

RESEARCH MEMORANDUM

Alignment of the TOEIC Bridge® Tests With the Educational Functioning Level Descriptors for English as a Second Language

AUTHORS

Mikyung Kim Wolf and Renka Ohta

ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey
Lord Chair in Measurement and Statistics

ASSOCIATE EDITORS

Usama Ali
Senior Measurement Scientist

Beata Beigman Klebanov
Principal Research Scientist, 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

Zuowei 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.

**Alignment of the TOEIC Bridge® Tests With the Educational Functioning Level Descriptors for
English as a Second Language**

Mikyung Kim Wolf & Renka Ohta

ETS Research Institute, ETS, Princeton, New Jersey, United States

May 2026

Corresponding author: Mikyung Kim Wolf, E-mail: mkwolf@ets.org

Suggested citation: Wolf, M. K., & Ohta, R. (2026). *Alignment of the TOEIC Bridge® tests with the educational functioning level descriptors for English as a second language* (Research Memorandum No. RM-26-03). ETS.
<https://doi.org/10.64634/1n934x31>

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

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

Action Editor: Jonathan Schmidgall

Reviewers: Maurice Hauck and Spiros Papageorgiou

Copyright © 2026 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, TOEIC, and TOEIC BRIDGE are registered trademarks of Educational Testing Service (ETS).

All other trademarks are the property of their respective owners.

Abstract

This report presents the findings of an alignment study evaluating the extent to which the TOEIC Bridge® tests reflect the language knowledge, skills, and abilities described in the National Reporting System (NRS) educational functioning level descriptors for English as a second language (ESL). The study was intended to provide critical validity evidence supporting the use of these assessments in adult education programs serving ESL learners, particularly for purposes of NRS accountability. Results indicate strong evidence of content alignment across all test forms examined. The report details the research-based methodology and offers practical recommendations for the effective use of the TOEIC Bridge tests in the NRS.

Keywords: alignment, adult basic education, educational functioning level, EFL, English as a second language, ESL, National Reporting System, NRS, TOEIC®, TOEIC Bridge®

Introduction

The purpose of this study is to evaluate the extent to which the TOEIC Bridge® tests cover the language knowledge, skills, and abilities described in the National Reporting System (NRS) educational functioning level (EFL) descriptors for English as a second language (ESL; hereafter, the NRS EFL descriptors for ESL). See the U.S. Department of Education’s Division of Adult Education and Literacy (DAEL) *Technical Assistance Guide for Performance Accountability under the Workforce Innovation and Opportunity Act* (DAEL, 2025) for more on the NRS EFL descriptors for ESL. The study examines two specific forms of the TOEIC Bridge tests: Forms A and B. The study results are expected to provide critical evidence to support the use of these tests in adult education programs serving ESL learners, particularly for NRS accountability purposes.

An evaluation based on test content is a fundamental component of the validation process, conceptualized as alignment in the field of educational measurement. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014) defined alignment as follows: “the degree to which the content and cognitive demands of test questions match targeted content and cognitive demands described in the test specifications” (p. 216).

In educational settings, alignment evaluation typically focuses on the extent to which the test content represents the knowledge, skills, and abilities described in learning standards. By doing so, test results can provide accurate information about students' achievement relative to the expectations described in those learning standards. Thus, alignment serves as a key form of content-oriented evidence that supports the validity of score interpretations and the intended uses of an assessment.

In the context of the present study, alignment refers to evaluating the degree to which the content of the TOEIC Bridge tests appropriately reflects the domain set forward in the NRS EFL descriptors for ESL. A high degree of alignment will allow for valid inferences about test takers' abilities and proficiency levels relative to the NRS EFL descriptors for ESL. A team of ETS researchers conducted the present alignment study involving panels of external content experts. This report details the study methodology and findings. Based on the results, we offer a set of recommendations for assessment use and further research, along with key lessons learned.

Research Background on Alignment Methods

Previous research has suggested that there are varied approaches to evaluating alignment depending on the purpose, assessment domain, and artifacts compared as the basis for alignment (Cook, 2005; Porter, 2002; Resnick et al., 2004; Webb, 1997; Webb, 2005). Although the knowledge, skills, and abilities represented in test content are a common dimension of evaluation, their depth and breadth are analyzed differently across various alignment methods. In subject-matter assessments such as English language arts, mathematics, and science, depth of knowledge is conceptualized as cognitive complexity and is analyzed in different ways depending on the subject area and alignment method (e.g., four levels in Webb's method and five levels in Porter's method).

Unlike subject-matter assessments, language proficiency standards postulate trajectories of language development, often delineated through language proficiency-level descriptors. Hence, when evaluating the alignment of language assessments, linguistic difficulty or language complexity levels are identified in lieu of cognitive complexity (Bailey et al., 2022; Cook, 2005; Johnson, 2005; Office of Elementary and Secondary Education, 2018; Wolf et al.,

2023). The U.S. Department of Education’s Office of Elementary and Secondary Education (2018) guidance document for state assessments explicitly calls for an analysis of language demands or language complexity instead of cognitive complexity when examining the alignment of English language proficiency (ELP) assessments with standards. Language researchers have also proposed useful strategies and approaches for identifying language complexity, drawing on second language acquisition theories, the communicative ability model, and systemic functional linguistics (Cook, 2005; Cook & McDonald, 2013; Sato et al., 2008; Wolf et al., 2022). For instance, the language complexity of test items and tasks can be analyzed in terms of specific language functions (e.g., explain, inquire, retell) and linguistic forms at the lexical, syntactic, and discourse aspects.

Based on the research described previously, the present study is concerned with the specific language knowledge and skills as well as the language complexity reflected in the TOEIC Bridge tests in relation to the NRS EFL descriptors for ESL. In addition, Webb’s alignment method is employed to guide the primary alignment criteria (e.g., content match, coverage).

TOEIC Bridge Tests

Developed by ETS, the TOEIC Bridge tests measure the English proficiency of young adults and adults across any or all four communication skills—listening, reading, speaking, and writing—in everyday and workplace contexts (ETS, n.d.). The constructs and content of the TOEIC Bridge tests were developed based on a careful target language use domain analysis of communicative tasks in the daily and workplace lives of adults. This development was also grounded in theoretical frameworks such as evidence-centered design (Mislevy et al., 2003) and the model of communicative language ability (Bachman & Palmer, 2010). A series of publications is available that illustrates the conceptualization of the constructs and the specific task types intended to measure test takers’ language knowledge, skills, and abilities (e.g., see Schmidgall et al., 2019, for the TOEIC Bridge design framework; Everson et al., 2019, for the TOEIC Bridge task design, pilot, and field-testing results).

The intended uses of the TOEIC Bridge tests include measuring the English proficiency of adult English learners to inform decision such as placement and employment. In instructional settings, the tests can also be used to inform teaching and to monitor learners’ progress. Unlike

the TOEIC tests, which assess beginning to advanced proficiency levels, the TOEIC Bridge tests are geared toward beginning to intermediate proficiency levels. TOEIC Bridge test scores are mapped to language proficiency levels of the Common European Framework of Reference for Languages (CEFR), Pre-A1 to B1 (Schmidgall, 2021).

Although the TOEIC Bridge tests have numerous parallel forms for operational use over the decades, two test forms—Form A and Form B—were included in this study. Table 1 summarizes the test form structure.

Table 1. Structure of the TOEIC Bridge Tests for Each Form

Domain	Number of items/tasks	Item/Task format	Names of task types (test parts)
Listening	50	Multiple-choice	Four Pictures Question-Response Conversations Talks
Reading	50	Multiple-choice	Sentence Completion Text Completion Reading Comprehension
Speaking	8	Constructed-response	Read a Short Text Aloud Describe a Photograph Listen and Retell Short Interaction Tell a Story Make and Support a Recommendation
Writing	9	Constructed-response	Build a Sentence Write a Sentence Respond to a Brief Message Write Narrative Respond to an Extended Message

NRS Education Functioning Level Descriptors for ESL

The NRS EFL descriptors for ESL were revised in 2017 to align with the new *English Language Proficiency Standards for Adult Education* (hereafter referred to as the ELP Standards; Office of Career, Technical and Adult Education, 2016). This document also identifies the College and Career Readiness (CCR) anchor standards that correspond to each ELP standard.

