largely unintelligible.</p> <p>A typical response exhibits one or more of the following.</p> <ul style="list-style-type: none"> • A minimal response of a few words is made; most of the prompt is missing. • The response is recognizable as an attempt to repeat the prompt, but it is mostly unintelligible.
0	<p>No response OR the response is entirely unintelligible OR there is no English in the response OR the content is entirely unconnected to the prompt (or consists only of phrases such as “I don’t know”).</p>

Scoring Guide: Express an Opinion	
5	The response fulfills the demands of the task and is readily intelligible, sustained, and coherent.
	<p>It is characterized by ALL of the following:</p> <ul style="list-style-type: none"> • The response is generally well developed; relationships between ideas are clear. • The speech is clear with a generally well-paced flow. It may include minor lapses or minor difficulties with pronunciation or intonation patterns that do not affect overall intelligibility. • The response demonstrates automatic and effective use of grammar, with good control of basic and complex structures. Some minor errors may be noticeable, but they do not obscure meaning. • The use of vocabulary is effective, with allowance for occasional minor inaccuracy.
4	The response addresses the task and is generally intelligible, sustained, and coherent.
	<ul style="list-style-type: none"> • The response is somewhat developed; relationships between ideas are mostly clear, with occasional lapses. • Minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times, although overall intelligibility is not significantly affected. • The response demonstrates fairly automatic and effective use of grammar but may be somewhat limited in the range of structures used. • The use of vocabulary is fairly effective. Some vocabulary may be inaccurate or imprecise.
3	The response addresses the task and is basically intelligible, but development is limited.
	<ul style="list-style-type: none"> • The response provides some development. However, it provides little elaboration, repeats itself with no new information, is vague, or is unclear. Relationships between ideas may be unclear at times. • The speech is basically intelligible, though listener effort may be needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places. • The response demonstrates limited control of grammar; for the most part, only basic sentence structures are used successfully. • The use of vocabulary is limited.
2	The response attempts to address the task but has significant limitations. Parts of the response may not be sustained or may be unintelligible or incoherent.
	<ul style="list-style-type: none"> • The response expresses only basic ideas and is vague or repetitious. Relationships between ideas may be unclear. • Consistent difficulties with pronunciation, stress, and intonation cause considerable listener effort; delivery is choppy, fragmented, or telegraphic; there may be long pauses and frequent hesitations. • Control of grammar severely limits the expression of ideas and clarity of connections among ideas. • The use of vocabulary is severely limited or highly repetitious.
1	The response is limited to reading the prompt or the directions aloud OR the response consists of isolated words or phrases, or mixtures of the first language and English.
0	No response OR no English in the response.

Appendix B: Writing Assessment Human Scoring Rubrics

Scoring Guide: Write a Brief Email	
5	<p>A fully successful response The response is well elaborated and shows consistent facility in language use.</p>
	<p>A typical response shows all of the following.</p> <ul style="list-style-type: none"> • The e-mail accomplishes the task and is fully elaborated with relevant supporting details. • A range of grammar and vocabulary is used effectively (e.g., to create a degree of fluidity, precision, idiomaticity, or expressiveness). • Language use is accurate, although minor errors or nonidiomatic uses may occur. However, the intended meaning is fully clear and cohesive throughout. • Appropriate social conventions (e.g., politeness, organization of information, and formulation of actions such as requests, etc.) are consistently used.
4	<p>A generally successful response The response is elaborated, although minor inconsistencies in language use are present.</p>
	<p>A typical response shows most of the following.</p> <ul style="list-style-type: none"> • The e-mail accomplishes the task and is supported with suitable elaboration. • A range of grammar and vocabulary is used appropriately, but the range or effectiveness may not be sustained throughout the response. • Inaccuracies in language use may be noticeable and may cause minor ambiguities in sentence-level meanings, but the overall response is clear. • Mostly appropriate social conventions are used, but usage may not be consistent.
3	<p>A partially successful response The response shows some elaboration and/or facility in language use, but it is limited in one or more aspects.</p>
	<p>A typical response partially accomplishes the task, but it may show either of the following patterns.</p> <ul style="list-style-type: none"> • The response shows a reasonable degree of elaboration, with limitations elsewhere. <ul style="list-style-type: none"> ○ A somewhat limited range of grammar and vocabulary is used; attempts at more complex structures or vocabulary are not entirely successful. ○ Limitations in range, accuracy, or mechanics may affect the readability or clarity of part or all of the response. • The response shows limited elaboration, which reduces the effectiveness of the message. <ul style="list-style-type: none"> ○ An adequate range of grammar and vocabulary is used; limitations in language use do not greatly affect the overall clarity of the message.
2	<p>A mostly unsuccessful response The response is limited in content and in facility of language use.</p>
	<p>A typical response shows some or all of the following.</p> <ul style="list-style-type: none"> • Very limited elaboration is present, which significantly affects the effectiveness of the e-mail. • A limited range of grammar and vocabulary is used, although some attempt is made to produce sentence-level language beyond simple clauses or basic vocabulary. • Errors in language use or mechanics may cause the intended meaning to be unclear.
1	<p>An unsuccessful response The content of the response is very limited, with little evidence of ability to produce extended text.</p>
	<p>A typical response shows some or all of the following.</p> <ul style="list-style-type: none"> • Little or no elaboration is provided. • The e-mail consists of telegraphic language (i.e., short and/or disconnected phrases and simple sentences). • Serious errors in language use may be present.
0	<p>The response is blank, rejects the topic, is not in English, is entirely copied from the prompt, is entirely unconnected to the prompt, or consists of arbitrary keystrokes.</p>