The NRS EFL descriptors for ESL are organized into three modalities, following the structure of the new ELP Standards: interpretive, productive, and interactive. Moving away from the traditional organization by four discrete language skills (i.e., listening, reading,

speaking, writing), the new ELP Standards highlight the integrated use of language skills for communication (Council of Chief State School Officers, 2014; Wolf et al., 2022). For example, the overarching descriptor for the productive modality states that “the learner’s ability to produce level-appropriate written and spoken text such that it meaningfully transmits meaning” (DAEL, 2025, p. B-20), combining speaking and writing skills.

The NRS EFL descriptors for ESL also specify the ELP Standards to which they are aligned. The descriptors for the productive modality align with ELP Standards 3, 4, 7, 9, and 10; those for the interpretive modality align with ELP Standards 1, 6, 7, and 8; those for the Interactive modality align with ELP Standards 2 and 5 (Office of Career, Technical and Adult Education, 2016). Therefore, the alignment findings between the NRS EFL descriptors and the TOEIC Bridge tests in this study also imply alignment between the new ELP Standards and the tests.

Evaluating alignment with the NRS EFL descriptors for ESL should account for some critical structural and conceptual features of the descriptors. The introduction to the NRS EFL descriptor document explicitly states that “the descriptors do not provide a complete or comprehensive delineation of all of the skills at any given level but rather provide a description of the most critical concepts and skills for the level” (DAEL, 2025, p. B-20). Due to this nature, alignment judges or raters may often need to take into account underlying skills needed to perform the example tasks included in the NRS EFL descriptors for ESL.

Another important aspect of the NRS EFL descriptors for ESL lies in the way they conceptualize complexity to distinguish proficiency levels. The NRS EFL descriptor document explains that text complexity (e.g., complexity in syntactic structures and vocabulary) and topic progression (e.g., familiarity and orientation toward academic contexts) are important indicators of progression across proficiency levels. The document further emphasizes that the descriptors are intended to be progressive across levels, indicating that attainment of a higher level presupposes mastery of the language knowledge and skills specified for all preceding levels. This characteristic of the descriptors is critical for alignment judges when evaluating levels of complexity or proficiency.

Methodology

Panels of Content Experts

Past alignment studies have typically included three to 10 experts on a panel for a specific subject matter and grade levels (Herman et al., 2007). In this study, we convened two panels, each consisting of six to seven content experts. One panel (Panel RW) focused on the reading and writing domains of the tests, while the other (Panel LS) focused on the listening and speaking domains. As alignment evaluation relies on expert judgment, it is of utmost importance to select highly qualified experts to serve on alignment panels.

We sent recruitment letters to educators in adult education ESL programs who had participated in a prior ETS research study involving the piloting of the TOEIC Bridge tests or in previous ETS standard-setting studies. Individuals who responded to the recruitment letter completed an intake survey about their professional backgrounds. The main selection criteria for this study included experience and familiarity with the NRS EFL descriptors for ESL as well as experience teaching adult ESL learners. Efforts were also made to ensure diversity in terms of location and types of institutions where panel members teach adult ESL learners (e.g., public schools, community centers, higher education institutes). Tables 2 and 3 summarize the backgrounds of the individual members of Panel RW and Panel LS, respectively. All panel members had experience teaching ESL learners in adult education programs (an average of 11 years of teaching experience) and had prior experience using NRS-approved assessments such as the Comprehensive Adult Student Assessment System Student Test of English Progress and Success (CASAS STEPS), the Test of Adult Basic Education Complete Language Assessment System – English (TABE CLAS-E), or the Basic English Skills Test (BEST) Literacy/Plus assessments.

Table 2. Background Characteristics of Panel Members for the Reading and Writing Domains

ID	State	Institution type	Area	Teaching (years)	Age	Gender	Ethnicity/race
RW1	Colorado	Public school district	Urban	6	51–60	Female	White
RW2	Indiana	Community college	Rural	3	31–40	Female	White
RW3	Colorado	Community center	Urban	8	31–40	Female	Hispanic
RW4	Pennsylvania	University	Urban	18	31–40	Male	White
RW5	Indiana	Public school district	Rural	6	51–60	Female	White
RW6	Ohio	Community center	Suburban	14	41–50	Female	White

Table 3. Background Characteristics of Panel Members for the Listening and Speaking Domains

ID	State	Institution type	Area	Teaching (years)	Age	Gender	Ethnicity/race
LS1	Michigan	Community center	Suburban	11	21–30	Female	White
LS2	Minnesota	Public school district	Urban	4	51–60	Female	White
LS3	New Jersey	Public library	Suburban	10	41–50	Male	Hispanic
LS4	Illinois	Community college	Suburban	21	51–60	Female	White
LS5	Illinois	Community college	Suburban	4	31–40	Female	White
LS6	Massachusetts	University	Urban	17	31–40	Male	White
LS7	New York	University	Urban	20	51–60	Female	White

Note. Panel member LS7 participated in the alignment rating process for Form A of the TOEIC Bridge tests but was unable to participate in rating the second test forms, Form B, due to personal reasons.

Alignment Criteria

The alignment process focused on three major criteria: content match, content coverage, and proficiency- or complexity-level coverage. These criteria are defined as follows:

- **Content match.** This refers to the specific language knowledge and skills assessed by the test items and their correspondence to those reflected in the NRS EFL descriptors for ESL.
- **Content coverage.** This refers to the distribution of specific language knowledge and skills assessed in the tests, indicating the extent to which the language knowledge and skills from the NRS EFL descriptors for ESL are covered.
- **Proficiency- or complexity-level coverage (“level,” for short).** This refers to the alignment between the six proficiency levels defined in the NRS EFL descriptors for

ESL and the level(s) covered by the test items. An item/task may align with a single level or span a range of levels.

To evaluate the degree of content match between the items and the NRS EFL descriptors for ESL, three rating codes were utilized as follows:

- Complete match. This rating indicates that the language knowledge, skills, and abilities measured by the item are completely aligned with those included in the EFL descriptors for ESL.
- Partial match. This rating indicates that the language knowledge, skills, and abilities measured in the items are partially aligned with the EFL descriptors for ESL. The item measures additional knowledge, skills, and abilities that are not reflected in the EFL descriptors for ESL.
- No match. This rating indicates that the language knowledge, skills, and abilities measured in the items are not included in the EFL descriptors for ESL.

The concepts of content match and content coverage are similar to the categorical concurrence criterion and the range and balance of representation criterion employed in Webb's (2005) alignment method. These criteria are concerned with not only the content correspondence but also the distribution of the content in the tests compared to the NRS EFL descriptors for ESL. Similarly, the level criterion in this study is similar to the cognitive complexity, or depth of knowledge, criterion in Webb's method. However, drawing from alignment research on language assessments, using the level category to evaluate language complexity or proficiency progression is more appropriate than using the depth of knowledge criterion in this study.

It is important to note that a single item or task can cover multiple proficiency levels because the NRS EFL descriptors for ESL include a set of key language skills or language functions across all proficiency levels. For example, the skill of constructing a claim is included in all six levels (phrased as "expressing an opinion" in Levels 1 and 2). Instead, levels are differentiated by linguistic complexity and topic familiarity. Considering that speaking and

writing items are scored using rubrics that capture a range of performance levels, it is reasonable to expect that a single item may align with multiple levels of proficiency.

Study Instruments

In order to help panel members familiarize themselves with the TOEIC Bridge tests and alignment rating procedures, a set of study instruments and materials was developed and utilized. This section briefly describes the content and purpose of these materials.

- Test blueprints and specifications. Based on a number of publications on the TOEIC Bridge test frameworks, constructs, and task designs, a blueprint and test specification document was created for each test domain (i.e., listening, reading, speaking, writing) of the TOEIC Bridge tests (i.e., a total of four test blueprint and specifications documents). Each document includes information about the test purposes, target test takers, definitions of the constructs, test structures (e.g., number of items per task type), task types, intended language knowledge and skills measured by task types, and sample items for each task type. For the speaking and writing documents, scoring rubrics and sample responses are also included. These materials are intended to help panel members better understand the intended constructs and content represented in the tests.
- Alignment protocol. This document was developed to provide guidelines for panel members to follow standardized procedures during the alignment rating process. It contains information about the definition of alignment from the education measurement field, the alignment method and criteria employed in the present study, and detailed step-by-step procedures for reviewing test items and assigning alignment ratings. The alignment workshop and training meetings were organized following the alignment protocol (see Appendix A for the training slides).
- Notes on salient features of the NRS EFL descriptors. We (the authors involved in the alignment study) created this notes document to support the development of a shared and coherent understanding of the NRS EFL descriptors for ESL among panel members. As emphasized in previous alignment research (Herman et al., 2007),

having a common interpretation of standards or level descriptors is essential, especially since such descriptors often include general or ambiguous language. Following approaches suggested in prior studies (e.g., Bailey et al., 2022; Cook & McDonald, 2013; Wolf et al., 2023), this notes document lists key language knowledge, skills, and functions described in the NRS descriptors for ESL with a particular focus on features that distinguish the levels of proficiency. Figure 1 displays an example of how salient features of a specific NRS proficiency level were extracted to create the notes document. Language functions that are explicitly stated in the NRS descriptors for ESL are included in the notes document as key language functions (e.g., identify the main topic, retell a few key details, as highlighted in Figure 1). This document was shared with panel members and used during group discussions to establish consensus on the salient features of the NRS EFL descriptors for ESL. During these discussions, panel members had the opportunity to provide suggestions about the salient features and key language functions.

Figure 1. An Example of Notes on Salient Features Based on the National Reporting System Descriptors

Level 3: High Beginning ESL (ELP Standards for AE Level 2)		
Interpretive: The ability to process, understand, interpret and/or engage with level-appropriate literary and informational written and spoken text to construct meaning (1, 6, 8)		
ELLs ready to exit the High Beginning ESL Level are able to identify the main topic in oral presentations and simple spoken and written texts and retell a few key details using an emerging set of strategies.		
ELLs ready to exit this level are able to, with support, identify the main argument an author or speaker makes. They can, with support, identify one reason an author or a speaker gives to support the argument.		
ELLs ready to exit this level are able to determine the meaning of frequently occurring words, phrases, and expressions in spoken and written texts about familiar topics, experiences, or events.		
<i>Note.</i> The identified salient features are highlighted in yellow.		
Excerpt from the notes document:		
Level	Text complexity	Contexts, topics, language functions
3	- High-frequency vocabulary - Short texts	- Familiar topics - Identify the main topic/argument in simple texts - Retell a few key details - Identify one reason to support the argument

- Alignment rating form. An alignment form was created for panel members to record their ratings for each item. The form included the following categories: (a) language knowledge and skills measured, (b) content match, (c) level match, (d) primary level, (e) missing skills, and (f) notes and comments. In the first category, panel members were asked to describe the language knowledge and skills required to answer the item correctly (i.e., intended skills). Since a single item may span multiple proficiency levels, panel members were instructed to indicate all levels that the item covers, as well as to identify the primary level they believed was best covered by the item. After completing the rating form for each test, panel members were also asked to note any language knowledge or skills included in the NRS EFL descriptors for ESL that were not assessed by the test, if any (i.e., the missing skills category). Figure 2

presents an example of a completed rating for an item, using the alignment rating form.

Figure 2. An Example of the Ratings

Item type	Item number	Skills measured	Content match	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Primary level	Notes	Missing skills in general
Talks	L40	main idea in a short text; frequent vocabulary; complex sentence	1		1	1	1			3		

- Readiness form. This form was developed to confirm that panel members felt prepared to make independent ratings for items following the alignment training session. Appendix B includes the readiness form used in this study.
- Evaluation survey. This survey was created to gather panel members' evaluations about the alignment meetings and procedures used in this study. The survey questions are provided in Appendix B.

Training and Rating Procedures

As noted earlier, Panel RW was involved in rating the reading and writing tests, and Panel LS was involved in rating the listening and speaking tests. Each panel followed the steps outlined in this section to complete the alignment ratings for both Forms A and B of the TOEIC Bridge and TOEIC® tests. All alignment meetings and rating activities (i.e., Steps 3 through 8) were conducted virtually over a 1-week period for each panel.

- Step 1: Familiarization with alignment, NRS EFL descriptors for ESL, and TOEIC Bridge tests. Prior to the first training meeting, panel members received the alignment protocol, NRS EFL descriptors for ESL, and the blueprints and test specification documents. They were asked to review these materials thoroughly.
- Step 2: Test familiarization by taking the TOEIC Bridge tests. Panel members completed the tests online using the provided test authorization codes, in order to gain firsthand experience with the test content and format.
- Step 3: Participation in alignment training. Two separate virtual training sessions were held: one for Panel RW and one for Panel LS. During these sessions, panel members discussed the Notes on Salient Features document to develop a clearer

understanding of the key aspects of the NRS EFL descriptors for ESL. They also received instruction on how to use the alignment protocol and rating form, practiced rating with sample items, and engaged in group calibration discussions. After completing the training, panel members were asked to complete the readiness form to confirm their preparedness to proceed with independent item ratings. All panel members signed the statement of readiness.

- Step 4: Independent alignment rating of Form A tests. Panel members independently completed alignment ratings for Form A of the TOEIC Bridge tests. The test materials were made accessible through a secured SharePoint site, with panel members using their unique credentials for access. Completed rating forms were submitted in Microsoft Excel format.
- Step 5: Alignment rating discussion meeting for Form A. After the facilitators compiled and compared all the ratings, they identified items with rating discrepancies. Items for which a majority of panel members (i.e., at least four out of six members, or five out of seven members) provided the same rating were considered to have agreement. Panel members then participated in a virtual meeting to discuss the remaining items to reach consensus ratings. This meeting also served as an additional training opportunity.
- Step 6: Independent alignment rating of Form B tests. After the discussion meeting, panel members were asked to independently complete the ratings for Form B tests.
- Step 7: Alignment rating discussion meeting for Form B. Using the same process, facilitators prepared a list of items with rating differences. Panel members discussed these items to reach consensus.
- Step 8: Alignment workshop evaluation. Panel members completed a survey to evaluate and provide feedback on the alignment workshop meetings and the alignment process.

Rater Reliability

To examine the rater reliability, percentage agreement was computed for both the content and level categories. Because each panel included six to seven raters, a majority rating for each item was considered an indicator of agreement (i.e., reaching at least 67% rater agreement threshold), adopting the approach used in past alignment research (e.g., Herman et al., 2007). Then, the percentage of the agreement was calculated for each domain of the tests (i.e., listening, reading, speaking, writing). Additionally, intraclass correlation coefficients were computed for the level category, as it involved a scale from one to six levels. The statistical software SPSS Version 29 was used. Table 4 summarizes the rater agreement results. Overall, the indices suggest acceptable levels of agreement.

Table 4. Rater Agreement Indices

TOEIC Bridge tests	Content rating	Level rating	
	Percent agreement	Percent agreement	Intraclass correlation coefficient
Listening	1.00	0.79	0.87
Reading	0.96	0.81	0.79
Speaking	0.88	0.69	0.82
Writing	0.94	0.78	0.93

Analysis of Final Ratings

As described previously, rating disagreements were resolved through discussion, and consensus ratings were obtained for those items. Thus, the final ratings reflect either majority or consensus agreement. The alignment analysis was conducted using these final ratings.

The counts and percentages of the items rated as complete, partial, or no content match were calculated for each test form. The language knowledge and skills identified by panel members for each item were analyzed to determine the number of items representing each type of knowledge and skill. For the level ratings, the counts and percentages of items per each level were computed by test form.

Based on the distribution of aligned language knowledge and skills in each test form, a gap analysis was conducted to identify any language knowledge and skills from the NRS EFL

descriptors for ESL that are not assessed by the tests. A qualitative summary of these findings was completed.

Finally, descriptive statistics were computed for the evaluation survey.

Results of Alignment

The results are presented by interpretive modalities (i.e., listening, reading) and productive modalities (i.e., speaking, writing), as the NRS EFL descriptors for ESL are organized into these integrated modalities. As described earlier, the results are based on majority or consensus ratings.

Alignment Results for TOEIC Bridge Listening and Reading Tests

As shown in Table 5, panel members rated all items in the TOEIC Bridge Listening and Reading tests (both Forms A and B) as aligned with the NRS EFL descriptors for ESL in terms of content. In other words, all items were judged to measure the language knowledge and skills reflected in the NRS EFL descriptors for ESL.