Scoring Guide: Write an Extended Email	
5	<p>A fully successful response The response is effective, is clearly expressed, and shows consistent facility in the use of language.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Elaboration that effectively supports the communicative purpose • Effective syntactic variety and precise, idiomatic word choice • Consistent use of appropriate social conventions (e.g., politeness, register, organization of information, and formulation of actions such as requests, refusals, criticisms, etc.) • Almost no lexical or grammatical errors other than those expected from a competent writer writing under timed conditions (e.g., common typos or common misspellings or substitutions like <i>there/their</i>)
4	<p>A generally successful response The response is mostly effective and easily understood. Language facility is adequate to the task.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Adequate elaboration to support the communicative purpose • Syntactic variety and appropriate word choice • Mostly appropriate social conventions • Few lexical or grammatical errors
3	<p>A partially successful response The response generally accomplishes the task. Limitations in language facility may prevent parts of the message from being fully clear and effective.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Elaboration that partially supports the communicative purpose • A moderate range of syntax and vocabulary • Some noticeable errors in structure, word forms, use of idiomatic language, and/or social conventions
2	<p>A mostly unsuccessful response The response reflects an attempt to address the task, but it is mostly ineffective. The message may be limited or difficult to interpret.</p>
	<p>A typical response may display the following.</p> <ul style="list-style-type: none"> • Limited or irrelevant elaboration • Some connected sentence-level language, with a limited range of syntax and vocabulary • An accumulation of errors in sentence structure and/or language use
1	<p>An unsuccessful response The response reflects an ineffective attempt to address the task. The message may be limited to the point of being unintelligible.</p>
	<p>A typical response may display the following.</p> <ul style="list-style-type: none"> • Very little elaboration, if any • Telegraphic language (i.e., short and/or disconnected phrases and sentences) with a very limited range of vocabulary • Serious and frequent errors in the use of language • Minimal original language; any coherent language is mostly borrowed from the stimulus.
0	<p>The response is blank, rejects the topic, is not in English, is entirely copied from the prompt, is entirely unconnected to the prompt, or consists of arbitrary keystrokes.</p>

Scoring Guide: Write About Your Perspective	
5	<p>A fully successful response The response is a relevant and very clearly expressed contribution to the online discussion, and it demonstrates consistent facility in the use of language.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Relevant and well-elaborated explanations, exemplifications, and/or details • Effective use of a variety of syntactic structures and precise, idiomatic word choice • Almost no lexical or grammatical errors other than those expected from a competent writer writing under timed conditions (e.g., common typos or common misspellings or substitutions like <i>there/their</i>)
4	<p>A generally successful response The response is a relevant contribution to the online discussion, and facility in the use of language allows the writer’s ideas to be easily understood.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Relevant and adequately elaborated explanations, exemplifications, and/or details • A variety of syntactic structures and appropriate word choice • Few lexical or grammatical errors
3	<p>A partially successful response The response is a mostly relevant and mostly understandable contribution to the online discussion, and there is some facility in the use of language.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Elaboration in which part of an explanation, example, or detail may be missing, unclear, or irrelevant • Some variety in syntactic structures and a range of vocabulary • Some noticeable lexical and grammatical errors in sentence structure, word form, or use of idiomatic language
2	<p>A mostly unsuccessful response The response reflects an attempt to contribute to the online discussion, but limitations in the use of language may make ideas hard to follow.</p>
	<p>A typical response displays the following.</p> <ul style="list-style-type: none"> • Ideas that may be poorly elaborated or only partially relevant • A limited range of syntactic structures and vocabulary • An accumulation of errors in sentence structure, word forms, or use
1	<p>An unsuccessful response The response reflects an ineffective attempt to contribute to the online discussion, and limitations in the use of language may prevent the expression of ideas.</p>
	<p>A typical response may display the following.</p> <ul style="list-style-type: none"> • Words and phrases that indicate an attempt to address the task but with few or no coherent ideas • Severely limited range of syntactic structures and vocabulary • Serious and frequent errors in the use of language • Minimal original language; any coherent language is mostly borrowed from the stimulus
0	<p>The response is blank, rejects the topic, is not in English, is entirely copied from the prompt, is entirely unconnected to the prompt, or consists of arbitrary keystrokes.</p>

Suggested Citation

Cid, J., Schmidgall, J., & Park, E. (2026). *TOEIC® Link™ assessments technical manual* (Research Report No. RR-26-04). ETS. <https://doi.org/10.64634/7e2pzg04>

Action Editor: Larry Davis

Reviewers: Spiros Papageorgiou and Jiyun Zu

ETS, the ETS logo, TOEIC, and TOEIC ESSENTIALS are registered trademarks of Educational Testing Service (ETS). TOEIC LINK is a trademark of ETS. All other trademarks are property of their respective owners.

Cover image by Mitchell Luo, Pexels

Find other ETS-published reports by searching the [ETS ReSEARCHER](#) database.