Regarding the types of language knowledge and skills measured, the test forms are found to include a range of key knowledge and skills that are represented in the NRS EFL descriptors for ESL. Tables 6 and 7 summarize the language knowledge and skills identified by panel members in the listening and reading tests, respectively. As shown in these tables, panel members identified a mix of simple and complex structures and vocabulary, as well as simple and complex information that test takers are required to process in each test form. Key language functions described in the NRS EFL descriptors for ESL—such as understanding and responding to questions, determining author’s or speaker’s points or central ideas, identifying details, and participating in oral or written exchanges—are represented both in the listening and reading test forms. Additionally, analyzing the development of ideas or reasoning (e.g., identifying implied meanings through analysis) is also present in the reading test forms.

Table 5. Percentage of TOEIC Bridge Listening and Reading Test Items Aligned to the NRS**Descriptors**

Domain	Number of total items	Form A	Form B
		Complete match	Complete match
		%	%
Listening	50	100%	100%
Reading	50	100%	100%

Table 6. Content Distribution of TOEIC Bridge Listening Test Forms Relative to the NRS**Descriptors**

Language knowledge and skills	Form A		Form B	
	Number of items	%	Number of items	%
Recognize the meaning of simple words/phrases	20	40%	20	40%
Determine the meaning of high-frequent vocabulary	50	100%	50	100%
General academic and content-specific words/phrases	24	48%	26	52%
Understand simple sentences	23	46%	23	46%
Understand compound or complex sentences	24	48%	24	48%
Understand and respond to questions	30	60%	30	60%
Understand simple information	30	60%	33	66%
Determine author's or speaker's point(s)/central idea(s)	50	100%	50	100%
Identify details	32	64%	24	48%
Label information	10	20%	8	16%
Analyze the development of ideas or reasoning	4	8%	4	8%
Participate in conversations (oral exchanges)	30	60%	30	60%

Table 7. Content Distribution of TOEIC Bridge Reading Test Items Relative to the NRS**Descriptors**

Language knowledge and skills	Form A		Form B	
	Number of items	%	Number of items	%
Recognize the meaning of simple words/phrases	5	10%	5	10%
Determine the meaning of high-frequent vocabulary	36	72%	36	72%
General academic and content-specific words/phrases	29	58%	26	52%
Understand simple sentences	32	64%	35	70%
Understand compound or complex sentences	23	46%	22	44%
Understand simple information	28	56%	29	58%
Determine author's or speaker's point(s)/central idea(s)	50	100%	50	100%
Identify details	16	32%	16	32%
Analyze the development of ideas or reasoning	8	16%	11	22%
Participate in written exchanges	21	42%	21	42%

With respect to the level or complexity category, panel members judged that the TOEIC Bridge Listening test (both Forms A and B) covered Levels 1 through 4 (see Table 8). For the TOEIC Bridge Reading test, panel members rated Form A as covering Levels 2 through 4 and Form B as covering Levels 2 through 5 (see Table 9). Only 1 out of 50 items was rated at Level 5. Although Level 1 was not explicitly mentioned for any reading items, Level 2 presumes mastery of Level 1 proficiency.

Table 8. Coverage of Levels in TOEIC Bridge Listening Test Items

Level	Form A			Form B		
	Number of items	%	Cumulative %	Number of items	%	Cumulative %
Level 1	3	6%	6%	3	6%	6%
Level 2	5	10%	16%	4	8%	14%
Level 3	24	48%	64%	25	50%	64%
Level 4	18	36%	100%	18	36%	100%

Table 9. Coverage of Levels in TOEIC Bridge Reading Test Items

Level	Form A			Form B		
	Number of items	%	Cumulative %	Number of items	%	Cumulative %
Level 2	7	14%	14%	6	12%	12%
Level 3	32	64%	78%	33	66%	78%
Level 4	11	22%	100%	10	20%	98%
Level 5	0	0%	0%	1	2%	100%

Alignment Results for TOEIC Bridge Speaking and Writing Tests

As presented in Table 10, panel members rated all items in the TOEIC Bridge Speaking and Writing tests (both Forms A and B) as aligned with the NRS EFL descriptors for ESL in terms of content. All items, considered together with their scoring rubrics, were judged to measure the language knowledge and skills reflected in the NRS EFL descriptors for ESL. During the alignment rating process, high agreement rates (i.e., five to six panel members rating each item as a complete match) were obtained for all items except two items in the speaking test. These two items were from the Read a Short Text Aloud task type. Panel members discussed this item type in detail to reach a consensus on complete match ratings. They agreed that although the NRS EFL descriptors for ESL do not explicitly mention pronunciation, the descriptors are not exhaustive, and the new ELP standard 10 overtly specifies the demonstration of command of standard English to communicate in level-appropriate speech and writing. Panel members also noted that this item type covers a range of linguistic complexity as well as underlying language skills for understanding an author/speaker's points and delivering presentations, all of which are included in the NRS EFL descriptors for ESL.

Table 10. Percentage of TOEIC Bridge Speaking and Writing Test Items Aligned to the NRS Descriptors

Domain	Number of total items	Form A	Form B
		Complete match %	Complete match %
Speaking	8	100%	100%
Writing	9	100%	100%

Regarding the types of language knowledge and skills measured, the test forms were found to include a range of key knowledge and skills represented in the NRS EFL descriptors for ESL. Tables 11 and 12 summarize the language knowledge and skills identified by panel members in the speaking and writing tests, respectively. As shown in these tables, panel members identified a mix of simple and complex structures and vocabulary, as well as simple and complex information that test takers are required to process in each test form. Key language functions described in the NRS EFL descriptors for ESL such as understanding and responding to questions, participating in conversations/written exchanges, constructing a claim/opinion, and recounting/retelling details are represented commonly in the speaking and writing test forms. Additionally, writing an informational text is also present in the writing test forms.

Table 11. Content Distribution of TOEIC Bridge Speaking Test Items Relative to the NRS Descriptors

Language knowledge and skills	Form A		Form B	
	Number of items	%	Number of items	%
Use simple words/phrases	2	25%	2	25%
Use high frequent words	7	88%	7	88%
General academic and content-specific words/phrases	4	50%	4	50%
Use simple sentences	8	100%	8	100%
Use compound or complex sentences	8	100%	8	100%
Understand/respond to questions	1	13%	1	13%
Communicate simple information	2	25%	2	25%
Deliver presentations	2	25%	2	25%
Participate in conversations	1	13%	1	13%
Construct a claim/opinion	1	13%	1	13%
Recount/retell information or details	3	38%	3	38%

Table 12. Content Distribution of TOEIC Bridge Writing Test Items Relative to the NRS**Descriptors**

Language knowledge and skills	Form A		Form B	
	Number of items	%	Number of items	%
Use simple words/phrases	3	33%	3	33%
Use high frequent words	8	89%	8	89%
General academic and content-specific words/phrases	3	33%	2	22%
Use simple sentences	6	67%	6	67%
Use compound or complex sentences	4	44%	4	44%
Understand/respond to questions	2	22%	2	22%
Communicate simple information	2	22%	2	22%
Construct a claim/opinion	2	22%	2	22%
Write informational text	1	11%	1	11%
Retell information or details	1	11%	1	11%
Participate in written exchanges	2	22%	2	22%

With respect to the level or complexity category, panel members judged that the TOEIC Bridge Speaking test (Forms A and B) covered Levels 2 through 5 (see Table 13). The majority of panel members considered one speaking task to extend up to Level 5. For the TOEIC Bridge Writing test, panel members rated both Forms A and B as covering Levels 2 through 4 (see Table 14). Although Level 1 was not explicitly mentioned for any speaking and writing items, Level 2 presumes mastery of Level 1 proficiency.

Table 13. Coverage of Levels in TOEIC Bridge Speaking Test Items

Level	Form A			Form B		
	Number of items	%	Cumulative %	Number of items	%	Cumulative %
Level 2	2	25%	25%	2	25%	25%
Level 3	3	38%	63%	3	38%	63%
Level 4	2	25%	88%	2	25%	88%
Level 5	1	13%	100%	1	13%	100%

Table 14. Coverage of Levels in TOEIC Bridge Writing Test Items

Level	Form A			Form B		
	Number of items	%	Cumulative %	Number of items	%	Cumulative %
Level 2	4	44%	44%	3	33%	33%
Level 3	2	22%	67%	3	33%	67%
Level 4	3	33%	100%	3	33%	100%

Panel Evaluation Survey Results

Panel members took the evaluation survey after completing all alignment sessions and ratings. The survey included questions related to their overall evaluation of the alignment training and discussion meetings, and materials used throughout the process. Additionally, panel members were asked to rate their level of comfort with their own alignment ratings. All survey questions used a 4-point Likert scale. Table 15 presents the mean rating for each question. In general, panel members' responses were positive, with most mean ratings falling between 3 and 4. Panel members particularly appreciated the materials designed to enhance understanding of the NRS EFL descriptors for ESL such as the highlighted key features and the notes documents (see Figure 1). Panel members also reported a high level of confidence in the adequacy of their own ratings. The survey results generally support the soundness of the alignment procedures used in this study, which, in turn, reinforces the credibility of the panel ratings.

Table 15. Summary of Evaluation Survey Results

Survey item/statement	Panel LS		Panel RW	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Overall evaluation (1 = <i>strongly disagree</i> ; 2 = <i>disagree</i> ; 3 = <i>agree</i> ; 4 = <i>strongly agree</i>)				
I understood the purpose of the study.	3.7	0.49	3.3	0.52
I understood the key KSAs described in the NRS EFL descriptors.	3.4	0.53	3.7	0.52
I understood the distinguishing features of the six EFL descriptor levels.	3.3	0.49	3.2	0.41
I became more familiar with the NRS EFL descriptors as I completed the alignment ratings.	4.0	0.00	3.8	0.41
I understood the intended language knowledge and skills assessed in the test items.	3.6	0.53	3.5	0.55
The instructions and explanations provided by the facilitators were clear.	3.4	0.53	3.3	0.52
The facilitators were able to answer questions, provide information, and clarify points when needed.	3.9	0.38	3.7	0.52
The explanations of consensus ratings for disagreement items were clear.	3.7	0.49	3.2	0.41
Helpfulness of materials (1 = <i>not helpful at all</i> ; 2 = <i>not helpful</i> ; 3 = <i>helpful</i> ; 4 = <i>very helpful</i>)				
Test specification and blueprint documents	3.4	0.53	3.7	0.52
Original EFL descriptor document	3.3	0.76	3.3	0.52
EFL descriptors in table format with key features highlighted	3.9	0.38	4.0	0.00
NRS Descriptor notes document	4.0	0.00	3.7	0.52
Panel meeting slides	3.7	0.49	3.2	0.75
Comfort level with ratings (1 = <i>very uncomfortable</i> ; 2 = <i>somewhat uncomfortable</i> ; 3 = <i>comfortable</i> ; 4 = <i>very comfortable</i>)				
Content match for reading items	3.9	0.38	3.7	0.52
Level match for reading items	3.9	0.38	3.3	0.52
Content match for writing items	3.6	0.53	3.7	0.52
Level match for writing items	4.0	0.00	3.7	0.52

Note. LS = listening/speaking; RW = reading/writing

Results of Gap Analysis

The findings on the types and complexity levels of the language knowledge and skills identified in the test items provide insight into any gaps between the TOEIC Bridge tests and the NRS EFL descriptors for ESL. This section summarizes those language knowledge and skills that are included in the NRS EFL descriptors for ESL but not assessed by the tests, if any. The summary is based on the alignment results for content and level coverage as well as on panel members' written comments on any missing skills collected through the alignment rating form.

Results of Gap Analysis for TOEIC Bridge Listening and Reading Tests

As presented in Tables 6 and 7, the TOEIC Bridge Listening and Reading tests encompass a range of linguistic and topic complexity with key language functions included in the interpretive and interactive modalities of the NRS EFL descriptors for ESL. One panel member noted that the ability of carrying out short research projects, which is mentioned across all six levels under the interactive modality, was not explicitly present in the tests. However, all panel members acknowledge that it would be difficult to assess students' ability to carry out research projects in standardized assessment settings with limited item types. Moreover, because conducting a research project involves not only language proficiency but also a range of additional knowledge and skills (e.g., content-specific knowledge and the ability to navigate relevant resources), measuring research skills falls beyond the construct of a language proficiency test.

The level coverage for the TOEIC Bridge Listening and Reading tests was found to extend up to Level 4. Thus, specific language functions included in Levels 5 and 6, such as summarizing a text and analyzing reasoning in persuasive spoken and written texts to evaluate whether the evidence is sufficient, are not assessed.

Results of Gap Analysis for TOEIC Bridge Speaking and Writing Tests

As summarized in Tables 11 and 12, the TOEIC Bridge Speaking and Writing tests featured a mix of simple to complex linguistic features and topics involving key language functions described in the NRS EFL descriptors for ESL. Like in the TOEIC Bridge Listening and Reading tests, panel members commented on the absence of assessing the ability to conduct research projects in standardized assessments, although this ability is described in the NRS EFL descriptors for ESL. Based on the content coverage results, the skill of providing a concluding statement in producing extended written or spoken texts also appears to be lacking in the tests. However, the Make and Support a Recommendation task type in the speaking test and the Respond to an Extended Message task type in the writing test may elicit students' ability to provide a concluding statement, although the scoring rubrics do not explicitly require a concluding statement in students' responses.

Regarding level coverage, some panel members initially assessed that certain task types in the speaking and writing tests (e.g., Tell a Story, Make and Support a Recommendation, Respond to an Extended Message) could cover Level 5 or higher. However, reviewing sample responses of the highest scores led them to agree on Level 4.

Table 16 presents a summary of the skills and abilities described in the NRS EFL descriptors for ESL but not assessed in the TOEIC Bridge tests.

Table 16. A Summary of Gap Analysis Results for the TOEIC Bridge Tests

Test	Absent skills and abilities	Weak coverage of levels
TOEIC Bridge Listening and Reading	Conducting research projects Summarizing a text (Level 5 and above) Analyzing reasoning in persuasive spoken and written texts to evaluate whether the evidence is sufficient (Level 5 and above)	Levels 5 and 6
TOEIC Bridge Speaking and Writing	Conducting research projects Providing a concluding statement in producing extended written or spoken texts	Levels 5 and 6 (given the expectations for the highest score in rubrics)

Summary and Discussion

The present study evaluated the alignment between the TOEIC Bridge tests and the NRS EFL descriptors for ESL. Drawn on past research in alignment methodology, the study employed a systematic approach that included the use of trained expert panels and standardized procedures. The study results provide strong evidence of content alignment for all test forms. Panel members consistently rated all items in the TOEIC Bridge tests as aligned with the NRS EFL descriptors for ESL in terms of content (i.e., key language knowledge, skills, abilities).

The TOEIC Bridge tests generally cover the core language skills and language functions manifested in the NRS EFL descriptors for ESL. Such skills and functions include understanding and responding to questions, identifying an author's/speaker's point, retelling/recounting key details, expressing an opinion/constructing a claim with supporting reasons, and analyzing the development of the ideas. These language functions are described across all six levels of the NRS EFL descriptors for ESL. What differentiates the levels lies in the degree of linguistic and topical complexity involved in performing those communicative language functions. For example, Level 1 includes the ability to express an opinion about a familiar topic, while Level 5 states "constructing a claim about a variety of topics," which demands greater sophistication in

language use. As the level increases, the complexity of sentence structures and the use of academic and content-specific vocabulary become more demanding. Panel members' judgments about content match and coverage were consistent across both Forms A and B of each test. This consistency likely reflects the deliberate design of parallel forms in accordance with the test and item specifications of the TOEIC Bridge tests.

Due to the way that proficiency levels are distinguished in the NRS EFL descriptors for ESL, panel members rated the TOEIC Bridge tests as encompassing Levels 1 through 4. Although the task types were judged to assess key language skills and functions across all levels, the linguistic and topical complexity of the TOEIC Bridge tests influenced panel members' judgments regarding level coverage.

With respect to the gap analysis results, panel members consistently identified the skill of carrying out research projects, listed under the interactive modality of the NRS EFL descriptors for ESL, as a missing skill in the TOEIC Bridge tests. However, panel members noted that this skill may not be appropriate for standardized assessment settings due to practical constraints. This particular skill may be better evaluated through performance-based tasks or classroom-based assessments. Additionally, skills such as summarizing a text and evaluating the sufficiency of evidence in persuasive arguments were found to be underrepresented in the tests. While certain task types (e.g., Listen and Retell, Make and Support a Recommendation) may partially elicit these abilities, they do not directly target them as their primary focus.

It is worth noting the challenges that panel members encountered during the alignment rating process. The NRS EFL descriptors for ESL explicitly state in their introduction that the descriptors do not comprehensively cover all language skills but instead focus on the most critical ones. Furthermore, the descriptors explain that the distinctions between levels reflect "increasingly complex and cognitively demanding language structures, academic vocabulary, and concepts" (DAEL, 2025, p. B-20), without clearly specifying the level distinction. Due to this inherent ambiguity, panel members often found it difficult to determine level coverage for test items. They frequently perceived that a single item could span multiple proficiency levels. To address this challenge, the present study developed and utilized a notes document that outlined key language skills and features distinguishing each level. Panel members reported in

the evaluation survey that this notes document was particularly helpful in guiding their rating process and establishing a coherent understanding of the descriptors. Given that panel members worked with varied curricula and student populations with different proficiency levels, they especially appreciated having a shared reference to support consistent interpretation of the NRS EFL descriptors for ESL.

Another challenge in the NRS EFL descriptors for ESL is that they do not reflect underlying, foundational language skills, even though such skills are included in the new *English Language Proficiency Standards for Adult Education* (Standards 8–10). The ELP Standards document explains that Standards 8–10 focus on micro-level linguistic features, including the demonstration of command of the conventions of standard English (Office of Career, Technical and Adult Education, 2016). However, the NRS descriptors are more focused on describing key communicative skills and language functions, as exemplified earlier. Panel members remarked that they often emphasize foundational language skills such as grammar, syntax, and vocabulary development, particularly in lower-level classes, as these are essential building blocks for performing communicative tasks. Due to the absence of specific descriptions about these foundational skills in the NRS EFL descriptors for ESL, panel members expressed difficulty in rating test items that target these areas. In the present alignment study, panel members underscored the importance of such foundational knowledge and skills, aligned with the ELP Standards, and incorporated this perspective when assigning alignment ratings to items in the TOEIC Bridge tests.

These challenges point to some limitations of the study. As the results of the alignment study are largely shaped by expert judgments and the methods employed, the study findings should be interpreted with caution. Although the present study made a deliberate effort to select highly qualified expert panel members and to implement systematic and standardized procedures, alignment judgments may vary depending on individual expertise, instructional contexts, and interpretive frameworks. Future research that compares the content of the TOEIC Bridge tests with other tests aligned with the NRS EFL descriptors for ESL may offer additional insights into their relative strengths and coverage. Additionally, studies involving a larger and more diverse panel of experts may help validate and extend the current findings. Examining

how learners at various NRS levels perform on the TOEIC Bridge tests in real instructional settings may also shed light on the practical implications of alignment for teaching and learning.

Conclusion and Recommendations

Although the TOEIC Bridge tests were not originally developed based on the NRS EFL descriptors for ESL, the present study provides compelling evidence of strong alignment with these descriptors. Designed to assess the ELP of adult English learners in everyday and workplace contexts, the TOEIC Bridge tests incorporate task types that stem from a careful domain analysis of target language use. The overlap in the intended target population, as well as the shared construct of communicative language abilities, seem to contribute to the alignment observed in this study.

Based on the findings and lessons learned from the study, we offer a few practical recommendations to support test users in understanding and using the TOEIC Bridge tests effectively in relation to the NRS EFL descriptors for ESL:

- Alignment reference guide. It would be beneficial to develop and provide a reference document that clearly explains how the constructs and task types of the TOEIC Bridge tests align with the NRS EFL descriptors for ESL. This document may include illustrative examples, mapped language functions, and alignment rationales to help users interpret test content in relation to the descriptors.
- Professional development resources. It would be important to provide training or orientation materials for educators and program administrators who use the tests. These resources should cover the NRS EFL descriptors, test constructs, and how to interpret test results in light of instructional planning and learner progress tracking.
- Score interpretation support. It would also be important to provide detailed guidance on how test scores can be interpreted in relation to the NRS levels. This guidance may include alignment tables, sample learner profiles, and suggested instructional follow-ups based on proficiency levels.
- Complementary assessment use. As the study findings indicate that certain skills such as conducting research projects, summarizing texts, and evaluating the

sufficiency of evidence in persuasive texts are not primarily assessed in the TOEIC Bridge tests, educators may need to supplement the tests with additional classroom-based evaluations to cover those skills.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
<https://doi.org/10.7916/D8CV4HB8>
- Bailey, A. L., Wolf, M. K., & Ballard, L. (2022). *English language proficiency standards and assessment alignment: Issues, insights, and innovations for guiding peer review* (Research Notes). ETS.
https://www.ets.org/Media/Research/pdf/Research_Notes_Bailey_Wolf_Ballard.pdf
- Cook, H. G. (2005). *Aligning English language proficiency tests to English language learning standards*. Council of Chief State School Officers.
- Cook, H. G., & MacDonald, R. (2013). *Tool to evaluate language complexity of test items* (WCER Working Paper No. 2013-5). Wisconsin Center for Education Research.
https://wcer.wisc.edu/docs/working-papers/Working_Paper_No_2013_05.pdf
- Council of Chief State School Officers. (2014). *English language proficiency (ELP) standards with correspondences to K–12 English language arts (ELA), mathematics, and science practices, K–12 ELA standards, and 6–12 literacy standards*. <https://elpa21.org/wp-content/uploads/2023/02/1.2.2-ELP-Standards-2014.pdf>
- Division of Adult Education and Literacy. (2025). *Technical assistance guide for performance accountability under the Workforce Innovation and Opportunity Act: National reporting system for adult education*. Office of Career, Technical, and Adult Education; U.S. Department of Education. <https://nrs.ed.gov/sites/default/files/2025-05/NRS-TA-Guide-April-2025-508.pdf>
- ETS. (n.d.). *TOEIC® tests*. <https://www.ets.org/toEIC/about/bridge.html>
- Everson, P., Duke, T., Garcia Gomez, P., Carter Grissom, E., Park, E., & Schmidgall, J. (2019). *Development of the redesigned TOEIC Bridge® tests* (Research Memorandum No. RM-19-10). ETS. <https://www.ets.org/Media/Research/pdf/RM-19-10.pdf>
- Johnson, D. F. (2005). *Aligning ELP assessments to ELP standards*. Harcourt Assessment.

- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62.
https://doi.org/10.1207/S15366359MEA0101_02
- Office of Career, Technical and Adult Education. (2016). *Adult English Language Proficiency Standards for Adult Education*. U.S. Department of Education.
<http://lincs.ed.gov/publications/pdf/elp-standards-adult-ed.pdf>
- Office of Elementary and Secondary Education. (2018). *A state's guide to the U.S. Department of Education's assessment peer review process*. U.S. Department of Education.
<https://www.ed.gov/sites/ed/files/2020/07/assessmentpeerreview.pdf>
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14. <https://doi.org/10.3102/0013189X031007003>
- Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment*, 9(1–2), 1–27.
<https://doi.org/10.1080/10627197.2004.9652957>
- Sato, E., Lagunoff, R., & Worth, P. (2008). *Language for achievement: A framework for academic English language*. WestEd.
- Schmidgall, J. (2021). *Mapping the redesigned TOEIC Bridge® test scores to proficiency levels of the Common European Framework of Reference for Languages* (Research Memorandum No. RM-21-01). ETS. <https://www.ets.org/Media/Research/pdf/RM-21-01.pdf>
- Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). *Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests—Framework paper* (Research Report No. RR-19-30). ETS.
<https://doi.org/10.1002/ets2.12267>
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Council of Chief State School Officers.
- Webb, N. L. (2005). *Web alignment tool (WAT) training manual*. Wisconsin Center for Education Research.

Wolf, M. K., Bailey, A. L., & Ballard, L. (2023). Aligning English language proficiency assessments to standards: Conceptual and technical issues. *TESOL Quarterly*, 57(2), 670–685.

<https://doi.org/10.1002/tesq.3199>

Wolf, M. K., Bailey, A. L., Ballard, L., Wang, Y., & Pogossian, A. (2022). Unpacking the language demands in academic content and English language proficiency standards for English learners. *International Multilingual Research Journal*, 17(1), 68–85.

<https://doi.org/10.1080/19313152.2022.2116221>

Appendix A: Alignment Training Presentation Slides, Sample for Panel RW

***ets**

TOEIC Bridge/TOEIC Tests and the NRS Educational Functioning Level Descriptors for ESL: Alignment Study Panel 1, Meeting 1

Facilitators: Miyoung Kim Wolf & Renka Ohta

1

Agenda

01	Introduction & Purposes of the study (10 min)
02	NRS EFL Descriptors for ESL (15 min)
03	Alignment rating procedure (15 min)
04	Sample rating (15 min)
05	Practice rating (20 min)
06	Next steps (10 min)

***ets**

2

Purposes of the Alignment Study

- To collect expert judgments on the degree of alignment between the TOEIC family tests and the **Educational Functioning Level (EFL) Descriptors for ESL** used in the **National Reporting System for Adult Education (NRS)**

***ets**

3

***ets**

New Educational Functioning Level Descriptors for English as a Second Language (ESL)¹⁷

EFL Descriptors for ESL

Introduction

In the National Reporting System for Adult Education (NRS), the Educational Functioning Level (EFL) descriptors are intended to guide teaching and assessment for adult learners. The descriptors for English as a second language (ESL) are divided into six educational functioning levels: Beginning ESL, Low Beginning ESL, High Beginning ESL, Low Intermediate ESL, High Intermediate ESL, and Advanced ESL. The descriptors do not provide a complete or comprehensive delineation of all of the skills at any given level but rather provide a description of the most critical concepts and skills for the level.

Although these narrative descriptors address the most critical concepts for assessment and instruction for adult learners, lesson plans and test items should be based on additional content concepts from state instructional frameworks and standards, as appropriate for the learner and state requirements.

The EFLs for ESL are organized into three modalities: interpretive, productive, and interactive. These modalities include the domains of reading, writing, speaking, and listening. These modalities allow for an integrated or holistic approach to teaching and assessing English language learners (ELLs) in the adult education setting.

- Interpretive** refers to the learner's ability to process, understand, interpret, or engage with level-appropriate literary and informational written and spoken text to construct meaning. For example, an ELL exiting from the Low Intermediate ESL classroom should be able to, with support, explain the reasons an author or a speaker gives to support a claim and identify one or two reasons an author or a speaker gives to support the main point.
- Productive** refers to the learner's ability to produce level-appropriate written and spoken text such that it meaningfully transmits meaning. For example, an ELL exiting from the Low Beginning ESL classroom should be able to, with support, communicate information and feelings about familiar texts, topics, and experiences.
- Interactive** refers to the learner's ability to process and produce level-appropriate written and spoken text interactively with the purpose of understanding, interpreting, engaging in, and meaning-making. For example, ELLs exiting from the High Beginning ESL classroom should be able to, with support, gather information from provided print and digital sources, record information in simple notes, and summarize data and information.

4

Understanding the EFL Descriptors for ESL

ENGLISH LANGUAGE PROFICIENCY STANDARDS FOR ADULT EDUCATION

Are Comprehensive to College and Career Readiness Standards for English Language Arts and Literacy, and Mathematical and Science Practices

New ELP Standards for Adult Education (2016)

New Educational Functioning Level Descriptors for English as a Second Language (ESL)¹⁷

Introduction

In the National Reporting System for Adult Education (NRS), the Educational Functioning Level (EFL) descriptors are intended to guide teaching and assessment for adult learners. The descriptors for English as a second language (ESL) are divided into six educational functioning levels: Beginning ESL, Low Beginning ESL, High Beginning ESL, Low Intermediate ESL, High Intermediate ESL, and Advanced ESL. The descriptors do not provide a complete or comprehensive delineation of all of the skills at any given level but rather provide a description of the most critical concepts and skills for the level.

- 3 modalities: Interpretive (listening, reading), productive (speaking, writing), and interactive -> focusing on the integrated use of language in communication
- 6 levels: (1) Beginning, (2) Low Beginning, (3) High Beginning, (4) Low Intermediate, (5) High Intermediate, and (6) Advanced ESL
- ELP progression: increasingly complex and cognitively demanding language structures, academic vocabulary, and concepts

***ets**

Alignment evaluation: Identify key language knowledge, skills, and abilities, as well as the features that distinguish each level

5

NRS levels: Salient features (examples)

Level	Linguistic complexity	Contexts, topics, language functions
1	- Simple words/phrases - A narrow range of vocabulary - Simple sentences	- Familiar topics - Understand and respond to simple questions - With support, communicate simple info and express a preference/opinion - Label key info
2	- A few words, phrases - High-frequency words - Routinized expressions - Simple texts	- Familiar topics - Identify an author/speaker's point - Understand and respond to simple questions (simple yes/no, some who-2) - Express an opinion and give a reason to it
3	- High-frequency vocabulary - Use common linking words (with support) - Emerging informal, informal registers - Some common academic and content-specific words - Simple and compound sentences - Short texts	- Familiar topics - Emerging common academic/social contexts - With support, short presentations - Write simple narratives or information texts - Construct a claim, given a reason - With support, retell a short sequence of events - Respond to simple questions and ask questions - Identify the main topic/argument in simple texts - Retell a few key details - Identify one reason to support the argument

***ets**

6



7

Alignment Categories

- (1) **Content match:** This refers to the specific language knowledge and skills assessed by the tests and their correspondence to those reflected in the EFL Descriptors for ESL.
- 1 – **Complete match:** The content (knowledge and skills) of the item is fully reflected in the descriptors.
 - 0.5 – **Partial match:** The content of the item is partially reflected in the level descriptors. The item also assesses other skills that are not included in the level descriptors.
 - 0 – **No match:** The content of the item is not reflected in the level descriptors.
- (2) **Proficiency level match:** This refers to the alignment between the six proficiency levels defined in the EFL Descriptors for ESL and the level(s) covered by the test items.

Note: Judging the content match and proficiency level match is likely to occur simultaneously.

ets

8

Excel rating files

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Item type	Item number	Skills measured	Content match	Notes	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Primary level	Notes	Missing skills to general
Real Pictures	101												
Real Pictures	102												

ets

9

Coding Procedure

- (1) **Skill identification (Column C):** Review each item and identify the main language knowledge, skills, and abilities it measures. Refer to the test specification and blueprint documents to understand the intended skills measured. For speaking and writing items, also review the scoring rubrics and sample responses to better understand what the items are intended to assess.
- Guiding question: *What language knowledge and skills would students need to answer the item correctly?*
- (2) **Degree of content match (Column D):** Note that an item may assess multiple skills. Select 1 – Complete match, 0.5 – Partial match, or 0 – No match.
- (3) **Level match (Columns F–J):** Determine the level(s) to which each item aligns, based on the EFL Descriptors for ESL. An item may align with a single level or span a range of levels.
- Enter 1 in the column(s) (Columns F through J) that correspond to the primary level(s) you judge the item to be aligned with.
 - Enter the **primary level** you judge the item to align with, if applicable in Column L.
- (4) **Missing skills (Column N):** After rating all items, reflect and describe any language knowledge, skills, or abilities that are included in the EFL Descriptors for ESL but are not assessed by the test, if any. Use Column N (Missing skills) for your overall comments.

ets

10



11

Reading Sample

Test Computer sample item:

Practice Item

Read the text and answer the questions.

1. (1) Mr. ...
(2) Ms. ...
(3) Mr. ...
(4) Ms. ...

2. (1) Please ask for help.
(2) Red is a bright color.
(3) They all arrive today.
(4) Put them on the desk.

3. (1) cheap
(2) cheap
(3) cheap
(4) cheap

Step 1: Identify

Language knowledge, skills, abilities required to answer the item correctly:

- Understand short correspondence or descriptive text
- Identify the author's points
- Common vocabulary
- Simple sentence structure
- Grammatical features (pronouns, tense, prepositional phrases)
- Organization of a short text

Step 2: Determine the content and level match against

EFL Descriptors for ESL – Interpretive, Interactive

- Content: Complete match
- Primary level: 3
- Other level(s): 2

Skills measured	Content match	Notes	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Primary level
Read text for general, author point, details, grammar, pronunciation	1				1				3

ets

12

Writing Sample

Directions: Write a short blog post about a problem you solved. Tell a story about it, including what the problem was and how you solved it.

In your story, you can describe people, places, actions, and feelings.

Sample response: I help a friend when they get hurt or sick. Last summer my friend got hurt in his leg I helped him in school to carry his bag. When I carry his bag he felt happy.

Language knowledge, skills, abilities required to answer the item correctly:

- Describe, narrate, and sequence events
- Recount a sequence of events
- Vocabulary
- Sentence structure
- Cohesion, common transitional words/phrases

EFL Descriptors for ESL – Productive, Interactive

- Content: Complete match
- Primary level: 4
- Other level(s): 3

Score	Response Description
3	The response successfully addresses the prompt and is marked by all of the following: - The response fully addresses the topic and task. - Clear meaning is clear, though minor grammatical errors that do not obscure meaning may be present. - The story is told in a logical sequence and any connectors are used appropriately. - The choice of vocabulary is appropriate to the topic of the prompt.
2	The response is partially effective at addressing the prompt. - The response partially addresses the topic or partially completes the task, and/or - Use of language structures contributes to meaning, though grammatical errors may occasionally obscure meaning, and/or - The logical sequence of the story is mostly clear, and/or - The choice of vocabulary is sometimes limited or inappropriate to the topic.
1	The response does not effectively address the prompt and exhibits one or more of the following: - The response is unrecognizable at addressing the topic or task, though it may contain some related words, and/or - The response is off topic or seriously underdeveloped - Frequent and serious grammatical errors interfere with comprehensibility of most of the responses, and/or - The choice of vocabulary is limited (use of isolated words), meaningless, and/or it relies on repetition of the prompt in most of the response.

ets

13

Writing Sample

Directions: Write a short blog post about a problem you solved. Tell a story about it, including what the problem was and how you solved it.

In your story, you can describe people, places, actions, and feelings.

Sample response: I help a friend when they get hurt or sick. Last summer my friend got hurt in his leg I helped him in school to carry his bag. When I carry his bag he felt happy.

Language knowledge, skills, abilities required to answer the item correctly:

- Describe, narrate, and sequence events
- Recount a sequence of events
- Vocabulary
- Sentence structure
- Cohesion, common transitional words/phrases

EFL Descriptors for ESL – Productive, Interactive

- Content: Complete match
- Primary level: 4
- Other level(s): 3

Score	Response Description
3	The response successfully addresses the prompt and is marked by all of the following: - The response fully addresses the topic and task. - Clear meaning is clear, though minor grammatical errors that do not obscure meaning may be present. - The story is told in a logical sequence and any connectors are used appropriately. - The choice of vocabulary is appropriate to the topic of the prompt.
2	The response is partially effective at addressing the prompt. - The response partially addresses the topic or partially completes the task, and/or - Use of language structures contributes to meaning, though grammatical errors may occasionally obscure meaning, and/or - The logical sequence of the story is mostly clear, and/or - The choice of vocabulary is sometimes limited or inappropriate to the topic.
1	The response does not effectively address the prompt and exhibits one or more of the following: - The response is unrecognizable at addressing the topic or task, though it may contain some related words, and/or - The response is off topic or seriously underdeveloped - Frequent and serious grammatical errors interfere with comprehensibility of most of the responses, and/or - The choice of vocabulary is limited (use of isolated words), meaningless, and/or it relies on repetition of the prompt in most of the response.

ets

14

Practice Sample 1

Skills measured

Content match

Level match

Reading sample item: Reading Comprehension

Used Car For Sale. Several used Corolla Custom, City one owner, low mileage. Car used to commute short distances to town. Brakes and tires replaced six months ago. Truck replaced two weeks ago. All conditioning works well, but heater takes a while to warm up. Brand new spare tire included. Priced to sell. Owner going overseas at the end of this month and must sell the car. Call Francisco Gonzalez at (848) 555-0135.

147. What is suggested about the car?

(A) It was recently repaired.
(B) It has had more than one owner.
(C) It is very fuel efficient.
(D) It has been on sale for six months.

ets

15

Practice Sample 2

Skills measured

Content match

Level match

Writing sample item: Write a Sentence

Directions: Write a 200-word sentence based on the picture. Use the 700 words or phrases under the picture. You may rearrange the items in the words and you can use them in any order. You have 30 seconds to write.

very heavy

Sample response: a score of 3
They are carrying a heavy boxes.

ets

16

Practice Sample 3

Skills measured

Content match

Level match

Reading sample item: Incomplete Sentences

Incomplete Sentences sample item:

Customer reviews indicate that many modern mobile devices are often unnecessarily _____.

(A) complication (C) complicate
(B) complicates (D) complicated

ets

17

Practice Sample 4

Skills measured

Content match

Level match

Writing sample item: Write a Narrative

Directions: Write a short blog post about a time when you helped a friend. Tell a story about it, including who you helped and what happened.

In your story, you can describe people, places, actions, and feelings.

You have 8 minutes to write. Write as much as you can in the time provided.

ets

18

Next steps

1) Readiness to Proceed Form:

<https://researchsurvey.ets.org/s3/Alignment-Workshop-Ready-to-Proceed-Form>

2) Complete the alignment judgement for TOEIC Bridge and TOEIC test form A's and enter your ratings into the excel rating files.

- Test Form Location: [ETS Alignment Study Panel 1](#)

*ets

19

Appendix B: Readiness Form and Evaluation Survey**ETS Alignment Workshop****READINESS TO PROCEED EVALUATION SHEET**

ID #: _____

The purpose of this form is to verify that you understand the general purpose of the alignment workshop and believe that you have received sufficient information and explanations to complete your alignment judgments.

	Yes	No
I understand the purpose of the alignment workshop.	☐ Y	☐ N
I understand the steps I am to follow to make my alignment judgments.	☐ Y	☐ N
I am ready to complete my alignment judgments.	☐ Y	☐ N

If you responded “No” to any of the statements, please indicate what additional information or explanations you need.

 (Date)

 (Signature)

 (Print Name)

Alignment Evaluation Survey

Note. The following survey questions were administered using an Alchemer online survey form.

Thank you for participating in the alignment study panel. Please take a few minutes to complete the alignment evaluation survey.

Part A. Overall Evaluation

Please indicate the degree to which you agree with each of the following statements.

(Scale: Strongly disagree, Disagree, Agree, Strongly agree)

- I understood the purpose of the study.
- I understood the key language knowledge, skills, and abilities described in the NRS EFL descriptors.
- I understood the distinguishing features of the six EFL descriptor levels (Levels 1 to 6).
- I became more familiar with the NRS EFL descriptors as I completed the alignment ratings.
- I understood the intended language knowledge and skills assessed in the test items.
- The instructions and explanations provided by the facilitators were clear.
- The facilitators were able to answer questions, provide information, and clarify points when needed.
- The explanations of consensus ratings for disagreement items were clear.
- There was adequate time to complete the alignment ratings.

Part B. Materials

Please indicate the degree to which the following materials were helpful during the alignment rating process.

(Scale: Not helpful at all, Not helpful, Helpful, Very helpful)

- Test specification and blueprint documents
- Original EFL descriptor document (PDF)

- EFL descriptors in table format with highlighting
- NRS level notes
- Panel meeting slides

Part C. Comfort Level With Ratings

Please indicate how comfortable you felt with your alignment ratings.

Select the panel you are on: Panel 1 (Reading/Writing), Panel 2 (Listening/Speaking)

(Scale: Very uncomfortable, Somewhat uncomfortable, Somewhat comfortable, Very comfortable)

- Content match for listening items
- Level match for listening items
- Content match for speaking items
- Level match for speaking items

Part D. Suggestions

- Please provide any comments or suggestions about the alignment meetings and process.
- We'd also appreciate your thoughts on any potential areas for improvement in the TOEIC/TOEIC Bridge tests to better serve the adult education ESL population. Please share any comments or suggestions.

